Tomlinson-Harashima Precoding for Broadcast Channels with Uncertainty

Michael Botros Shenouda and Timothy N. Davidson, *Member, IEEE,*

***Abstract*— We consider the design of Tomlinson-Harashima (TH) precoders for broadcast channels in the presence of channel uncertainty. For systems in which uplink-downlink reciprocity is used to obtain a channel estimate at the transmitter, we present a robust design based on a statistical model for the channel uncertainty. We provide a convex formulation of the design problem subject to two types of power constraints: a set of constraints on the power transmitted from each antenna and a total power constraint. For the case of the total power constraint, we present a closed-form solution for the robust TH precoder that incurs essentially the same computational cost as the corresponding designs that assume perfect channel knowledge. For systems in which the receivers feed back quantized channel state information to the transmitter, we present a robust design based on a bounded model for the channel uncertainty. We provide a convex formulation for the TH precoder that maximizes the performance under the worst-case channel uncertainty subject to both types of power constraints. We also present a conservative robust design for this type of channel uncertainty that has reduced computational complexity for the case of power constraints on individual antennas and leads to a closed-form solution for the total power constraint case. Simulation studies verify our analytical results and show that the robust TH precoders can significantly reduce the rather high sensitivity of broadcast transmissions to errors in channel state information.**

***Index Terms*— Tomlinson-Harashima precoding, broadcast channel, channel uncertainty, robust precoding.**

I. INTRODUCTION

TOMLINSON-HARASHIMA (TH) precoding was originally introduced as a non-linear transmitter equalization technique for single-input single-output (SISO) channels with inter-symbol interference (ISI) [1], [2]. For ISI channels, TH precoding works by pre-subtracting the interference that the previous symbols would otherwise create at the receiver. The same principle can be applied to broadcast channels in which the transmitter uses multiple antennas to transmit independent data streams to decentralized users. In that scenario, TH precoding pre-subtracts previously precoded symbols intended for other receivers, thus performing spatial (pre-) equalization, rather than temporal equalization. This interference presubtraction at the transmitter is well suited to the broadcast scenario because the decentralized nature of the receivers prevents joint processing of the received signals.

A fundamental assumption of TH precoding is the availability of perfect Channel State Information (CSI) at the transmitter. Perfect CSI enables the transmitter to precisely pre-subtract the terms that would interfere at the receiver. Based on the assumption of perfect CSI at the transmitter, several different approaches for designing TH precoders for broadcast channels have been proposed, including zero-forcing designs [3]–[6], minimum mean square error (MMSE) [7], [8] and designs with independent mean square error constraints [9]. However, in practical broadcasting systems, such as the downlink of a cellular system, the CSI available at the transmitter is generally imperfect. In systems in which the transmitter exploits uplink-downlink reciprocity to estimate the downlink channel, the uncertainty in the CSI at the transmitter is dominated by the effect of estimation errors, and in systems in which each receiver quantizes its channel information and feeds it back to the transmitter, the uncertainty in the CSI at the transmitter is dominated by the effect of quantization errors. In both scenarios, the mismatch between the actual CSI and the transmitter's estimate of the CSI can result in serious degradation of the performance of TH precoding. Analysis of TH precoding for uncertain ISI channels has shown that the received signal to noise ratio (SNR) saturates as the transmitted power is increased [10]. Moreover, precoding for the broadcast channel is highly sensitive to CSI mismatch. Even for a simple linear precoding scheme, recent results have shown that the receivers' signal to interference plus noise ratios (SINRs) saturates with increasing transmitted power [11].

Motivated by the sensitivity of both broadcast channels and TH precoding to channel uncertainty, we design, herein, robust TH precoders that explicitly take into account the nature of CSI mismatch. For scenarios in which CSI mismatch results from estimation errors, we propose a statistical model for the channel uncertainty. We provide a convex formulation for the robust TH precoder that maximizes the average performance under two possible power constraints: a set of constraints on the power transmitted from each antenna, and a total power constraint. Furthermore, we obtain a closed-form solution for the robust precoder under the total power constraint. When the CSI mismatch results from quantization errors, the transmitter can bound the dominant errors using its knowledge of the quantization codebooks. For this scenario, we design a robust TH precoder that maximizes the worst-case performance. For this minimax strategy, we provide convex formulations for precise and conservative approaches to the worst-case design

Manuscript received 15 June 2006; revised 19 February 2007. This work was supported in part by Communications and Information Technology, Ontario. The work of the second author is also supported by the Canada Research Chairs Program. A preliminary version of a portion of this work appears in *Proc. 44th Ann. Allerton Conf. Communications, Control, Computing*, Monticello, IL, Sept. 2006.

The authors are with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada (e-mail: {botrosmw,davidson}@mcmaster.ca).

Digital Object Identifier 10.1109/JSAC.2007.070911.

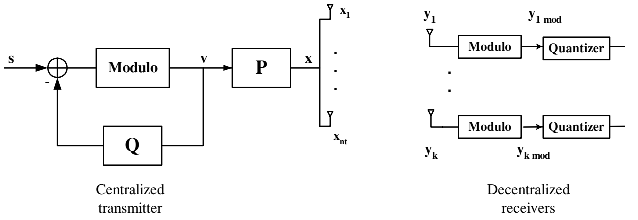

Fig. 1. Multiple-input single-output downlink system with Tomlinson-Harashima precoding at the transmitter.

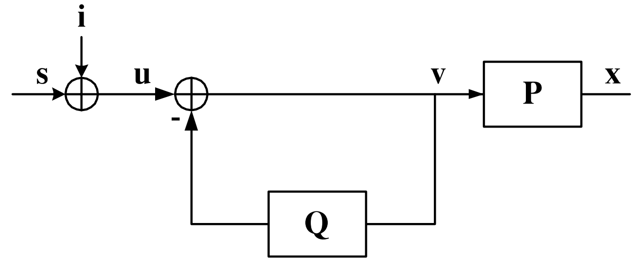

Fig. 2. Equivalent linear model for the transmitter.

under the two types of power constraints. The conservative approach provides an alternative design that achieves performance close to that of the true worst-case design, yet has a lower computational cost for the case of power constraints on individual antennas, and leads to a closed-form solution for the case of a total power constraint.

Previous approaches to the design of robust TH precoders have considered the statistical model of channel uncertainty in scenarios that involve equalization of a SISO ISI channel [12], or precoding for a point-to-point MIMO system [13]. For the broadcast channel, a zero-forcing robust design under the statistical uncertainty model was proposed in [14]. Besides considering the additional bounded model of channel uncertainty, the approach proposed herein enables power constraints on individual antennas to be incorporated, in addition to a constraint on the total transmitted power.

Our notation is as follows: Boldface type is used to denote matrices and vectors; $[\mathbf{Q}]_{i,j}$ denotes the element at the intersection of the i^{th} row and j^{th} column of the matrix $\mathbf{Q}$; and $\mathbf{P}^H$ denotes the conjugate transpose of the matrix $\mathbf{P}$. The notation $\|\mathbf{x}\|$ denotes the Euclidean norm of vector $\mathbf{x}$, while $\|\mathbf{E}\|$ denotes the spectral norm (maximum singular value) of the matrix $\mathbf{E}$. The term $\text{tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$, $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$, and for symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \geq \mathbf{B}$ denotes the fact that $\mathbf{A} - \mathbf{B}$ is positive semidefinite. The notation $\text{Diag}(\mathbf{x})$ denotes the diagonal matrix whose diagonal elements are the elements of $\mathbf{x}$.

II. System Model

We consider the downlink of a multiuser cellular communication system with n_t antennas at the transmitter and K users, each with one receive antenna. We consider downlink systems in which Tomlinson-Harashima (TH) precoding is used at the transmitter for multi-user interference pre-subtraction. As shown in Fig. 1, interference pre-subtraction and channel spatial equalization are performed at the transmitter using a feedback precoding matrix $\mathbf{Q} \in \mathbb{C}^{K \times K}$ and a feedforward precoding matrix $\mathbf{P} \in \mathbb{C}^{n_t \times K}$. The vector $\mathbf{s} \in \mathbb{C}^K$ contains the data symbol destined for each user, and we assume that s_k is chosen from a square QAM constellation $\mathcal{S}$ with cardinality M. The Voronoi region, $\mathcal{V}$, of the constellation is a square whose side length is D; i.e., $D = \sqrt{M}\, d$, where d is the distance between two successive constellation points along any of the basis directions.

In absence of the modulo operation, the output symbols of the feedback loop in Fig. 1, v_k, would be generated successively according to the following relation

$$v_k = s_k - \sum_{j=1}^{k-1} [\mathbf{Q}]_{k,j} v_j, \tag{1}$$

where, at the k^{th} step, only the previously precoded symbols $v_1, .., v_{k-1}$ are subtracted. Hence, $\mathbf{Q}$ is a strictly lower triangular matrix. The summation in equation (1) suggests that the magnitude of v_k may grow beyond the boundaries of $\mathcal{V}$. The role of the modulo operation is to bring the magnitude back inside the boundaries of $\mathcal{V}$. (For square QAM symbols, the modulo operation corresponds to performing separate modulo-D operations on the real and imaginary parts of v_k.) The effect of the modulo operation is equivalent to the addition of the complex quantity $i_k = i_k^{re}\, D + j\, i_k^{imag}\, D$ to v_k, where $i_k^{re},\ i_k^{imag} \in \mathbb{Z}$, and $j = \sqrt{-1}$. Using this observation, we obtain the standard linear model of the transmitter that does not involve a modulo operation, as shown in Fig. 2; e.g., [15]. In this model, the constellation of the modified data symbols in the vector $\mathbf{u} = \mathbf{s} + \mathbf{i}$ is simply the periodic extension of the original constellation $\mathcal{S}$ along the real and imaginary axes. From this equivalent model, it is clear that $\mathbf{v}$ is linearly related to the modified data vector $\mathbf{u}$,

$$\mathbf{v} = (\mathbf{I} + \mathbf{Q})^{-1}\mathbf{u}. \tag{2}$$

Following the feedback processing, the vector $\mathbf{v}$ is then linearly precoded to produce the vector of transmitted signals:

$$\mathbf{x} = \mathbf{P}\mathbf{v}. \tag{3}$$

The signals received at each user, y_k, can be written in the vector form:

$$\begin{aligned}\mathbf{y} &= \mathbf{H}\mathbf{x} + \mathbf{n} \\ &= \mathbf{H}\mathbf{P}(\mathbf{I} + \mathbf{Q})^{-1}\mathbf{u} + \mathbf{n},\end{aligned}$$

where the k^{th} row of $\mathbf{H}$ is the vector of channel coefficients for the k^{th} user and $\mathbf{n}$ is the vector of additive noise which is assumed to have zero mean and a covariance matrix $\mathbf{R}_n = \text{E}\{\mathbf{n}\mathbf{n}^H\}$. The transmitter's estimate of the channel will be denoted by $\hat{\mathbf{H}}$, and the error between the actual channel $\mathbf{H}$ and $\hat{\mathbf{H}}$ will be denoted by $\mathbf{E}$; i.e., $\mathbf{H} = \hat{\mathbf{H}} + \mathbf{E}$.

At the receivers, the modulo operation is necessary to eliminate the effect of the periodic extension of the constellation. Following the modulo operation, the receivers use a quantizer of the original constellation $\mathcal{S}$ to detect the received symbols. Since the transmitter processing and channel propagation may result in a non-unity gain, the receivers have to compensate for this gain g_k by dividing by the factor g_k prior to the modulo operation; e.g. [7], [16]. Alternatively, the modulo operation and the quantizer can incorporate this gain by assuming a scaled constellation $g_k\mathcal{S}$ instead of $\mathcal{S}$; e.g. [17]. We will assume that the second approach is implemented by the receivers. In this paper, we will focus on precoder design, and, analogous to existing broadcast precoder designs for the case of perfect channel knowledge (c.f. [16], [17]), we will choose g_k to be a function of the channel estimate, $\hat{\mathbf{H}}$.[1]

A. The performance metric

Motivated by [7], [19], the performance metric that we will use in our designs is the mean squared error (MSE) between the received signal $\mathbf{y}$ and a scaled version of the modified data symbols $\mathbf{u} = \mathbf{s} + \mathbf{i}$. In the absence of noise, this is equivalent to the MSE between $\mathbf{y}_{\text{mod}}$, the vector of inputs to the decision devices at the (decentralized) receivers, and the data symbols $\mathbf{s}$. (Recall that the receivers perform a modulo g_kD operation before the decision device.) This performance metric can be written as:

$$\begin{aligned}\text{MSE}(\mathbf{E},\mathbf{P},\mathbf{Q}) &= \text{E}\{\|\mathbf{G}\mathbf{u}-\mathbf{y}\|^2\}\\ &= \text{E}\{\|\mathbf{G}(\mathbf{I}+\mathbf{Q})\mathbf{v}-(\mathbf{H}\mathbf{P}\mathbf{v}+\mathbf{n})\|^2\}\\ &= \text{tr}\{(\mathbf{G}\mathbf{R}_v^{\frac{1}{2}}+\mathbf{G}\tilde{\mathbf{Q}}-(\hat{\mathbf{H}}+\mathbf{E})\tilde{\mathbf{P}})^H\\ &\quad\times(\mathbf{G}\mathbf{R}_v^{\frac{1}{2}}+\mathbf{G}\tilde{\mathbf{Q}}-(\hat{\mathbf{H}}+\mathbf{E})\tilde{\mathbf{P}})\}+\text{tr}(\mathbf{R}_n),\end{aligned} \tag{4}$$

where $\mathbf{G} = \text{Diag}([g_1, g_2, \ldots, g_K])$, $\mathbf{R}_v = \text{E}\{\mathbf{v}\mathbf{v}^H\}$, $\tilde{\mathbf{P}} = \mathbf{P}\mathbf{R}_v^{\frac{1}{2}}$ and $\tilde{\mathbf{Q}} = \mathbf{Q}\mathbf{R}_v^{\frac{1}{2}}$. The second term in (4) is independent of $\mathbf{P}$ and $\mathbf{Q}$, and therefore only the first term will appear in the objectives below.

In order to employ the expression in (4) we need to obtain (an approximation of) the auto-correlation matrix, $\mathbf{R}_v$, of the output of the transmitter's modulo operation, $\mathbf{v}$. To do so, we will make the standard observation that the elements of $\mathbf{v}$ are almost uncorrelated and uniformly distributed over the Voronoi region of the constellation, $\mathcal{V}$, and that such a model becomes more precise as M increases [15, Th. 3.1], [20]. Furthermore, for a square constellation, we have $\text{E}\{|v_k|^2\} = \frac{M}{M-1}\text{E}\{|s_k|^2\}$ for $k = 2, \ldots, K$, [15]. Using these properties and the assumption that $\text{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$, it follows that $\mathbf{R}_v$ can be accurately approximated by a diagonal matrix with $[\mathbf{R}_v]_{1,1} = 1$ and $[\mathbf{R}_v]_{k,k} = \frac{M}{M-1}$, $k = 2, \ldots, K$; see also [3].

B. Transmitter power constraints

We will consider two possible power constraints: a constraint on the total transmitted power $\text{E}\{\mathbf{x}^H\mathbf{x}\} \leq P_{\text{total}}$ and a set of constraints on the power transmitted from each antenna $\text{E}\{|x_i|^2\} \leq P_i,\ i = 1, \ldots, n_t$. Power constraints on individual antennas are of considerable interest, because in most practical implementations each antenna has its own power amplifier. Individual antenna power constraints are also appropriate in multi-cell scenarios, in which each antenna belongs to a different base station. A transmitter design for a broadcast channel with such "per-antenna" constraints was considered in [21] under the assumption that perfect CSI is available at the transmitter. Without loss of generality, we can assume that $\text{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$. Hence, the total power constraint simplifies to $\text{tr}(\tilde{\mathbf{P}}^H\tilde{\mathbf{P}}) \leq P_{\text{total}}$, and the set of power constraints on individual antennas simplifies to $\|\tilde{\mathbf{p}}_i\|^2 \leq P_i,\ i = 1, \ldots, n_t$, where $\tilde{\mathbf{p}}_i$ is the i^{th} row of $\tilde{\mathbf{P}}$.

C. Channel uncertainty models

We will develop design formulations for robust precoders under two models of channel uncertainty:

- For systems with reciprocity in which the transmitter performs channel estimation, the error $\mathbf{E}$ between the actual channel $\mathbf{H}$ and the estimated channel $\hat{\mathbf{H}}$ can often be accurately modelled as a Gaussian random variable with zero mean and $\text{E}\{\mathbf{E}^H\mathbf{E}\} = \sigma_E^2\mathbf{I}$, [22]. For this stochastic model of uncertainty, robust precoding based on the average MSE will be presented in Section III.
- In the second model, the error is assumed to be deterministically bounded, $\|\mathbf{E}\| \leq \Delta$. This model is suitable for certain systems that involve quantization of the channel information; see, e.g., [23]. For this model of uncertainty, robust minimax precoding based on the worst-case MSE will be presented in Sections IV and V.

III. Statistically Robust Tomlinson-Harashima Precoding

In this section, we present an approach to robust Tomlinson-Harashima precoder design for the first model of uncertainty. Our goal is to design a precoder that minimizes the average value, over the channel estimation errors, of the MSE in (4). For the case of individual power constraints on each antenna, we will show that the design problem is equivalent to a convex conic programming problem that can be efficiently solved using interior points methods. For the case where the power constraints on each individual antenna are replaced by a total power constraint, we will derive a closed-form expression for the optimal precoder.

A. Statistically robust TH Precoding with individual antenna power constraints

The average, over channel estimation error $\mathbf{E}$, of the MSE in (4) can be written as:

$$\text{E}_{\mathbf{E}}\{\text{MSE}(\mathbf{E},\tilde{\mathbf{P}},\tilde{\mathbf{Q}})\} = \text{E}_{\mathbf{E}}\{\|\text{vec}(\mathbf{G}\mathbf{R}_v^{\frac{1}{2}}+\mathbf{G}\tilde{\mathbf{Q}}-(\hat{\mathbf{H}}+\mathbf{E})\tilde{\mathbf{P}})\|^2\},$$

where the vec operator stacks the columns of the input matrix into one vector. We will write $\text{vec}(\mathbf{G}\mathbf{R}_v^{\frac{1}{2}}+\mathbf{G}\tilde{\mathbf{Q}}-(\hat{\mathbf{H}}+\mathbf{E})\tilde{\mathbf{P}}) = \mathbf{b} + \text{vec}(\mathbf{G}\tilde{\mathbf{Q}}) - \mathbf{A}\tilde{\mathbf{p}}$, where

$$\mathbf{b} = \text{vec}(\mathbf{G}\mathbf{R}_v^{\frac{1}{2}}), \tag{5}$$

$$\tilde{\mathbf{p}} = \text{vec}(\tilde{\mathbf{P}}), \tag{6}$$

$$\mathbf{A} = \mathbf{I}\otimes(\hat{\mathbf{H}}+\mathbf{E}) = \hat{\mathbf{A}} + \mathbf{\Delta}_A, \tag{7}$$

with $\hat{\mathbf{A}} = \mathbf{I}\otimes\hat{\mathbf{H}}$ and $\mathbf{\Delta}_A = \mathbf{I}\otimes\mathbf{E}$. Because of the strictly lower triangular structure of $\mathbf{G}\tilde{\mathbf{Q}}$, $\text{vec}(\mathbf{G}\tilde{\mathbf{Q}})$ will contain only $L = K(K-1)/2$ non-zero elements. We will find it convenient to write $\text{vec}(\mathbf{G}\tilde{\mathbf{Q}})$ as the product of a matrix $\mathbf{D} \in \{0,1\}^{K^2\times L}$ and a vector $\tilde{\mathbf{q}} \in \mathcal{C}^{L\times 1}$ that contains only

[1] For the case of statistically-robust linear precoding, a joint design of the precoder and receiver under a zero-forcing constraint appears in [14], and a jointly optimal minimum mean square error design appears in [18].

the non-zero elements of vec($\mathbf{G}\tilde{\mathbf{Q}}$). The matrix $\mathbf{D}$ is given by:

$$\mathbf{D}^T = \left[\mathbf{0}_{L\times 1}, \mathbf{e}_1, \ldots, \mathbf{e}_{K-1} \mid \mathbf{0}_{L\times 2}, \mathbf{e}_K, \ldots, \mathbf{e}_{2K-3} \mid \ldots \mid \mathbf{0}_{L\times K}\right], \tag{8}$$

where $\mathbf{e}_i$ is the i^{th} column of $\mathbf{I}$. By using this notation, we avoid adding equality constraints to the subsequent formulations of the problem. The average MSE can now be written as:

$$\begin{aligned} \mathrm{E}_{\mathbf{E}}\{\mathrm{MSE}(\mathbf{E}, \tilde{\mathbf{P}}, \tilde{\mathbf{Q}})\} &= \mathrm{E}_{\mathbf{E}}\{(\mathbf{b} + \mathbf{D}\tilde{\mathbf{q}} - \mathbf{A}\tilde{\mathbf{p}})^H \\ &\quad \times (\mathbf{b} + \mathbf{D}\tilde{\mathbf{q}} - \mathbf{A}\tilde{\mathbf{p}})\} \\ &= \|\hat{\mathbf{A}}\tilde{\mathbf{p}} - \mathbf{D}\tilde{\mathbf{q}} - \mathbf{b}\|^2 \\ &\quad + \tilde{\mathbf{p}}^H \mathrm{E}_{\mathbf{E}}\{{\mathbf{\Delta}_A}^H \mathbf{\Delta}_A\}\tilde{\mathbf{p}} \\ &= \|\hat{\mathbf{A}}\tilde{\mathbf{p}} - \mathbf{D}\tilde{\mathbf{q}} - \mathbf{b}\|^2 + \sigma_E^2\|\tilde{\mathbf{p}}\|^2, \end{aligned} \tag{9}$$

where we have used the fact that ${\mathbf{\Delta}_A}^H\mathbf{\Delta}_A = \mathbf{I}\otimes\mathbf{E}^H\mathbf{E}$. Using the above expression for the average MSE, the design problem reduces to:

$$\min_{\tilde{\mathbf{p}},\tilde{\mathbf{q}}} \quad f_0(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) = \|\hat{\mathbf{A}}\tilde{\mathbf{p}} - \mathbf{D}\tilde{\mathbf{q}} - \mathbf{b}\|^2 + \sigma_E^2\|\tilde{\mathbf{p}}\|^2 \tag{10a}$$

$$\text{subject to} \quad \|\mathbf{S}^i\tilde{\mathbf{p}}\|^2 \leq P_i, \quad i = 1, \ldots, n_t, \tag{10b}$$

where $\mathbf{S}^i$ is the selection matrix that selects from $\tilde{\mathbf{p}}$ the elements corresponding to the i^{th} row of $\tilde{\mathbf{P}}$. The elements of $\mathbf{S}^i$ can be written as $[\mathbf{S}^i]_{m,n} = \delta(i + (m-1)n_t - n)$, where $\delta(k)$ is the Kronecker delta.

Since $\tilde{\mathbf{q}}$ in (10) is unconstrained, we can obtain a closed-form expression for the optimal value for $\tilde{\mathbf{q}}$ for each value of $\tilde{\mathbf{p}}$, and hence we can reduce the dimensionality of the optimization problem. Indeed, by solving the optimality condition $\nabla_{\tilde{\mathbf{q}}} f_0 = \mathbf{0}$, [24], we obtain the following expression for the optimal value of $\tilde{\mathbf{q}}$:

$$\tilde{\mathbf{q}} = (\mathbf{D}^H\mathbf{D})^{-1}\mathbf{D}^H(\hat{\mathbf{A}}\tilde{\mathbf{p}} - \mathbf{b}) = \mathbf{D}^H\hat{\mathbf{A}}\tilde{\mathbf{p}}, \tag{11}$$

where we used the facts that $\mathbf{D}^H\mathbf{D} = \mathbf{I}$ and $\mathbf{D}^H\mathbf{b} = \mathbf{0}$, which follow from (5) and (8). The design of the statistically robust TH precoder under individual antenna power constraints can now be written as:

$$\min_{\tilde{\mathbf{p}}} \quad \|\mathbf{J}\hat{\mathbf{A}}\tilde{\mathbf{p}} - \mathbf{b}\|^2 + \sigma_E^2\|\tilde{\mathbf{p}}\|^2 \tag{12a}$$

$$\text{subject to} \quad \|\mathbf{S}^i\tilde{\mathbf{p}}\|^2 \leq P_i \quad i = 1, \ldots, n_t, \tag{12b}$$

where

$$\mathbf{J} = \mathbf{I} - \mathbf{D}\mathbf{D}^H = \mathrm{diag}(\underbrace{1, 0, \ldots, 0}_{K}, \underbrace{1, 1, \ldots, 0}_{K}, \ldots, \underbrace{1, 1, \ldots, 1}_{K}). \tag{13}$$

Equation (12a) shows that the robust precoding problem is a form of constrained regularized least-squares, in which the core term in the objective is the MSE for the estimated channel (the first term) and the regularizing term in the objective is the total precoder power (the second term). This regularizing term captures the impact of the interference at the receivers that results from imperfect channel state information at the transmitter. Indeed, the consequences of this interference have been previously observed in the saturation of the bit error performance with increasing transmitted power [13], [14]. Similar saturation effects with increasing transmitted power were observed in the SINR of the received signals; e.g. [11]. In the model that we consider, the receiver processing gains are function of the channel estimate only, and as a result the optimal precoder does not necessarily use all the available transmission power; i.e., the constraint in (12b) is not necessarily active at optimality. In fact, in the next section we will provide conditions under which the precoder in (12) does not use all the available transmission power. This does not entail significant loss of performance because of the saturation effects mentioned above. A joint design of the precoder and the receivers' processing gains can be obtained in analogous way to that for the robust statistical design for linear transceivers presented in [18].

The problem in (12) is convex. In particular, it can be rewritten as the following rotated second order cone program (SOCP) [25]:

$$\min_{\tilde{\mathbf{p}},t_1,t_2} \quad t_1 + \sigma_E^2 t_2 \tag{14a}$$

$$\text{subject to} \quad \|\mathbf{J}\hat{\mathbf{A}}\tilde{\mathbf{p}} - \mathbf{b}\|^2 \leq t_1, \tag{14b}$$

$$\|\tilde{\mathbf{p}}\|^2 \leq t_2, \tag{14c}$$

$$\|\mathbf{S}_i\tilde{\mathbf{p}}\|^2 \leq P_i \quad i = 1, \ldots, n_t. \tag{14d}$$

This problem can be efficiently solved using self-dual interior point methods [26]–[28].

B. Closed-form solution for statistically robust TH precoding with a total power constraint

In this section we address the problem of designing a statistically robust TH precoder under a constraint on the total transmitted power. That is, we wish to solve the problem

$$\min_{\tilde{\mathbf{p}}} \quad \|\mathbf{J}\hat{\mathbf{A}}\tilde{\mathbf{p}} - \mathbf{b}\|^2 + \sigma_E^2\|\tilde{\mathbf{p}}\|^2 \tag{15a}$$

$$\text{subject to} \quad \|\tilde{\mathbf{p}}\|^2 \leq P_{\text{total}}. \tag{15b}$$

(The optimal $\tilde{\mathbf{q}}$ remains that given by equation (11).) Although the convex optimization formulation in (14) can be slightly modified to obtain an efficiently-solvable formulation of this design problem, we can actually obtain a closed-form solution.

The design problem in (15) is a regularized least squares problem with an inequality constraint on the Euclidean norm of the variable. By solving the Karush-Kuhn-Tucker optimality conditions for the problem in (15) and by using the fact that $\mathbf{J}\mathbf{b} = \mathbf{b}$, we can obtain (see Appendix I for the details) the following closed-form solution for the optimal $\tilde{\mathbf{p}}$:

$$\tilde{\mathbf{p}}^{\text{opt}} = \begin{cases} (\hat{\mathbf{A}}^H\mathbf{J}^2\hat{\mathbf{A}} + (\lambda + \sigma_E^2)\mathbf{I})^{-1}\hat{\mathbf{A}}^H\mathbf{b} & \lambda > 0, \\ (\hat{\mathbf{A}}^H\mathbf{J}^2\hat{\mathbf{A}} + \sigma_E^2\mathbf{I})^{-1}\hat{\mathbf{A}}^H\mathbf{b} & \text{otherwise,} \end{cases} \tag{16}$$

where λ is the Lagrange multiplier associated with the inequality constraint (15b). The value of λ is given by the unique positive root (if any) of the equation:

$$f(x) = \mathrm{tr}(\hat{\mathbf{A}}^H\mathbf{b}\mathbf{b}^H\hat{\mathbf{A}}(\hat{\mathbf{A}}^H\mathbf{J}^2\hat{\mathbf{A}} + (\sigma_E^2 + x)\mathbf{I})^{-2})) - P_{\text{total}} = 0, \tag{17}$$

Moreover, this root can be found using bisection search.

The closed-form expression in (16) for the robust precoder contains two cases. The case that is applicable depends on the root of equation (17). If a positive root exists, then it will be unique and λ will be equal to the value of this root. Hence, the first case of the optimal solution will be applicable and the transmitted power constraint in (15b) is satisfied with equality. If no positive root exists, then the alternate expression for $\tilde{\mathbf{p}}^{\text{opt}}$ in (16) is applicable and $\lambda = 0$. In that case, the constraint in (15b) will not be active at optimality.

Comparing the optimal solution for the robust TH precoder design given in (16) to the MMSE-TH precoder design method that assumes perfect channel knowledge [7], it can be seen that the robust solution has essentially the same computational cost. To show this, we observe first the special structure of the matrix $\mathbf{J}\hat{\mathbf{A}}$. From (7), it can be seen that $\hat{\mathbf{A}}$ is block diagonal with all blocks equal to $\hat{\mathbf{H}}$. Using (13), we can see that $\mathbf{JA}$ is a block diagonal matrix with the first i rows of the i^{th} block being the first i rows of $\hat{\mathbf{H}}$ and remaining rows being zero. The computational complexity of calculating $\tilde{\mathbf{p}}^{\text{opt}}$ is dominated by the complexity of calculating the regularized inversion of these blocks, and hence is essentially same as the MMSE-TH precoding method in [7] whose expression also involves the regularized inversion of same blocks. Hence, robustness can be incorporated into the TH precoder without significant additional computational cost.

By using the singular value decomposition (SVD) of the matrix $\mathbf{J}\hat{\mathbf{A}}$, we can obtain simpler expressions for the optimal $\tilde{\mathbf{p}}$ and equation (17). This will also enable us to obtain necessary and sufficient condition for the robust TH precoder to use all the allowable transmission power. Let the SVD of $\mathbf{J}\hat{\mathbf{A}}$ be given by:

$$\mathbf{J}\hat{\mathbf{A}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H = \begin{bmatrix}\mathbf{U}_1 & \mathbf{U}_2\end{bmatrix}\begin{bmatrix}\mathbf{\Sigma}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}_1^H\\ \mathbf{V}_2^H\end{bmatrix}. \quad (18)$$

If we substitute the SVD of $\mathbf{J}\hat{\mathbf{A}}$ into (16) and use $\tilde{\mathbf{b}}_1 = \mathbf{U}_1^H\mathbf{b}$, the expression for $\tilde{\mathbf{p}}^{\text{opt}}$ can be written as (see Appendix I):

$$\tilde{\mathbf{p}}^{\text{opt}} = \begin{cases} \mathbf{V}_1(\mathbf{\Sigma}_1^2 + (\sigma_E^2 + \lambda)\mathbf{I})^{-1}\mathbf{\Sigma}_1\tilde{\mathbf{b}}_1 & \lambda > 0,\\ \mathbf{V}_1(\mathbf{\Sigma}_1^2 + \sigma_E^2\mathbf{I})^{-1}\mathbf{\Sigma}_1\tilde{\mathbf{b}}_1 & \text{otherwise,}\end{cases} \quad (19)$$

and the equation for calculating the Lagrange multiplier reduces to:

$$f(x) = \tilde{\mathbf{b}}_1^H\,\mathbf{\Sigma}_1^2\,(\mathbf{\Sigma}_1^2 + (x+\sigma_E^2)\mathbf{I})^{-2}\,\tilde{\mathbf{b}}_1 - P_{\text{total}} \quad (20)$$

$$= \sum_{i=1}^{r}\beta_i\frac{\sigma_i^2}{(\sigma_i^2 + (x+\sigma_E^2))^2} - P_{\text{total}} = 0, \quad (21)$$

where r is the rank of $\mathbf{J}\hat{\mathbf{A}}$, β_i is the i^{th} diagonal element of $\tilde{\mathbf{b}}_1\tilde{\mathbf{b}}_1^H$ and σ_i is the i^{th} diagonal element of $\mathbf{\Sigma}_1$. The next lemma shows that there is a threshold value for P_{total} above which there will be no positive root for (21).

Theorem 1: Let $N = \sum_{i=1}^{r}\beta_i\frac{\sigma_i^2}{(\sigma_i^2+\sigma_E^2)^2}$. A necessary and sufficient condition for (17) to have a unique positive root is $P_{\text{total}} < N$.

Proof: See Appendix II. ■

This result shows that by using only the parameters of the design problem, we can decide which case of the optimal solution in (16) or (19) is applicable.

IV. Minimax Robust Tomlinson-Harashima Precoding

In this section we present a robust Tomlinson-Harashima precoder design that does not rely on a statistical model of channel uncertainty, but merely assumes that the channel uncertainty lies within a given bound. This type of uncertainty can arise in systems in which the channel knowledge is quantized at the receiver and fed back to the transmitter. The transmitter in this case can bound the quantization errors through its knowledge of the codebooks used by the receiver. For this type of channel uncertainty, our goal is to design the Tomlinson-Harashima precoder to minimize the worst-case MSE over all admissible channel uncertainties $\mathbf{E}$, subject to power constraints on each antenna, or a total power constraint.

The robust TH precoder that minimizes the worst-case performance under individual antenna power constraints, can be stated as the following minimax optimization problem:

$$\min_{\tilde{\mathbf{P}},\tilde{\mathbf{Q}}}\ \max_{\|\mathbf{E}\|\le\Delta}\ \text{tr}\{(\mathbf{G}\mathbf{R}_v^{\frac{1}{2}} + \mathbf{G}\tilde{\mathbf{Q}} - (\hat{\mathbf{H}}+\mathbf{E})\tilde{\mathbf{P}})^H \times (\mathbf{G}\mathbf{R}_v^{\frac{1}{2}} + \mathbf{G}\tilde{\mathbf{Q}} - (\hat{\mathbf{H}}+\mathbf{E})\tilde{\mathbf{P}})\} \quad (22a)$$

$$\text{subject to}\quad \|\tilde{\mathbf{p}}_i\|^2 \le P_i, \quad i = 1,\ldots,n_t, \quad (22b)$$

$$\tilde{\mathbf{Q}}_{i,j} = 0, \quad 1 \le i \le j \le K. \quad (22c)$$

This problem can be re-cast as the following single minimization problem:

$$\min_{\tilde{\mathbf{P}},\tilde{\mathbf{Q}},t}\quad t \quad (23a)$$

$$\text{s.t.}\quad \text{tr}\{(\mathbf{G}\mathbf{R}_v^{\frac{1}{2}} + \mathbf{G}\tilde{\mathbf{Q}} - (\hat{\mathbf{H}}+\mathbf{E})\tilde{\mathbf{P}})^H \times (\mathbf{G}\mathbf{R}_v^{\frac{1}{2}} + \mathbf{G}\tilde{\mathbf{Q}} - (\hat{\mathbf{H}}+\mathbf{E})\tilde{\mathbf{P}})\} \le t \quad \forall\|\mathbf{E}\|\le\Delta, \quad (23b)$$

along with constraints (22b) and (22c). The expression in (23b) represents an infinite set of constraints, one for each $\mathbf{E}$ that satisfies $\|\mathbf{E}\| \le \Delta$. Each of these constraints is non-linear in the decision variables $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{Q}}$. The result of the following lemma will be useful in formulating this constraint as a Linear Matrix Inequality (LMI) constraint.

Lemma 1: Let $\mathbf{M} \in \mathcal{C}^{K\times K}$ be a Hermitian matrix such that $\mathbf{M} = ((\mathbf{C}+\mathbf{\Delta}_C)\mathbf{X} - \mathbf{F})^H((\mathbf{C}+\mathbf{\Delta}_C)\mathbf{X} - \mathbf{F})$. Then there exists $s, \lambda \ge 0$ and $\mathbf{Z} \ge \mathbf{0}$ such that the constraint $\text{tr}(\mathbf{M}) \le t$, $\forall\,\|\,\mathbf{\Delta}_G\| \le \Delta$ is equivalent to the following LMI representation:

$$\begin{bmatrix} \mathbf{Z} + s\mathbf{I} & (\mathbf{CX}-\mathbf{F})^H & -\Delta\mathbf{X}^H\\ (\mathbf{CX}-\mathbf{F}) & (1-\lambda)\mathbf{I} & \mathbf{0}\\ -\Delta\mathbf{X} & \mathbf{0} & \lambda\mathbf{I}\end{bmatrix} \ge \mathbf{0}, \quad (24)$$

$$t - Ks - \text{tr}(\mathbf{Z}) \ge 0. \quad (25)$$

Proof: The proof involves the application of properties of the Schur Complement to results in [29] and [30]. A similar proof for the case in which the constraint $\text{tr}(\mathbf{M}) \le t$ is replaced by $\lambda_{\max}(\mathbf{M}) \le t$, where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue, can be found in [31]. ■

Using the above lemma, the design of the minimax robust TH precoder under individual antenna power constraints can be formulated as the following Semidefinite Programming (SDP) problem:

$$\min_{\tilde{\mathbf{P}},\tilde{\mathbf{Q}},\mathbf{Z},s,t,\lambda} \quad t \tag{26a}$$

$$\text{s. t.} \begin{bmatrix} \mathbf{Z}+s\mathbf{I} & \mathbf{X}(\tilde{\mathbf{P}},\tilde{\mathbf{Q}})^H & -\Delta\tilde{\mathbf{P}}^H \\ \mathbf{X}(\tilde{\mathbf{P}},\tilde{\mathbf{Q}}) & (1-\lambda)\mathbf{I} & \mathbf{0} \\ -\Delta\tilde{\mathbf{P}} & \mathbf{0} & \lambda\mathbf{I} \end{bmatrix} \geq \mathbf{0}, \tag{26b}$$

$$t - Ks - \text{tr}(\mathbf{Z}) \geq 0, \tag{26c}$$

$$\mathbf{Z} \geq \mathbf{0}, \tag{26d}$$

$$\|\tilde{\mathbf{p}}_i\|^2 \leq P_i, \quad i=1,\ldots,n_t, \tag{26e}$$

$$\tilde{\mathbf{Q}}_{i,j} = 0, \quad 1 \leq i \leq j \leq K, \tag{26f}$$

where we have used the notation $\mathbf{X}(\tilde{\mathbf{P}},\tilde{\mathbf{Q}}) = \mathbf{G}\mathbf{R}_v^{\frac{1}{2}} + \mathbf{G}\tilde{\mathbf{Q}} - \hat{\mathbf{H}}\tilde{\mathbf{P}}$ to simplify (26b). This problem can be efficiently solved using general purpose implementations of interior point methods, such as SeDuMi [28]. In a similar way, the convex formulation for the minimax robust TH precoder under a total power constraint can be obtained by replacing (26e) by $\|\tilde{\mathbf{p}}\|^2 \leq P_{\text{total}}$.

V. Conservative Approach for Minimax Robust Tomlinson-Harashima Precoding

Although the SDP in (26) can be efficiently solved, its computational cost is significantly larger than that of the designs that were obtained for the statistical model of the channel uncertainty. To address this computational imbalance, in this section we will present a conservative minimax robust design that has a computational cost that is similar to that of the statistically robust designs. The proposed design is "conservative" in the sense that we admit uncertainties into the model that do not arise in practice.

In order to derive the conservative design, we observe that the expression for the MSE in (4) can be re-written as:

$$\text{MSE}(\mathbf{E},\tilde{\mathbf{p}},\tilde{\mathbf{q}}) = \|(\hat{\mathbf{A}}+\boldsymbol{\Delta}_A)\tilde{\mathbf{p}} - \mathbf{D}\tilde{\mathbf{q}} - \mathbf{b}\|^2, \tag{27}$$

where, as in Section II, $\boldsymbol{\Delta}_A = \mathbf{I}\otimes\mathbf{E}$. Therefore, the inner maximization in (22a) is equivalent to maximization of (27) over all $\boldsymbol{\Delta}_A$ with $\|\boldsymbol{\Delta}_A\| \leq \Delta$ and the block structure implicit in the fact that $\boldsymbol{\Delta}_A = \mathbf{I}\otimes\mathbf{E}$. The computational bottleneck in the precise minimax approach arises from the block structure of $\boldsymbol{\Delta}_A$. As we will show below, if we drop the structural constraints and simply minimize the maximum value of (27) over all $\|\boldsymbol{\Delta}_A\| \leq \Delta$, we can obtain designs with computational costs similar to the corresponding designs for the statistical model of the channel uncertainty.

The conservative minimax robust TH precoder under individual antenna power constraints is the solution to the following minimax optimization problem:

$$\min_{\tilde{\mathbf{p}},\tilde{\mathbf{q}}} \max_{\|\boldsymbol{\Delta}_A\|\leq\Delta} \|(\hat{\mathbf{A}}+\boldsymbol{\Delta}_A)\tilde{\mathbf{p}} - \mathbf{D}\tilde{\mathbf{q}} - \mathbf{b}\| \tag{28a}$$

$$\text{subject to} \quad \|\mathbf{S}^i\tilde{\mathbf{p}}\| \leq \sqrt{P_i}, \quad i=1,\ldots,n_t. \tag{28b}$$

Consider the inner maximization in (28a). This problem is convex and the maximum value is given by $\|\hat{\mathbf{A}}\tilde{\mathbf{p}} - \mathbf{D}\tilde{\mathbf{q}} - \mathbf{b}\| + \Delta\|\tilde{\mathbf{p}}\|$, [32]. Hence, the design problem reduces to:

$$\min_{\tilde{\mathbf{p}},\tilde{\mathbf{q}}} \|\hat{\mathbf{A}}\tilde{\mathbf{p}} - \mathbf{D}\tilde{\mathbf{q}} - \mathbf{b}\| + \Delta\|\tilde{\mathbf{p}}\| \tag{29a}$$

$$\text{subject to} \quad \|\mathbf{S}^i\tilde{\mathbf{p}}\| \leq \sqrt{P_i}, \quad i=1,\ldots,n_t. \tag{29b}$$

The above formulation of the problem is similar to that of the statistically robust design in (10). It differs only in the power of the norms in the objective. Similar to our approach in the statistical robustness case, we can perform the optimization with respect to $\tilde{\mathbf{q}}$ first. By doing so, we obtain the same expression for the optimal $\tilde{\mathbf{q}}$, namely (11). The remaining design problem is:

$$\min_{\tilde{\mathbf{p}}} \|\mathbf{J}\hat{\mathbf{A}}\tilde{\mathbf{p}} - \mathbf{b}\| + \Delta\|\tilde{\mathbf{p}}\| \tag{30a}$$

$$\text{subject to} \quad \|\mathbf{S}^i\tilde{\mathbf{p}}\| \leq \sqrt{P_i}, \quad i=1,\ldots,n_t. \tag{30b}$$

The above formulation shows that the conservative minimax design for TH precoder is equivalent to another form of constrained regularized least squares problem. The optimization problem (30) is convex and can be formulated as the following second-order cone programming problem (SOCP):

$$\min_{\tilde{\mathbf{p}},\tau_1,\tau_2} \tau_1 + \Delta\tau_2 \tag{31a}$$

$$\text{subject to} \quad \|\mathbf{J}\hat{\mathbf{A}}\tilde{\mathbf{p}} - \mathbf{b}\| \leq \tau_1, \tag{31b}$$

$$\|\tilde{\mathbf{p}}\| \leq \tau_2, \tag{31c}$$

$$\|\mathbf{S}^i\tilde{\mathbf{p}}\| \leq \sqrt{P_i} \quad i=1,\ldots,n_t. \tag{31d}$$

Similar to the conic formulation of the statistically robust precoding problem, the above optimization problem can be efficiently solved using self-dual interior point methods.

A similar formulation for the conservative minimax design problem with a total power constraint can be obtained from (30) by replacing (30b) by $\|\tilde{\mathbf{p}}\| \leq \sqrt{P_{\text{total}}}$. Although the resulting problem can be efficiently solved as a SOCP, a closed-form solution can also be obtained using techniques analogous to those used in [33].

VI. Simulation Studies

We simulated the proposed robust Tomlinson-Harashima precoders using QPSK signaling over an independent Rayleigh fading channel. The coefficients of the $\mathbf{H}$ matrix are modelled as being independent rotationally-symmetric complex Gaussian random variables with zero mean and unit variance, and the elements of the vector $\mathbf{n}$ are modelled as being independent rotationally symmetric complex Gaussian random variables with zero mean and equal variance, i.e., $\mathbf{R}_n$ is a diagonal matrix with equal diagonal entries. Following existing designs for the case of perfect channel state information at the transmitter (c.f. [7], [16], [17]), we will choose the same scaling factor for each receiver; i.e., $g_k = g$. Since our design is based on an MSE approach, we will choose g to be equal to the receiver scaling factor that would have been obtained for the MMSE TH precoder design in [7] if $\hat{\mathbf{H}}$ had been the actual channel. We plot the average bit error rate (BER) over all users against the signal-to-noise-ratio, which is defined as the ratio of the total average transmitted power to the total noise power; i.e., $\text{SNR} = \text{E}_{\mathbf{H}}\{\text{tr}(\mathbf{P}\mathbf{R}_v\mathbf{P}^H)\}/\text{tr}(\mathbf{R}_n)$. We compare the performance of our proposed designs with: two zero-forcing TH precoding designs, that in [5], [6] (denoted ZF1-THP) and that in [3], [4] (ZF2-THP); the minimum mean square error TH precoding design (MMSE-THP) in [7]; and the statistically robust zero-forcing TH precoding design (RZF-THP) in [14]. For these methods, the SNR is equivalent to

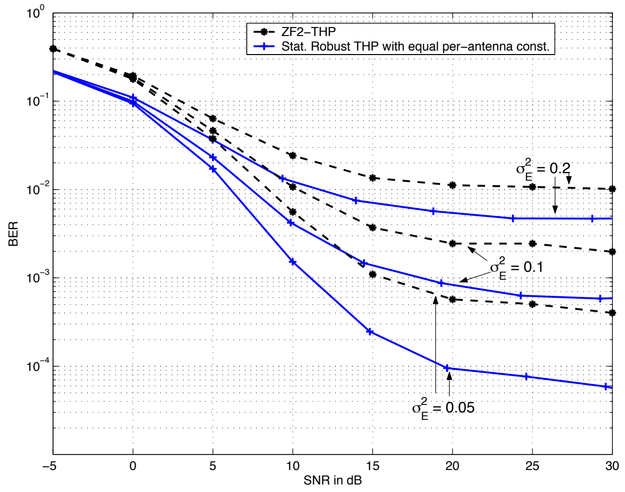

Fig. 3. Comparison between the performance of the proposed statistically robust precoder with equal antenna power constraints and the zero-forcing TH precoder (ZF2-THP) in [3], [4] for values of channel uncertainty $\sigma_E^2 = 0.05, 0.1$ for a system with $n_t = 4$ and $K = 3$, using QPSK signaling.

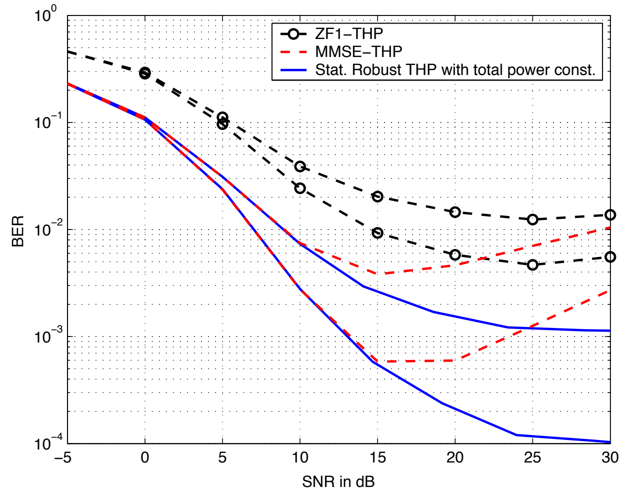

Fig. 4. Comparison between the performance of the proposed statistically robust precoder with a total power constraint, and the ZF1-THP [5], [6] and MMSE-THP [7] designs for values of channel uncertainty $\sigma_E^2 = 0.05, 0.1$ for a system with $n_t = 4$ and $K = 4$, using QPSK signaling.

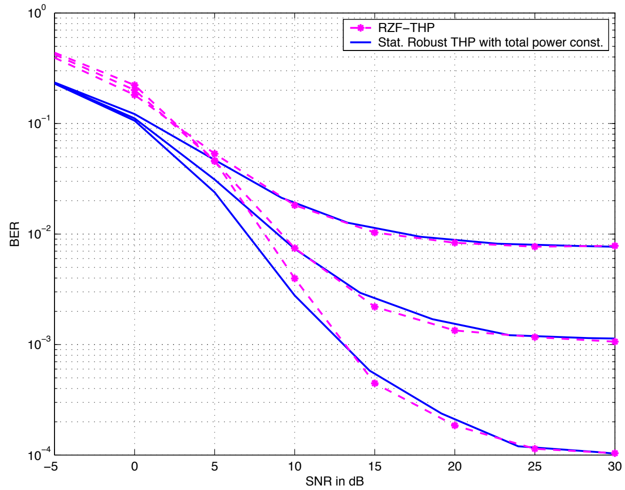

Fig. 5. Comparison between the performance of the proposed statistically robust precoder with a total power constraint and the robust zero-forcing TH precoder (RZF-THP) in [14] for values of channel uncertainty $\sigma_E^2 = 0.05, 0.1, 0.2$ for a system with $n_t = 4$ and $K = 4$, using QPSK signaling.

$\mathrm{SNR} = P_{\text{total}}/\mathrm{tr}(\mathbf{R}_n)$. All TH precoding strategies assume that the users are ordered in some sense. Since finding the optimal ordering will involve an exhaustive search over $K!$ possible arrangements, a suboptimal ordering is usually employed. We will choose the suboptimal ordering proposed for MMSE-TH precoder design in [7], and will use it for all methods, including the proposed robust precoders. User ordering is performed at the transmitter using the transmitter's channel estimate $\hat{\mathbf{H}}$.

A. Statistically robust Tomlinson-Harashima precoding

To model the error $\mathbf{E}$ between the actual channel and the estimated channel at the transmitter, $\mathbf{E}$ is generated from a zero-mean Gaussian distribution with $\mathrm{E}\{\mathbf{E}^H\mathbf{E}\} = \sigma_E^2\mathbf{I}$. This model is appropriate for a scenario in which the uplink power is controlled so that the received SNRs on the uplink are equal and independent of the downlink SNR. An example of such a scenario is when the downlink SNR is increased by simply allowing the base station to transmit with greater power. In Fig. 3 we compare the performance of the statistically robust precoder with equal antenna power constraints proposed in Section III-A with that of the zero-forcing design with a total power constraint (ZF2-THP) in [3], [4], for a system with 4 transmit antennas and 3 users. The performance of each method is plotted for values of $\sigma_E^2 = 0.05, 0.1, 0.2$. It can be seen that the effect of noise is dominant at low SNR, while channel uncertainty dominates at high SNR, where the robust precoding approach performs significantly better than the zero-forcing approach. This performance advantage of the proposed robust approach is obtained in spite of the imposition of individual antenna power constraints, as distinct from the total power constraints imposed in [3], [4].

In Fig. 4 we compare the performance of the statistically robust precoder with a total power constraint proposed in Section III-B with that of the ZF1-THP in [5], [6], and the MMSE-THP in [7] for a system with 4 transmit antennas and 4 users. The performance of each method is plotted for values of $\sigma_E^2 = 0.05, 0.1$. The figure shows that in the presence of channel uncertainty, both the zero-forcing (ZF1-THP) and MMSE (MMSE-THP) designs have the same performance limit at high SNR. This is due to the fact that the expression for the precoding matrices in the MMSE method involves the addition of a regularization term whose value is inversely proportional to $P_{\text{total}}/\mathrm{tr}(\mathbf{R}_n)$; see [7]. In Fig. 5 we compare the performance of the statistically robust precoder under a total power constraint with that of the robust zero-forcing method (RZF-THP) introduced in [14], for a system with 4 transmit antennas and 4 users. Each method is plotted for values of $\sigma_E^2 = 0.05, 0.1, 0.2$. Our proposed design shows improvements in the low and moderate SNR regions. Also, for a given target BER, the proposed robust precoder will achieve this target value with less power, as can be seen from next figure.

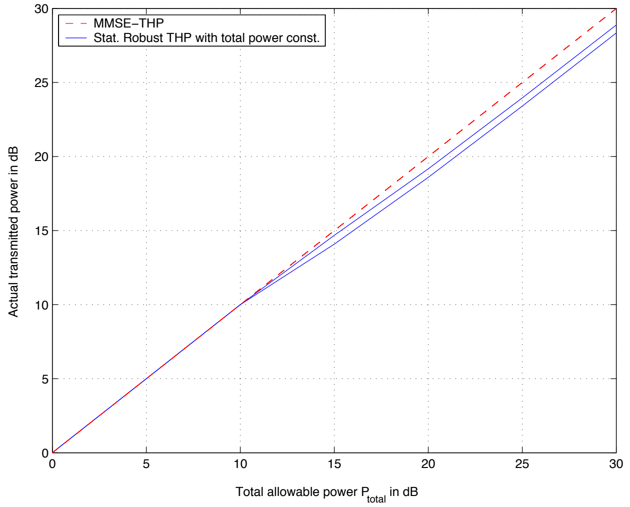

Fig. 6. Average of actual transmitted power versus total allowable power (P_{total}) for the statistically robust precoder with a total power constraint and the MMSE-THP [7] for values of $\sigma_E^2 = 0.05, 0.1$ (from the top downwards) for a system with $n_t = 4$ and $K = 4$, using QPSK signaling.

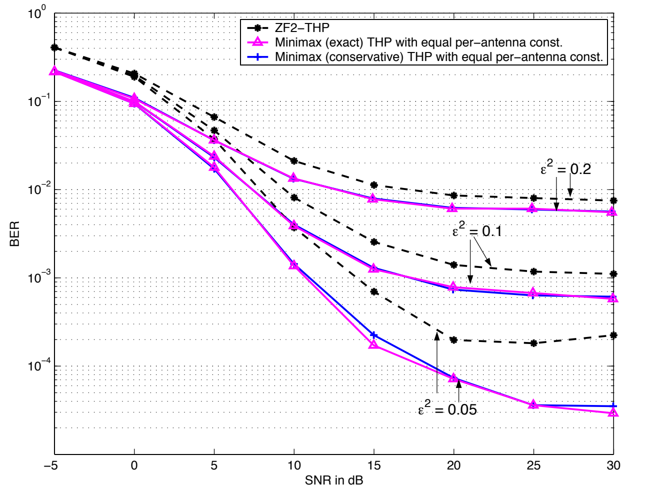

Fig. 7. Comparison between the performance of the exact and conservative minimax precoding methods with equal per-antenna constraints, and the ZF2-THP [3], [4], in the presence of uniformly distributed quantization errors with $\epsilon^2 = 0.05, 0.1, 0.2$ for a system with $n_t = 4$ and $K = 3$, using QPSK signaling.

In Fig. 6 we plot the average of the actual transmitted power against the total allowable power P_{total} for system with $n_t = 4$, $K = 4$ and $\sigma_E^2 = 0.05, 0.1$ for the statistically robust precoder with a total power constraint. It can be seen that the proposed robust precoder does not necessarily use all the allowable power. This is because the interference that results from channel uncertainty increases with the transmitted power and becomes the dominant source of error at high SNRs.

B. Minimax robust precoding

In systems that use feedback to provide the transmitter with a quantized version of the CSI, the information available to the transmitter will include the designed quantization codebooks and the statistics of the error resulting from the use of these codebooks; e.g., $\mathrm{E}\{(\mathbf{h}_k - \hat{\mathbf{h}}_k)(\mathbf{h}_k - \hat{\mathbf{h}}_k)^H\} = \epsilon^2$. Since we assume that each user's channel is independent from the others, the transmitter can model the error matrix $\mathbf{E}$ as being zero mean with independent rows and second order statistics given by $\mathrm{E}\{\mathbf{E}^H\mathbf{E}\} = \epsilon^2\mathbf{I}$. Thus we have $\|\mathrm{E}\{\mathbf{E}^H\mathbf{E}\}\| = \epsilon^2$. To simulate quantization errors, we will generate matrices $\mathbf{E}$ whose elements are drawn from uniform distribution with the above statistics; e.g., [23]. Given that the transmitter will have access to ϵ, and since $\Delta^2 = \|\mathbf{E}^H\mathbf{E}\|$, an appropriate choice for Δ is ϵ.

In Fig. 7 we compare the performance of the exact minimax robust precoder with equal antenna power constraints proposed in Section IV, the conservative minimax precoder with equal antenna power constraints proposed in Section V, and the ZF2-THP design [3], [4] for a system with $n_t = 4$ and $K = 3$. The performance of each method is plotted for values of $\epsilon^2 = 0.05, 0.1, 0.2$. Although the ZF2-THP was designed with a total power constraint, it can be seen from Fig. 7 that the proposed robust methods, with more restrictive per-antenna power constraints, provide better performance in the presence of uncertainty. It can also be seen that performance of the conservative minimax robust TH precoder is very close to that of the computationally more demanding precise minimax one.

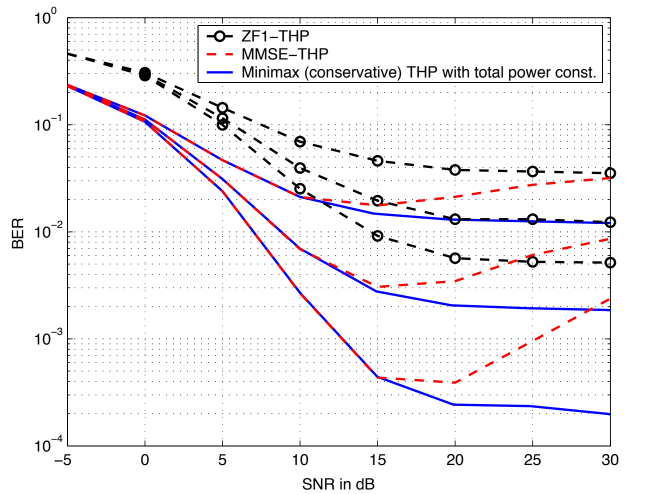

Fig. 8. Comparison between the performance of the conservative minimax precoding method with a total power constraint, and the ZF1-THP [5], [6] and MMSE-THP [7] methods in the presence of uniformly distributed quantization errors with $\epsilon^2 = 0.05, 0.1, 0.2$ for a system with $n_t = 4$ and $K = 4$, using QPSK signaling.

In Fig. 8 we compare the performance of the conservative minimax robust precoder with a total power constraint with that of the ZF1-THP [5], [6] and MMSE-THP [7] methods for a system with $n_t = 4$ and $K = 3$. The performance of each method is plotted for values of $\epsilon^2 = 0.05, 0.1, 0.2$. Fig. 8 reveals qualitatively similar trends to those for the statistically robust design in Fig. 4

VII. Conclusion

We have proposed design techniques for robust Tomlinson-Harashima (TH) precoders for the multiple-input single-output broadcast channel. The techniques were developed for two models of the channel uncertainty that were tailored to particular schemes for acquiring channel state information at the transmitter. We presented convex formulations for robust

TH precoders for each uncertainty model under two possible power constraints: a set of constraints on the power transmitted from each antenna, and a total power constraint. Furthermore, we presented closed-form solutions for robust TH precoders under a total power constraint, demonstrating that robustness can be incorporated into the TH precoder without significantly increasing the computational cost of the design (or implementation). Our simulation studies showed that in the presence of channel uncertainty the proposed robust TH precoders provide significantly better performance than existing designs that assume that accurate channel state information is available to the transmitter, especially at higher SNRs.

APPENDIX I
CLOSED-FORM SOLUTION TO (15)

In this appendix we derive a closed-form solution to (15), the statistically robust precoder design problem with the total power constraint. This problem is a regularized least squares problem with constrained design variables.[2] The Lagrangian associated with this optimization problem is given by:

$$L(\tilde{\mathbf{p}}, \lambda) = (\mathbf{J}\hat{\mathbf{A}}\tilde{\mathbf{p}}-\mathbf{b})^H(\mathbf{J}\hat{\mathbf{A}}\tilde{\mathbf{p}}-\mathbf{b})+\sigma_E^2(\tilde{\mathbf{p}}^H\tilde{\mathbf{p}})+\lambda(\tilde{\mathbf{p}}^H\tilde{\mathbf{p}}-P_{\text{total}}). \quad (32)$$

Since the optimization problem in (15) is convex and strictly feasible, then the KKT conditions are necessary and sufficient conditions for the optimal solution [25]. The KKT conditions associated with (15) are:

$$\nabla_{\tilde{\mathbf{p}}} L(\tilde{\mathbf{p}}, \lambda) = 2(\hat{\mathbf{A}}^H\mathbf{J}^2\hat{\mathbf{A}}\tilde{\mathbf{p}} - \mathbf{A}^H\mathbf{J}\mathbf{b} + (\sigma_E^2 + \lambda)\tilde{\mathbf{p}}) = 0, \quad (33)$$

$$\tilde{\mathbf{p}}^H\tilde{\mathbf{p}} - P_{\text{total}} \le 0, \quad (34)$$

$$\lambda \ge 0, \quad (35)$$

$$\lambda(\tilde{\mathbf{p}}^H\tilde{\mathbf{p}} - P_{\text{total}}) = 0. \quad (36)$$

The complementarity condition in (36) suggests two cases for the optimal solution.

Case 1: $\tilde{\mathbf{p}}^H\tilde{\mathbf{p}} - P_{\text{total}} < 0$, and hence $\lambda = 0$.

In this case the inequality constraint on the transmitted power is strictly satisfied at optimality and the optimal solution is the same as the solution of the unconstrained problem. Using the fact that $\mathbf{J}\mathbf{b} = \mathbf{b}$, the optimal solution can be written as:

$$\tilde{\mathbf{p}} = (\hat{\mathbf{A}}^H\mathbf{J}^2\hat{\mathbf{A}} + \sigma_E^2\mathbf{I})^{-1}\hat{\mathbf{A}}^H\mathbf{b}. \quad (37)$$

Case 2: $\lambda < 0$, and hence $\tilde{\mathbf{p}}^H\tilde{\mathbf{p}} - P_{\text{total}} = 0$.

In this case the power constraint is active at optimality and $\tilde{\mathbf{p}}$ will depend on the optimal value of the Lagrange multiplier. Using (33) and the fact that $\mathbf{J}\mathbf{b} = \mathbf{b}$, the optimal solution can be written as:

$$\tilde{\mathbf{p}} = (\hat{\mathbf{A}}^H\mathbf{J}^2\hat{\mathbf{A}} + (\lambda + \sigma_E^2)\mathbf{I})^{-1}\hat{\mathbf{A}}^H\mathbf{b}, \quad (38)$$

where the value of λ can be found by setting $\tilde{\mathbf{p}}^H\tilde{\mathbf{p}} = P_{\text{total}}$. That is,

$$\begin{aligned} P_{\text{total}} &= \tilde{\mathbf{p}}^H\tilde{\mathbf{p}} = \text{tr}(\tilde{\mathbf{p}}\tilde{\mathbf{p}}^H) \\ &= \text{tr}(\hat{\mathbf{A}}^H\mathbf{b}\mathbf{b}^H\hat{\mathbf{A}}(\hat{\mathbf{A}}^H\mathbf{J}^2\hat{\mathbf{A}} + (\sigma_E^2 + \lambda)\mathbf{I})^{-2})), \end{aligned} \quad (39)$$

[2]A closed-form solution of the corresponding constrained least squares problem without regularization can be found in [34].

where (39) is obtained by cyclic permutation of the order of multiplication of the matrices in the argument of the trace.

Using the SVD of $\mathbf{J}\hat{\mathbf{A}}$, a simpler equation for finding λ can be obtained. To derive that expression, let $\hat{\mathbf{A}} = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^H$ and define $\tilde{\mathbf{b}}_1 = \mathbf{U}_1^H\mathbf{b}$. Equation (39) can then be re-written as:

$$P_{\text{total}} = \text{tr}(\mathbf{\Sigma}_1\tilde{\mathbf{b}}_1\tilde{\mathbf{b}}_1^H\mathbf{\Sigma}_1(\mathbf{\Sigma}_1^2 + (\sigma_E^2 + \lambda)\mathbf{I})^{-2}). \quad (40)$$

Let β_i be the i^{th} diagonal element of $\tilde{\mathbf{b}}\tilde{\mathbf{b}}^H$ and σ_i be the i^{th} diagonal element of $\mathbf{\Sigma}_1$. Then (40) can be written as:

$$f(\lambda) = \sum_{i=1}^{r} \beta_i \frac{\sigma_i^2}{(\sigma_i^2 + (\lambda + \sigma_E^2))^2} - P_{\text{total}} = 0, \quad (41)$$

where r is the rank of $\hat{\mathbf{A}}$.

Now we show that if the root of (41) is positive, then it is unique. Let $\nu = \sigma_E^2 + \lambda$. Then the first two KKT conditions can be written as:

$$(\hat{\mathbf{A}}^H\hat{\mathbf{A}} + \nu\mathbf{I})\tilde{\mathbf{p}} = \hat{\mathbf{A}}^H\mathbf{b}, \quad (42)$$

$$\tilde{\mathbf{p}}^H\tilde{\mathbf{p}} = P_{\text{total}}. \quad (43)$$

For the above system of equations, the following property [34] will be useful in showing the uniqueness of ν if it is a positive solution of equations (42) and (43).

Lemma 2: Let $(\tilde{\mathbf{p}}_1, \nu_1)$ and $(\tilde{\mathbf{p}}_2, \nu_2)$ be two solutions to the pair of equations (42) and (43) such that $P_{\text{total}} > 0$. Then

$$-(\frac{\nu_1 + \nu_2}{2})\|\tilde{\mathbf{p}}_1 - \tilde{\mathbf{p}}_2\|^2 = \|\hat{\mathbf{A}}(\tilde{\mathbf{p}}_1 - \tilde{\mathbf{p}}_2)\|^2. \quad (44)$$

Furthermore, if $\nu_1 \neq \nu_2$ then $\hat{\mathbf{A}}\tilde{\mathbf{p}}_1 \neq \hat{\mathbf{A}}\tilde{\mathbf{p}}_2$.

Proof: See [34]. ■

From the previous lemma, we can see that if ν_1 and ν_2 are two distinct roots, then $-(\nu_1 + \nu_2)$ can only be positive. It cannot be zero since $\hat{\mathbf{A}}(\tilde{\mathbf{p}}_1 - \tilde{\mathbf{p}}_2) \neq 0$. Therefore, there are the two possible cases: both ν_1 and ν_2 are negative, or one of the roots is positive and the second is negative. Hence, if ν is positive then it is unique. Since $\nu = \lambda + \sigma_E^2$, then any positive λ that solves equations (42) and (43) is unique.

APPENDIX II
PROOF OF THEOREM 1

We begin by observing that $f(x)$ is differentiable on $(0, \infty)$ with derivative

$$f'(x) = -2\sum_{i=1}^{r} \beta_i \frac{\sigma_i^2}{(\sigma_i^2 + \sigma_E^2 + x)^3}.$$

Since each β_i is non-negative, for any positive x_1, the derivative $f'(x_1)$ is non-positive and hence $f(x)$ is non-increasing on $(0, \infty)$. The limits of $f(x)$ as x tends to 0 and ∞ are given by

$$\lim_{x\to 0} f(x) = \sum_{i=1}^{r} \beta_i \frac{\sigma_i^2}{(\sigma_i^2 + \sigma_E^2)^2} - P_{\text{total}}, \quad (45)$$

$$\lim_{x\to\infty} f(x) = -P_{\text{total}}, \quad (46)$$

respectively. Since $f(x)$ is a continuous non-increasing function over the interval $(0, \infty)$, a necessary and sufficient condition to have a root in this interval is for the above limits to have opposite signs. For values of P_{total} less than $N = \sum_{i=1}^{r} \beta_i \frac{\sigma_i^2}{(\sigma_i^2+\sigma_E^2)^2}$, the two limits will have different signs and at least one root will exist in the interval $(0, \infty)$. The results of Appendix I show that this root is unique.

REFERENCES

[1] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electron. Lett.*, vol. 7, pp. 138–139, Mar. 1971.

[2] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, no. 4, pp. 774–780, Aug. 1972.

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

[4] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "MIMO precoding for decentralized receivers," in *Proc. IEEE Int. Symp. Inform. Theory*, Lausanne, Switzerland, Jul. 2002, p. 496.

[5] M. Joham, J. Brehmer, A. Voulgarelis, and W. Utschick, "Multiuser spatio-temporal Tomlinson-Harashima precoding for frequency selective vector channels," in *Proc. ITG Wkshp Smart Antennas*, Munich, Mar. 2004, pp. 208–215.

[6] R. Hunger, M. Joham, and W. Utschick, "Extension of linear and nonlinear transmit filters for decentralized receivers," in *Proc. European Wireless Conf.*, Nicosia, April 2005, pp. 40–46.

[7] M. Joham, J. Brehmer, and W. Utschick, "MMSE approaches to multiuser spatio-temporal Tomlinson-Harashima precoding," in *Proc. 5th ITG Int. Conf. Source and Channel Coding*, Erlangen, Jan. 2004, pp. 387–394.

[8] M. Schubert and S. Shi, "MMSE transmit optimization with interference pre-compensation," in *Proc. IEEE Veh. Tech. Conf.*, Stockholm, May 2005, vol. 2, pp. 845–849.

[9] F. C.-H. Fung, W. Yu, and T. J. Lim, "Non-linear multi-user precoding for multi-antenna downlink channels using independent MSE constraints," in *Proc. 22nd Biennial Symposium on Communications*, Kingston, Ontario, May 2004.

[10] W. Shi and R. Wesel, "The effect of mismatch on decision-feedback equalization and Tomlinson-Harashima precoding," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1998, pp. 1743–1747.

[11] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 5045–5059, Nov. 2006.

[12] A. P. Liavas, "Tomlinson-Harashima precoding with partial channel knowledge," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 5–9, Jan. 2005.

[13] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "Tomlinson-Harashima precoding in space-time transmission for low-rate backward channel," in *Proc. Int. Zurich Sem. Broadband Communications, Access, Transmission, Networking*, Zurich, Feb. 2002, pp. 71–76.

[14] R. Hunger, F. Dietrich, M. Joham, and W. Utschick, "Robust transmit zero-forcing filters," in *Proc. ITG Wkshp Smart Antennas*, Munich, Mar. 2004, pp. 130–137.

[15] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. New York: Wiley, 2002.

[16] R. Fischer and C. Windpassinger, "Improved MIMO precoding for decentralized receivers resembling concepts from lattice reduction," in *Proc. IEEE Global Telecommun. Conf.*, San Francisco, Dec. 2003, vol. 4, pp. 1852–1856.

[17] B. Hochwald, C. Peel, and A. Swindlehurst, "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.

[18] M. Botros Shenouda and T. N. Davidson, "Statistically robust transceiver design for broadcast channels with uncertainty," in *Proc. IEEE Canadian Conf. Elec. & Comp. Engineering*, Vancouver, April 2007.

[19] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Linear and nonlinear preequalization/equalization for MIMO systems with long-term channel state information at the transmitter," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 373–378, Mar. 2004.

[20] G. D. Forney Jr. and M. V. Eyuboğlu, "Combined equilization and coding using precoding," *IEEE Commun. Mag.*, vol. 29, pp. 25–34, Dec. 1991.

[21] T. Lan and W. Yu, "Input optimization for multi-antenna broadcast channels with per-antenna power constraints," in *Proc. IEEE Global Telecommun. Conf.*, Dallas, Dec. 2004, pp. 420–424.

[22] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[23] A. Pascual-Iserte, D. Palomar, A. Perez-Neira, and M. Lagunas, "A robust maximin approach for MIMO communications with imperfect channel channel state information based on convex optimization," *IEEE Trans. Signal Processing*, vol. 45, pp. 346–360, Jan. 2006.

[24] D. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proc Parts F and H*, vol. 130, no. 1, pp. 11–16, Feb. 1983.

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

[26] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, no. 1, pp. 193–228, 1998.

[27] F. Alizadeh and D. Goldfarb, "Second-order cone programming," *Math. Program.*, vol. 95, no. 1, pp. 3–51, Jan 2003.

[28] J. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Methods Softw.*, vol. 11, pp. 625–653, 1999.

[29] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Processing*, vol. 53, no. 1, pp. 168–181, Jan. 2005.

[30] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. Philadelphia: SIAM, 2001.

[31] M. Botros Shenouda and T. N. Davidson, "Robust linear precoding for uncertain MISO broadcast channels," in *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing*, Toulouse, May 2006, pp. IV-37–IV-40.

[32] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, "Paramter estimation in the presence of bounded data uncertainties," *SIAM J. Matrix Anal. Appl.*, vol. 19, no. 1, pp. 235–252, Jan. 1998.

[33] M. Botros Shenouda and T. N. Davidson, "Minimax linear precoding for MISO broadcast channels with bounded uncertainty," in *Proc. IEEE Global Telecommun. Conf.*, San Francisco, Nov. 2006.

[34] W. Gander, "Least squares with quadratic constraints," *Numer. Math.*, vol. 36, no. 3, pp. 291–307, Sept 1997.

Michael Botros Shenouda received the B.Sc. (Hons. 1) degree in 2001 and the M.Sc. degree in 2003, both in electrical engineering and both from Cairo University, Egypt. He is currently working toward the Ph.D. degree at the Department of Electrical and Computer Engineering, McMaster University, Canada. His main areas of interest include wireless and MIMO communication, robust and convex optimization, and signal processing algorithms. Mr. Botros Shenouda was awarded an IEEE Student Paper Award at ICASSP 2006, and was a finalist in the student paper competition at ICASSP 2007.

Tim Davidson (M96) received the B.Eng. (Hons. I) degree in Electronic Engineering from the University of Western Australia (UWA), Perth, in 1991 and the D.Phil. degree in Engineering Science from the University of Oxford, U.K., in 1995.

He is currently an Associate Professor in the Department of Electrical and Computer Engineering at McMaster University, Hamilton, Ontario, Canada, where he holds the (Tier II) Canada Research Chair in Communication Systems. His research interests lie in the general areas of communications, signal processing and control. He has held research positions at the Communications Research Laboratory at McMaster University, the Adaptive Signal Processing Laboratory at UWA, and the Australian Telecommunications Research Institute at Curtin University of Technology, Perth, Western Australia.

Dr. Davidson was awarded the 1991 J. A. Wood Memorial Prize [for the most outstanding (UWA) graduand in the pure and applied sciences] and the 1991 Rhodes Scholarship for Western Australia. He is currently serving as an Associate Editor of the *IEEE Transactions on Signal Processing* and the *IEEE Transactions on Circuits and Systems II*, and as a Guest Editor for an upcoming issue of the *IEEE Journal on Selected Topics in Signal Processing*.