Robust Transmit Eigen Beamforming Based on Imperfect Channel State Information

Ayman Abdel-Samad, Timothy N. Davidson, *Member, IEEE*, and Alex B. Gershman, *Fellow, IEEE*

***Abstract*—Transmit beamforming is a powerful technique for enhancing the performance and increasing the throughput of wireless communication systems that employ multiple antennas at the transmitter. A major drawback of most existing transmit beamforming techniques is that they require nearly perfect knowledge of the channel at the transmitter, which is typically not available in practice. Transmitter designs that address the imperfect channel state information (CSI) problem commonly use statistical models for the channel and/or mismatch between the presumed and actual transmitter CSI. Since these approaches are model based, they can suffer from mismodeling. In this paper, a more robust framework is proposed in which no statistical assumptions are made about the CSI mismatch or the channel. The goal is to design a transmitter that has the best performance under the worst-case CSI mismatch. The transmitter designed herein achieves this goal for all CSI mismatches below a certain threshold level. The proposed design combines beamforming along the eigenvectors of the (deterministic) autocorrelation of the channel matrix perceived by the transmitter and power loading across those beams. While the power-loading algorithm resembles conventional water-filling to some degree, it explicitly incorporates robustness to the CSI mismatch, and the water level can be determined in a simple systematic way.**

***Index Terms*—Eigen-beamforming, imperfect channel state information, robust transmit beamforming, water-filling.**

I. Introduction

ANTENNA arrays at the transmitter or receiver (or both) provide effective means to improve the capacity and reliability of wireless links. Traditionally, base stations use antenna arrays to enhance link performance, but with the rapid advancement of technology, mobile units have recently started to share this advantage. There are two main techniques that are used to exploit transmit antenna arrays. The first one is space–time coding, which provides diversity in fading environments [1], [2]. The second is transmit beamforming/precoding, which provides spatially matched transmission or (in multiuser scenarios) mitigates interference [3]–[6]. These approaches represent two extremes regarding the channel knowledge, because space–time coding assumes that no channel state information (CSI) is available at the transmitter, while beamforming requires nearly perfect transmitter CSI.

In practical wireless communications systems, there are several factors that restrict the accuracy of the CSI available at the transmitter. For example, there are always channel estimation errors that are caused by the limited duration of the training sequence. In practice, the CSI imperfections are often dominated by the errors induced by imperfect (quantized, erroneous, and/or outdated) feedback.

Recently, there has been a growing interest in studying the gap between the aforementioned two extremes of CSI, i.e., the cases when only some partial or inaccurate knowledge of the channel at the transmitter side is available [7]–[13]. The designs that address this problem commonly use statistical models for the channel and/or the mismatch between the presumed and actual transmitter CSI. However, these approaches are model based, and therefore, can suffer from mismodeling of the CSI or channel statistics.

In this paper (see also [14] and [15]), we propose an efficient approach to the design of robust transmit beamformers which does not require any statistical model of the CSI mismatch or the channel. Our robust beamformers are designed based on the idea of the worst-case performance optimization, i.e., they offer the best receiver SNR in the worst-case CSI mismatch scenario.[1] We show that as long as the size of the CSI mismatch is below a certain threshold, the solution to the considered worst-case optimization problem reduces to eigen beamforming along the eigenvectors of the (deterministic) autocorrelation of the transmitter's estimate of the spatial channel matrix. Combined with eigen beamforming, robust power loading across the different beams is performed using a spatial water-filling-type technique, where the water level is computed in a simple way.

The rest of the paper is organized as follows. In Section II, the system model is presented for single and multiple receive antenna systems. Transmit beamforming in the perfect CSI case is addressed in Section III, while the imperfect CSI case is considered in Section IV. Simulation results are presented in Section V, and conclusions are drawn in Section VI.

Manuscript received September 27, 2004; revised May 25, 2005. This work was supported in part by the Wolfgang Paul Award Program of the Alexander von Humboldt Foundation and German Ministry of Education and Research; Premier's Research Excellence Award Program of the Ministry of Energy, Science, and Technology (MEST) of Ontario; Natural Sciences and Engineering Research Council (NSERC) of Canada; and Communication and Information Technology Ontario (CITO). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Behrouz Farhang-Boroujeny.

A. Abdel-Samad is with the IC Design Department, MIMOS Berhad, Kuala Lumpur, 57000 Malaysia.

T. N. Davidson is with Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada.

A. B. Gershman is with the Communication Systems Group, Darmstadt University of Technology, Darmstadt, D-64283 Germany.

Digital Object Identifier 10.1109/TSP.2006.872537

[1]Similar criteria have previously been used in several related applications, including the design of robust adaptive beamformers and multiuser detectors [16]–[19]. See also a recent worst-case multi-antenna transmitter design approach of [20], which is closely related to our work and which was proposed independently of [14] and [15].

II. SYSTEM MODEL

Let us consider a single-user wireless communication system with N_t antennas at the transmitter and one antenna at the receiver. The model is similar to that considered in [12] and is summarized here for completeness. At the ith $(i = 1, \ldots, N_t)$ transmit antenna, each transmitted symbol $s(k)$ modulates the code $\mathbf{c}_i = [c_i(0), \ldots, c_i(P_c - 1)]^T$ of length P_c $(P_c \geq N_t)$, where $(\cdot)^T$ denotes transpose, to generate the chip sequence $u_i(n) = \sum_{k=-\infty}^{\infty} s(k) c_i(n - kP_c)$. This chip sequence modulates a chip waveform that satisfies the "root-Nyquist" property with respect to the chip duration T_c. The resulting waveform is then transmitted over the channel. The baseband-equivalent channel is assumed to be flat in frequency and sufficiently slowly varying for the complex channel coefficients $\{h_i\}_{i=1}^{N_t}$ between each transmit antenna and the receiver to be considered constant over the symbol duration $P_c T_c$. The received signal is modeled as being contaminated by zero-mean additive white Gaussian noise (AWGN), and after chip-matched filtering and (synchronized) sampling at the chip rate, the received sequence can be written as

$$x(n) = \sum_{k=\infty}^{\infty} \sum_{i=1}^{N_t} h_i c_i(n - kP_c) s(k) + w(n) \tag{1}$$

where $w(n)$ is a zero-mean white circular complex Gaussian sequence of variance N_0. To write (1) in matrix form, we define the $P_c \times 1$ vectors $\mathbf{x}(k) = [x(kP_c), \ldots, x(kP_c + P_c - 1)]^T$ and $\mathbf{w}(k) = [w(kP_c), \ldots, w(kP_c + P_c - 1)]^T$, as well as the $N_t \times 1$ channel vector $\mathbf{h} = [h_1, \ldots, h_{N_t}]^T$ and the $P_c \times N_t$ beamforming matrix $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_{N_t}]^T$. Equation (1) can then be written as $\mathbf{x}(k) = \mathbf{C}\mathbf{h}s(k) + \mathbf{w}(k)$. As we will focus on symbol-by-symbol detection, we omit the symbol index k and simply deal with the model

$$\mathbf{x} = \mathbf{C}\mathbf{h}s + \mathbf{w}. \tag{2}$$

The matrix $\mathbf{C}$ in (2) can be alternatively viewed as a precoder or transmit beamformer [12].

The receiver is assumed to have perfect knowledge of $\mathbf{C}$ and $\mathbf{h}$. Thus, maximum ratio combining (MRC) can be used. The output of the MRC is given by

$$y = \mathbf{h}^{\mathcal{H}} \mathbf{C}^{\mathcal{H}} \mathbf{x} = \mathbf{h}^{\mathcal{H}} \mathbf{C}^{\mathcal{H}} \mathbf{C}\mathbf{h}s + \mathbf{h}^{\mathcal{H}} \mathbf{C}^{\mathcal{H}} \mathbf{w} \tag{3}$$

where $(\cdot)^{\mathcal{H}}$ denotes Hermitian transpose. The signal-to-noise ratio (SNR) at the MRC receiver output can be expressed as

$$\text{SNR} = \frac{E\left\{|\mathbf{h}^{\mathcal{H}} \mathbf{C}^{\mathcal{H}} \mathbf{C}\mathbf{h}s|^2\right\}}{E\left\{|\mathbf{h}^{\mathcal{H}} \mathbf{C}^{\mathcal{H}} \mathbf{w}\mathbf{w}^{\mathcal{H}} \mathbf{C}\mathbf{h}|\right\}} = \frac{E_s}{N_0} \mathbf{h}^{\mathcal{H}} \mathbf{C}^{\mathcal{H}} \mathbf{C}\mathbf{h} \tag{4}$$

where $E_s = E\{|s|^2\}$ is the average energy of the signal constellation and, as defined earlier, $N_0/2$ is the noise variance per dimension. In this paper, we will focus on optimizing the SNR under various scenarios and, for the sake of simplicity and without any loss of generality, we will drop the constant factor E_s/N_0 from the SNR expression.

To extend the model to a system with N_r receive antennas, we let $\mathbf{h}_j$ be the channel vector for each receive antenna, and arrange the $\mathbf{h}_j$'s into a matrix $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_{N_r}]$. Similar to the single-receive-antenna case, the received signal at the jth antenna is $\mathbf{x}_j = \mathbf{C}\mathbf{h}_j s + \mathbf{w}_j$, and it follows that the MRC output is given by

$$y = \sum_{j=1}^{N_r} \left[\mathbf{h}_j^{\mathcal{H}} \mathbf{C}^{\mathcal{H}} \mathbf{C}\mathbf{h}_j s + \mathbf{h}_j^{\mathcal{H}} \mathbf{C}^{\mathcal{H}} \mathbf{w}_j\right]. \tag{5}$$

The receiver SNR at the output of the MRC is the sum of individual antenna SNRs, i.e.,

$$\text{SNR} = \sum_{j=1}^{N_r} \mathbf{h}_j^{\mathcal{H}} \mathbf{C}^{\mathcal{H}} \mathbf{C}\mathbf{h}_j = \operatorname{tr}\left\{\mathbf{H}^{\mathcal{H}} \mathbf{C}^{\mathcal{H}} \mathbf{C}\mathbf{H}\right\}. \tag{6}$$

III. PERFECT CSI CASE

We will first design the optimal matrix $\mathbf{C}$ in the simpler case when the transmitter has perfect CSI. Our objective is to obtain $\mathbf{C}$ that maximizes the receiver SNR while maintaining the average power constraint across the transmit antennas. Although the solution to this problem is well known (e.g., [21] and [22] and references therein), we will use this section to introduce some of the tools required to deal with the more difficult problem of imperfect CSI in the next section. Our key notational convention will be to represent the eigen decompositions of the positive semidefinite matrices $\mathbf{C}^{\mathcal{H}}\mathbf{C}$ and $\mathbf{H}\mathbf{H}^{\mathcal{H}}$ by

$$\mathbf{C}^{\mathcal{H}}\mathbf{C} = \mathbf{U}_c \mathbf{D}_c \mathbf{U}_c^{\mathcal{H}} \tag{7}$$

$$\mathbf{H}\mathbf{H}^{\mathcal{H}} = \mathbf{U}_h \mathbf{D}_h \mathbf{U}_h^{\mathcal{H}} \tag{8}$$

where $\mathbf{U}_c = [\mathbf{u}_{c_1}, \ldots, \mathbf{u}_{c_{N_t}}]$ and $\mathbf{U}_h = [\mathbf{u}_{h_1}, \ldots, \mathbf{u}_{h_{N_t}}]$ are unitary. Here d_{c_i} and d_{h_i} denote the ith eigenvalue of $\mathbf{C}^{\mathcal{H}}\mathbf{C}$ and $\mathbf{H}\mathbf{H}^{\mathcal{H}}$, respectively, so that

$$\mathbf{D}_c = \operatorname{diag}\{\mathbf{d}_c\}, \quad \mathbf{D}_h = \operatorname{diag}\{\mathbf{d}_h\} \tag{9}$$

where $\mathbf{d}_c = [d_{c_1}, \ldots, d_{c_{N_t}}]^T$ and $\mathbf{d}_h = [d_{h_1}, \ldots, d_{h_{N_t}}]^T$. Without loss of generality, we can assume that $d_{c_1} \geq d_{c_2} \geq \cdots \geq d_{c_{N_t}}$ and $d_{h_1} \geq d_{h_2} \geq \cdots \geq d_{h_{N_t}}$.

The problem of maximizing the receiver SNR while satisfying a unit average power constraint across the transmit antennas can be written as

$$\max_{\mathbf{C}} \operatorname{tr}\left\{\mathbf{H}^{\mathcal{H}} \mathbf{C}^{\mathcal{H}} \mathbf{C}\mathbf{H}\right\} \quad \text{s.t.} \quad \operatorname{tr}\{\mathbf{C}^{\mathcal{H}}\mathbf{C}\} \leq 1. \tag{10}$$

Since both the received SNR and the power constraint depend on $\mathbf{C}$ through $\mathbf{C}^{\mathcal{H}}\mathbf{C}$, it suffices to optimize with respect to $\mathbf{U}_c$ and $\mathbf{D}_c$. In particular, the objective function can be rewritten as

$$\operatorname{tr}\left\{\mathbf{H}^{\mathcal{H}} \mathbf{C}^{\mathcal{H}} \mathbf{C}\mathbf{H}\right\} = \operatorname{tr}\left\{\mathbf{D}_h \mathbf{U}_h^{\mathcal{H}} \mathbf{U}_c \mathbf{D}_c \mathbf{U}_c^{\mathcal{H}} \mathbf{U}_h\right\}. \tag{11}$$

To simplify (10), we define the unitary matrix $\mathbf{U}_{hc} = \mathbf{U}_h^{\mathcal{H}} \mathbf{U}_c$, the positive semidefinite matrix

$$\bar{\mathbf{C}} = \mathbf{U}_{hc} \mathbf{D}_c \mathbf{U}_{hc}^{\mathcal{H}} \tag{12}$$

and $\bar{\mathbf{c}}$ and $\mathbf{d}_{\bar{c}}$ to be the vectors of diagonal elements and eigenvalues of $\bar{\mathbf{C}}$, respectively (note that $\mathbf{d}_{\bar{c}} = \mathbf{d}_c$). Accordingly, the problem in (10) is equivalent to

$$\max_{\mathbf{U}_c, \mathbf{D}_c} \sum_{i=1}^{N_t} d_{h_i} \bar{c}_i \ \text{ s.t.} \quad \text{tr}\{\mathbf{D}_c\} \leq 1. \tag{13}$$

Once the optimal $\mathbf{U}_c$ and $\mathbf{D}_c$ are obtained, the optimal matrix $\mathbf{C}$ can be expressed as

$$\mathbf{C} = \mathbf{\Phi} \mathbf{D}_c^{1/2} \mathbf{U}_c^{\mathcal{H}} \tag{14}$$

where the columns of $\mathbf{\Phi}$ are orthonormal. Note that the degrees of freedom offered by this matrix can be used to increase the symbol rate through the use of orthogonal space–time block codes [12], [13].

We can always maximize a function by first maximizing over some of the variables and then maximizing over the remaining ones [23]. Therefore, we can solve (13) by first finding the $\mathbf{U}_c$ that maximizes the objective function in (13) for a given $\mathbf{D}_c$, and then finding the optimal $\mathbf{D}_c$. The fact (demonstrated below) that the optimal $\mathbf{U}_c$ is independent of $\mathbf{D}_c$ leads to a particularly simple solution.

Recall that the eigenvectors of $\mathbf{C}^{\mathcal{H}}\mathbf{C}$ and $\mathbf{H}\mathbf{H}^{\mathcal{H}}$ are arranged so that their corresponding eigenvalues are in nonincreasing order. However, the diagonal elements of $\bar{\mathbf{C}}$ will not necessarily be in nonincreasing order. The following lemma provides the desirable ordering.

Lemma 1: For any two given positive real vectors $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_n]^T$ and $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_n]^T$, the permutation π^* that maximizes the sum $\sum_{i=1}^{n} \alpha_{\pi(i)} \beta_i$ is such that $\alpha_{\pi(i)}$ and β_i are in the same order. That is, $\forall\, i, j \in \{1, 2 \ldots, n\}$ if $\alpha_{\pi^*(i)} > \alpha_{\pi^*(j)}$ then $\beta_i > \beta_j$.

Proof: The proof follows the same line as the proof of Theorem 2 in [9] which involves minimizing the sum of α_i/β_i. The details are provided in Appendix I. ■

Using this lemma, we conclude that

$$\sum_{i=1}^{N_t} d_{h_i} \bar{c}_i \leq \sum_{i=1}^{N_t} d_{h_i} \bar{c}_{\pi(i)} \tag{15}$$

where $\pi(i)$ is a permutation such that $\bar{c}_{\pi(i)}$ is arranged in a nonincreasing order. In order to exploit the relation between the eigenvalues d_{c_i} of $\bar{\mathbf{C}}$, and its diagonal elements $\bar{c}_i$, we recall that the diagonal elements of a Hermitian matrix majorize the eigenvalues, where a real vector $\mathbf{a} \in \mathbb{R}^n$ is said to majorize another vector $\mathbf{d} \in \mathbb{R}^n$ if and only if the sum of the k smallest entries of $\mathbf{a}$ is greater than or equal to the sum of the k smallest entries of $\mathbf{d}$ for $k = 1, \ldots, n$, and the sums of all entries of $\mathbf{a}$ and $\mathbf{d}$ are equal. More formally, we have Lemma 2.

Lemma 2 (e.g., [24, Th. 4.3.26]): For an $n \times n$ Hermitian matrix $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathcal{H}}$, the vector of diagonal entries $[a_{1,1} \ldots, a_{n,n}]^T$ majorizes the vector of (nonincreasingly ordered) eigenvalues $[d_{1,1}, \ldots, d_{n,n}]^T$.

Applying this lemma to $\bar{\mathbf{C}}$, we have that

$$\sum_{i=1}^{k} \bar{c}_{\pi(i)} \leq \sum_{i=1}^{k} d_{c_i}. \tag{16}$$

The final lemma required for the solution of (13) is as follows.

Lemma 3: If positive real vectors $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_n]^T$, $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_n]^T$, and $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_n]^T$ have their components arranged in nonincreasing order, and if $\boldsymbol{\alpha}$ majorizes $\boldsymbol{\beta}$ then

$$\sum_{i=1}^{n} \alpha_i \gamma_i \leq \sum_{i=1}^{n} \beta_i \gamma_i. \tag{17}$$

Proof: The proof follows the same line as the proof of [9, Th. 3], which involves dividing α_i and β_i by γ_i rather than multiplying. In this proof, Lemma 2 is used. The framework of the proof is outlined in Appendix II. ■

Applying Lemma 3 to (15), we have that

$$\sum_{i=1}^{N_t} d_{h_i} \bar{c}_i \leq \sum_{i=1}^{N_t} d_{h_i} \bar{c}_{\pi(i)} \leq \sum_{i=1}^{N_t} d_{h_i} d_{c_i} \tag{18}$$

with obvious equality when $\mathbf{U}_{hc} = \mathbf{I}$, i.e., when $\mathbf{U}_c = \mathbf{U}_h$. Therefore, for *any* $\mathbf{D}_c$, the optimal $\mathbf{U}_c$ in (13) is $\mathbf{U}_h$.

Having found the optimal $\mathbf{U}_c$, the optimization problem in (13) reduces to

$$\max_{\sum_{i=1}^{N_t} d_{c_i} \leq 1} \sum_{i=1}^{N_t} d_{h_i} d_{c_i}. \tag{19}$$

This is a linear programming problem and the maximum is achieved when $d_{c_1} = 1$ and $d_{c_i} = 0$ $(i = 2, \ldots N_t)$. Thus, the optimal beamformer in the sense of maximizing the receiver SNR with perfect channel knowledge at the transmitter is a one-directional eigen beamformer along the principal eigenvector of the (deterministic) autocorrelation of the spatial channel matrix [12], [21], [22]. That is, the optimal $\mathbf{C}$ is given by

$$\mathbf{C} = \mathbf{\Phi} \mathbf{G}^{\mathcal{H}} \tag{20}$$

where $\mathbf{G} = [\mathbf{u}_{h_1}, \mathbf{0}, \ldots, \mathbf{0}]$ and, as in (14), the columns of $\mathbf{\Phi}$ are orthonormal.

IV. Imperfect CSI Case

We now consider the case when the transmitter does not have exact CSI, but has an estimate $\hat{\mathbf{H}}$ of the channel matrix $\mathbf{H}$. Thus, the CSI error is given by $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_{N_r}] = \mathbf{H} - \hat{\mathbf{H}}$. We want to design the matrix $\mathbf{C}$ that maximizes the receiver SNR for the worst-case error under the condition that the error matrix is bounded in its norm by some constant ϵ, i.e., $\|\mathbf{E}\|_F \leq \epsilon$, where $\| \cdot \|_F$ denotes the Frobenius matrix norm. The receiver SNR can be written as

$$\begin{aligned}\text{SNR} &= \sum_{j=1}^{N_r} \left(\hat{\mathbf{h}}_j + \mathbf{e}_j\right)^{\mathcal{H}} \mathbf{C}^{\mathcal{H}} \mathbf{C} \left(\hat{\mathbf{h}}_j + \mathbf{e}_j\right) \\ &= \text{tr}\left\{ \left(\hat{\mathbf{H}} + \mathbf{E}\right)^{\mathcal{H}} \mathbf{C}^{\mathcal{H}} \mathbf{C} \left(\hat{\mathbf{H}} + \mathbf{E}\right) \right\}\end{aligned} \tag{21}$$

where $\hat{\mathbf{H}}$ is known at the transmitter, but $\mathbf{E}$ is not. For a given presumed channel matrix $\hat{\mathbf{H}}$ and beamforming matrix $\mathbf{C}$, the worst-case SNR at the receiver is given by

$$\text{SNR}_w = \min_{\|\mathbf{E}\|_F \leq \epsilon} \sum_{j=1}^{N_r} \left(\hat{\mathbf{h}}_j + \mathbf{e}_j\right)^{\mathcal{H}} \mathbf{C}^{\mathcal{H}} \mathbf{C} \left(\hat{\mathbf{h}}_j + \mathbf{e}_j\right). \tag{22}$$

The output of this minimization will depend only on $\hat{\mathbf{H}}$, $\mathbf{C}$, and ϵ. We would like to design $\mathbf{C}$ in order to maximize this worst-case performance. Thus, our problem can be written as

$$\max_{\mathrm{tr}\{\mathbf{C}^{\mathcal{H}}\mathbf{C}\}\leq 1} \min_{\|\mathbf{E}\|_F\leq\epsilon} \sum_{j=1}^{N_r} \left(\hat{\mathbf{h}}_j+\mathbf{e}_j\right)^{\mathcal{H}} \mathbf{C}^{\mathcal{H}}\mathbf{C}\left(\hat{\mathbf{h}}_j+\mathbf{e}_j\right). \quad (23)$$

A. Worst-Case Mismatch

Our first step in finding the solution of (23) will be to solve the inner minimization problem to determine the worst-case mismatch and the worst-case SNR. Using the Lagrange multiplier method (e.g., [25]), the Lagrangian $\mathcal{L}$ can be written as

$$\mathcal{L}=\sum_{j=1}^{N_r}\left(\hat{\mathbf{h}}_j+\mathbf{e}_j\right)^{\mathcal{H}}\mathbf{C}^{\mathcal{H}}\mathbf{C}\left(\hat{\mathbf{h}}_j+\mathbf{e}_j\right)+\lambda\left(\|\mathbf{E}\|_F^2-\epsilon^2\right) \quad (24)$$

where λ is the Lagrange multiplier. Differentiating $\mathcal{L}$ with respect to $\mathbf{e}_j^*$ and setting the result to zero gives

$$\mathbf{C}\mathbf{C}^{\mathcal{H}}\hat{\mathbf{h}}_j+\mathbf{C}\mathbf{C}^{\mathcal{H}}\mathbf{e}_j+\lambda\mathbf{e}_j=\mathbf{0} \quad (25)$$

where $(\cdot)^*$ denotes complex conjugate. From the latter equation, we have that the jth column of the worst-case mismatch matrix $\mathbf{E}_w$ is given by

$$\mathbf{e}_{w_j}=-(\mathbf{C}^{\mathcal{H}}\mathbf{C}+\lambda\mathbf{I})^{-1}\mathbf{C}^{\mathcal{H}}\mathbf{C}\hat{\mathbf{h}}_j. \quad (26)$$

Using (7), we can rewrite (26) as

$$\mathbf{e}_{w_j}=-\mathbf{U}_c\,\mathrm{diag}\left\{\frac{d_{c_i}}{d_{c_i}+\lambda}\right\}\mathbf{U}_c^{\mathcal{H}}\hat{\mathbf{h}}_j \quad (27)$$

and the worst-case mismatch matrix $\mathbf{E}_w$ can be expressed as

$$\mathbf{E}_w=-\mathbf{U}_c\,\mathrm{diag}\left\{\frac{d_{c_i}}{d_{c_i}+\lambda}\right\}\mathbf{U}_c^{\mathcal{H}}\hat{\mathbf{H}} \quad (28)$$

where the value of the Lagrange multiplier λ can be determined by substituting (28) in $\|\mathbf{E}\|_F^2=\epsilon^2$. That is, λ can be obtained by solving

$$\epsilon^2=\mathrm{tr}\left\{\hat{\mathbf{H}}^{\mathcal{H}}\mathbf{U}_c\mathrm{diag}\left\{\left(\frac{d_{c_i}}{d_{c_i}+\lambda}\right)^2\right\}\mathbf{U}_c^{\mathcal{H}}\hat{\mathbf{H}}\right\}. \quad (29)$$

It is clear from this equation that, for a specific $\hat{\mathbf{H}}$ and $\mathbf{C}$, the value of λ is a monotonically decreasing function of ϵ. The two limiting cases are $\lambda=0$, which corresponds to $\epsilon=\|\hat{\mathbf{H}}\|_F$ (i.e., uncertainty is of the same size as the channel estimate itself), and $\lambda=\infty$, which corresponds to $\epsilon=0$ (i.e., perfect CSI). It is worth noting that $\mathbf{e}_{w_j}$ is not necessarily in the opposite direction of the presumed channel vector $\hat{\mathbf{h}}_j$, as evident from (27). This is because the diagonal elements $(d_{c_i}/(d_{c_i}+\lambda))$ are generally not equal, except for the cases when the d_{c_i}'s are equal or when $\lambda=0$.

By substituting (28) into (21), the worst-case SNR can be expressed as

$$\mathrm{SNR}_w=\mathrm{tr}\left\{\mathbf{D}_{\hat{h}}\mathbf{U}_{\hat{h}c}\mathrm{diag}\left\{d_{c_i}\left(\frac{\lambda}{d_{c_i}+\lambda}\right)^2\right\}\mathbf{U}_{\hat{h}c}^{\mathcal{H}}\right\} \quad (30)$$

where $\mathbf{U}_{\hat{h}c}=\mathbf{U}_{\hat{h}}^{\mathcal{H}}\mathbf{U}_c$, with $\mathbf{D}_{\hat{h}}$ and $\mathbf{U}_{\hat{h}}$ being the eigenvalue and eigenvector matrices of $\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathcal{H}}$, respectively. For given $\hat{\mathbf{H}}$, $\mathbf{C}$, and ϵ, the value of λ can be obtained by simplifying (29) to

$$\epsilon^2=\mathrm{tr}\left\{\mathbf{D}_{\hat{h}}\mathbf{U}_{\hat{h}c}\mathrm{diag}\left\{\left(\frac{d_{c_i}}{d_{c_i}+\lambda}\right)^2\right\}\mathbf{U}_{\hat{h}c}^{\mathcal{H}}\right\} \quad (31)$$

$$=\sum_{i=1}^{N_t}\bar{h}_{i,i}\left(\frac{d_{c_i}}{d_{c_i}+\lambda}\right)^2 \quad (32)$$

where $\bar{h}_{i,i}$ are the diagonal elements of the matrix $\bar{\mathbf{H}}=\mathbf{U}_{\hat{h}c}^{\mathcal{H}}\mathbf{D}_{\hat{h}}\mathbf{U}_{\hat{h}c}$. Using (32), we have that

$$\epsilon^2\prod_{i=1}^{N_t}(\lambda+d_{c_i})^2=\sum_{i=1}^{N_t}\left[\bar{h}_{i,i}d_{c_i}^2\prod_{\substack{j=1\\ j\neq i}}^{N_t}\left(\lambda+d_{c_j}\right)^2\right] \quad (33)$$

and hence the value of λ that satisfies (29) is the largest positive root of the polynomial

$$\prod_{i=1}^{N_t}\left(\lambda^2+2d_{c_i}\lambda+d_{c_i}^2\right) -\sum_{i=1}^{N_t}\left[\frac{\bar{h}_{i,i}d_{c_i}^2}{\epsilon^2}\prod_{\substack{j=1\\ j\neq i}}^{N_t}\left(\lambda^2+2d_{c_j}\lambda+d_{c_j}^2\right)\right]=0. \quad (34)$$

The coefficients of this polynomial can be easily computed using simple convolutions of the sequences $\{1,d_{c_i}\}$ along with a limited number of multiplications and additions, and the desired root can be computed using standard methods.

B. Robust Power Loading for Eigen Beamformers

Having found the worst-case mismatch and the corresponding worst-case SNR, the remaining task is to find the transmitter matrix $\mathbf{C}$ that maximizes the worst-case SNR. As was the case in the perfect CSI scenario, the optimization of $\mathbf{C}$ can be decomposed into finding the optimal beam directions $\mathbf{U}_c$ and obtaining the optimal power loading $\mathbf{D}_c$. However, unlike the perfect CSI scenario, the structure of the optimal $\mathbf{U}_c$ is dependent on the choice of $\mathbf{D}_c$ when ϵ cannot be regarded as being small (see Section IV-C). This means that we cannot optimize $\mathbf{U}_c$ and $\mathbf{D}_c$ separately. That said, one might suspect that for small values of epsilon, the optimal choice for the beam directions $\mathbf{U}_c$ is still the eigenmatrix $\mathbf{U}_{\hat{h}}$. That is, one might suspect that the eigen beamformer remains optimal. Indeed, this is the case when the uncertainty in the CSI is modeled using a standard statistical model [12], [21], [22], rather than the deterministically bounded model in (23). In this subsection, we will make the choice $\mathbf{U}_c=\mathbf{U}_{\hat{h}}$ and will derive a closed-form expression for the robust power loading of these eigen-beams. In Section IV-C we will validate this choice for $\mathbf{U}_c$ by providing sufficient conditions on the size of the uncertainty under which the proposed transmitter provides the maximum worst-case SNR, i.e., sufficient conditions under which the proposed transmitter solves (23).

Eigen beamforming corresponds to making the choice $\mathbf{U}_c = \mathbf{U}_{\hat{h}}$, and for that choice, the problem in (23) of maximizing the worst-case SNR reduces to

$$\max_{\mathbf{d}_c} \quad \sum_{i=1}^{N_t} d_{\hat{h}_i} \left[d_{c_i} \left(\frac{\lambda}{d_{c_i} + \lambda} \right)^2 \right] \tag{35a}$$

$$\text{subject to} \quad \sum_{i=1}^{N_t} d_{c_i} \leq 1, \text{ and } d_{c_i} \geq 0, \quad \forall i \tag{35b}$$

where, according to the derivation of the worst-case SNR (cf. (32)), λ has to satisfy

$$\sum_{i=1}^{N_t} d_{\hat{h}_i} \left(\frac{d_{c_i}}{d_{c_i} + \lambda} \right)^2 = \epsilon^2. \tag{36}$$

We will solve (35) by finding an analytic solution to the Karush–Kuhn–Tucker (KKT) optimality conditions (see, for example, [23] and [25] for discussion of these conditions). In order to use the KKT conditions in their standard form, we define

$$f = -\sum_{i=1}^{N_t} d_{\hat{h}_i} d_{c_i} \left(\frac{\lambda}{\lambda + d_{c_i}} \right)^2 \tag{37}$$

and modify the objective in (35a) to $\min\ f$. The Lagrangian of that problem can be written as

$$\mathcal{L} = f + \nu \left(\sum_{j=1}^{N_t} d_{c_j} - 1 \right) - \sum_{j=1}^{N_t} \mu_j d_{c_j} \tag{38}$$

where ν and the μ_j's are the Lagrange multipliers. In Appendix III, we show that

$$\frac{\partial f}{\partial d_{c_i}} = -d_{\hat{h}_i} \left(\frac{\lambda}{\lambda + d_{c_i}} \right)^2 \tag{39}$$

and, hence

$$\frac{\partial \mathcal{L}}{\partial d_{c_i}} = \nu - \mu_i - d_{\hat{h}_i} \left(\frac{\lambda}{\lambda + d_{c_i}} \right)^2. \tag{40}$$

Therefore, the complete KKT conditions can be written as

$$d_{c_i} \geq 0, \quad \sum_{j=1}^{N_t} d_{c_j} = 1, \quad \mu_i \geq 0, \quad \mu_i d_{c_i} = 0, \tag{41a}$$

$$\nu = \mu_i + d_{\hat{h}_i} \left(\frac{\lambda}{\lambda + d_{c_i}} \right)^2, \quad i = 1, \ldots, N_t. \tag{41b}$$

From (41b), we observe that μ_i acts as a slack variable. By eliminating it, we can rewrite (41) as

$$d_{c_i} \geq 0, \quad \sum_{j=1}^{N_t} d_{c_j} = 1 \tag{42a}$$

$$\left[\nu - d_{\hat{h}_i} \left(\frac{\lambda}{\lambda + d_{c_i}} \right)^2 \right] d_{c_i} = 0 \tag{42b}$$

$$\nu \geq d_{\hat{h}_i} \left(\frac{\lambda}{\lambda + d_{c_i}} \right)^2, \qquad i = 1, \ldots, N_t. \tag{42c}$$

If $\nu < d_{\hat{h}_i}$, (42c) implies that $d_{c_i} > 0$ which, according to (42b), means that $\nu = d_{\hat{h}_i}(\lambda/(\lambda + d_{c_i}))^2$. Solving for d_{c_i} we have $d_{c_i} = \lambda\left(\sqrt{d_{\hat{h}_i}/\nu} - 1\right)$. On the other hand, if $\nu \geq d_{\hat{h}_i}$, then we cannot have $d_{c_i} > 0$ because this would mean that $\nu \geq d_{\hat{h}_i} > d_{\hat{h}_i}(\lambda/(\lambda + d_{c_i}))^2$, which violates the complementary slackness condition in (42b). Therefore, if $\nu \geq d_{\hat{h}_i}$, then $d_{c_i} = 0$. In other words

$$d_{c_i} = \begin{cases} \lambda \left(\sqrt{\frac{d_{\hat{h}_i}}{\nu}} - 1 \right), & \nu < d_{\hat{h}_i} \\ 0, & \nu \geq d_{\hat{h}_i} \end{cases} \tag{43a}$$

$$= \lambda \max \left\{ \left(\sqrt{\frac{d_{\hat{h}_i}}{\nu}} - 1 \right), 0 \right\}. \tag{43b}$$

Another way of expressing d_{c_i} would be to write it as

$$d_{c_i} = \frac{\lambda}{\sqrt{\nu}} \max \left\{ \left[\left(\|\hat{\mathbf{H}}\|_F - \sqrt{\nu} \right) - \left(\|\hat{\mathbf{H}}\|_F - \sqrt{d_{\hat{h}_i}} \right) \right], 0 \right\}. \tag{44}$$

Equation (44) describes a water-filling-type solution. The water-filling procedure is illustrated in Fig. 1. The channel is flooded to a water-level of $(\|\hat{\mathbf{H}}\|_F - \sqrt{\nu})$, and power, d_{c_i}, is loaded in the direction of eigenvector $\mathbf{u}_{\hat{h}_i}$ if the water level is above the threshold $\left(\|\hat{\mathbf{H}}\|_F - \sqrt{d_{\hat{h}_i}}\right)$. The multiplying constant $\lambda/\sqrt{\nu}$ is chosen so that the solution satisfies the condition $\sum_{j=1}^{N_t} d_{c_j} = 1$.

In order to complete the solution to (35), we need to determine the values of λ and ν. To that end, we observe that the power constraint must hold with equality (i.e., $\sum_{j=1}^{N_t} d_{c_j} = 1$), and (36) must also hold. Therefore, to determine λ and ν, we must solve the two nonlinear simultaneous equations given by

$$\sum_{j=1}^{N_t} \max \left\{ \left(\sqrt{\frac{d_{\hat{h}_i}}{\nu}} - 1 \right), 0 \right\} = \frac{1}{\lambda} \tag{45}$$

and the substitution of (43) into (36). We will obtain ν by inserting (43) into(36) and will then substitute that value into (45) to determine λ. The details are as follows.

Since the eigenvalues of $\hat{\mathbf{H}}$ are arranged in nonincreasing order, d_{c_i}'s in (43) are also arranged in nonincreasing order. Therefore, there exists an $n \in \{1, 2, \ldots, N_t\}$ such that for the optimal d_{c_i}'s in (43), $d_{c_n} > 0$ and $d_{c_{n+1}} = 0$, where we define $d_{c_{N_t+1}} = 0$ (note that this means that $d_{\hat{h}_{n+1}} \leq \nu < d_{\hat{h}_n}$). Therefore, for the d_{c_i}'s in (43), equation (36) can be rewritten as

$$\begin{aligned} \epsilon^2 &= \sum_{i=1}^{N_t} d_{\hat{h}_i} \left(\frac{d_{c_i}}{\lambda + d_{c_i}} \right)^2 \\ &= \sum_{i=1}^{n} d_{\hat{h}_i} \left(\frac{\lambda \left(\sqrt{\frac{d_{\hat{h}_i}}{\nu}} - 1 \right)}{\lambda + \lambda \left(\sqrt{\frac{d_{\hat{h}_i}}{\nu}} - 1 \right)} \right)^2 \\ &= \left(\sum_{i=1}^{n} d_{\hat{h}_i} \right) - 2\sqrt{\nu} \left(\sum_{i=1}^{n} \sqrt{d_{\hat{h}_i}} \right) + n\,\nu. \end{aligned} \tag{46}$$

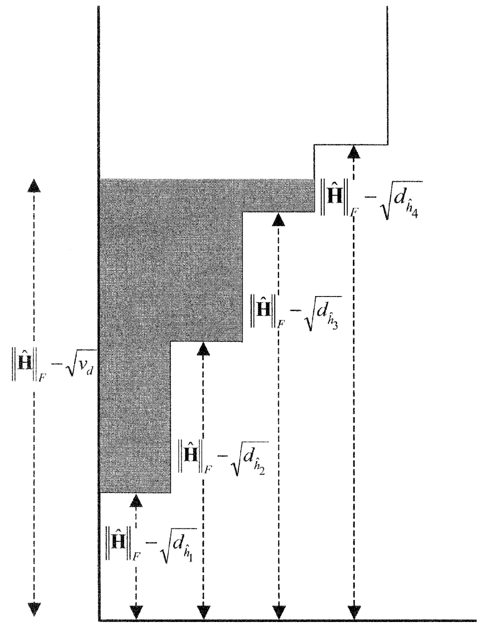


Fig. 1. Illustration of water-filling.

Fortunately, (46) is independent of λ and represents a second-order equation with respect to $\sqrt{\nu}$. As the largest of the two positive roots would result in $\mathbf{d}_c = \mathbf{0}$, we take the smallest root and obtain

$$\nu(n) = \left[\frac{\tilde{d}_{\hat{h}_n} - \sqrt{\tilde{d}^2_{\hat{h}_n} + n\left(\epsilon^2 - \bar{d}_{\hat{h}_n}\right)}}{n}\right]^2 \tag{47}$$

where

$$\tilde{d}_{\hat{h}_n} = \sum_{i=1}^{n} \sqrt{d_{\hat{h}_i}}, \quad \bar{d}_{\hat{h}_n} = \sum_{i=1}^{n} d_{\hat{h}_i}. \tag{48}$$

Therefore, all that remains is to determine the value of n for which $d_{\hat{h}_{n+1}} \le \nu(n) < d_{\hat{h}_n}$. Once we have determined n, we choose $\nu = \nu(n)$, and subsequently compute λ using (45) and d_{c_i} using (43).

The semianalytic solution to the robust power loading of eigen beamformers developed above yields some appealing insight. In particular, we observe that the objective function is given by

$$-f = \sum_{i=1}^{N_t} d_{\hat{h}_i} d_{c_i} \left(\frac{\lambda}{\lambda + d_{c_i}}\right)^2. \tag{49}$$

Using the expression for d_{c_i} in (43), we have that

$$-f = \sum_{i=1}^{n} d_{\hat{h}_i} d_{c_i} \left(\frac{\lambda}{\lambda + \lambda\left(\sqrt{\frac{d_{\hat{h}_i}}{\nu}} - 1\right)}\right)^2 = \nu.$$

In addition, substituting (43) into (36) reveals that

$$\epsilon^2 = \sum_{i=1}^{n} \left(\sqrt{d_{\hat{h}_i}} - \sqrt{\nu}\right)^2. \tag{50}$$

Hence, for the case of perfect CSI, for which $\epsilon = 0$, we have $n = 1$ (and $\nu = d_{\hat{h}_1}$), and hence optimal performance is obtained by loading only one of the eigen beams.[2] Since $\nu(n)$ is a decreasing function (cf. (50)), loading only one eigenbeam will continue to be optimal in the presence of small CSI mismatches that satisfy $\epsilon \le \sqrt{d_{\hat{h}_1}} - \sqrt{d_{\hat{h}_2}}$. For larger mismatches, loading the two dominant eigen beams will provide better performance. In general, whenever $\xi_{n-1} \le \epsilon < \xi_n$, where

$$\xi_k^2 = \sum_{i=1}^{k} \left(\sqrt{d_{\hat{h}_i}} - \sqrt{d_{\hat{h}_{k+1}}}\right)^2 \tag{51}$$

we will have power allocated to the n dominant eigen beams. This result links the (robust) power loading of the eigen beams to the spread of the eigenvalues of the (deterministic) autocorrelation $\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathcal{H}}$ of the transmitter's estimate of the spatial channel matrix $\hat{\mathbf{H}}$. If the disparity of the eigenvalues is high, the transmitter will continue to use one-directional beamforming for a larger value of ϵ, meaning that such beamforming will tolerate more CSI imperfections while still offering the best worst-case SNR performance. On the contrary, if the eigenvalues are close in value, the transmitter will have to switch to higher directional beamforming even for very small CSI imperfections.

The computational tasks that must be completed in order to implement the proposed transmission scheme are the eigen decomposition of $\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathcal{H}}$ and the computation of the power loading. The computational cost is dominated by that of the eigen decomposition of $\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathcal{H}}$, which requires $O(N_t^3)$ operations. An alternative approach to transmission in the presence of uncertainty that employs a statistical model of the uncertainty [12] also requires an eigen decomposition of an $N_t \times N_t$ matrix, and hence the computational complexity of the proposed transmission scheme is of the same order as that of the scheme in [12]. That said, the one-directional beamformer in [21], [22] can be computed in $O(N_t^2)$ operations.

C. Optimality of Eigen Beamforming

In the previous subsection we derived a robust power loading algorithm for eigen beamforming. That is, we maximized the worst-case SNR under the constraint that $\mathbf{U}_c = \mathbf{U}_{\hat{h}}$. In this subsection we derive conditions on $\hat{\mathbf{H}}$ and ϵ under which the derived algorithm solves the original max-min SNR problem; i.e., conditions under which our loaded eigen beamformer maximizes the worst-case SNR in the absence of the constraint that $\mathbf{U}_c = \mathbf{U}_{\hat{h}}$. First, we will provide a sufficient condition under which eigen beamforming is optimal. This condition is dependent on the choice of the power loading. We will also provide a (weaker) sufficient condition under which eigen beamforming is optimal independent of the power allocated to each beam.

Our approach to the derivation of the first sufficient condition is to assume that the unitary matrix $\mathbf{U}_{c_*} \ne \mathbf{U}_{\hat{h}}$ and the power loading $\mathbf{d}_{c_*}$ provide a larger worst-case SNR than that achieved by the combination of the eigen beamformer $\mathbf{U}_c = \mathbf{U}_{\hat{h}}$ and the power loading in (43) (we will use a subscript $(\cdot)_*$ to denote quantities associated with the presumed optimal solution). We

[2]One-directional eigen beamforming also provides the optimal average SNR when using standard statistical models for the CSI mismatch [21], [22], but it does not necessarily provide the minimum average symbol error rate [12].

will show that for values of ϵ less than a certain threshold, this assumption leads to a contradiction, and hence that for values of ϵ below this threshold, the choice of $\mathbf{U}_c = \mathbf{U}_{\hat{h}}$ is optimal.

From (30), the worst-case SNR for the presumed optimal solution is

$$\mathrm{SNR}_{w_*} = \mathrm{tr}\left\{\mathbf{D}_{\hat{h}}\mathbf{U}_{\hat{h}c_*}\mathrm{diag}\left\{d_{c_{i*}}\left(\frac{\lambda_*}{d_{c_{i*}}+\lambda_*}\right)^2\right\}\mathbf{U}^{\mathcal{H}}_{\hat{h}c_*}\right\} \tag{52}$$

where the parameter λ_* must satisfy

$$\epsilon^2 = \mathrm{tr}\left\{\mathbf{D}_{\hat{h}}\mathbf{U}_{\hat{h}c_*}\mathrm{diag}\left\{\left(\frac{d_{c_{i*}}}{d_{c_{i*}}+\lambda_*}\right)^2\right\}\mathbf{U}^{\mathcal{H}}_{\hat{h}c_*}\right\}. \tag{53}$$

Let $\mathbf{f}_*$ and $\mathbf{d}_{f_*}$ be the vectors of diagonal elements and eigenvalues, respectively, of the matrix

$$\mathbf{F}_* = \mathbf{U}_{\hat{h}c_*}\mathrm{diag}\left\{\left(\frac{d_{c_{i*}}}{d_{c_{i*}}+\lambda_*}\right)^2\right\}\mathbf{U}^{\mathcal{H}}_{\hat{h}c_*} \tag{54}$$

and note that $d_{f_{i*}} = (d_{c_{i*}}/(d_{c_{i*}}+\lambda_*))^2$. Also, note that the function $g(x) = (x/(x+a))^2$ has a derivative $(\partial g/\partial x) = (2xa/(x+a)^3)$ that is always positive for positive x and a. This means that for any λ_* and $\mathbf{d}_{c_*}$, the elements of $\mathbf{d}_{f_*}$ are in nonincreasing order. Using the definition of $\mathbf{f}_*$, (53) can be rewritten as

$$\epsilon^2 = \sum_{i=1}^{N_t} d_{\hat{h}_i} f_{i*} \tag{55}$$

and using Lemma 1, we have that

$$\epsilon^2 = \sum_{i=1}^{N_t} d_{\hat{h}_i} f_{i*} \leq \sum_{i=1}^{N_t} d_{\hat{h}_i} f_{\pi(i)*} \tag{56}$$

where π is a permutation such that $f_{\pi(i)}$'s are in a nonincreasing order. Continuing in the same manner as in the perfect CSI case and using Lemmas 2 and 3, we can write

$$\epsilon^2 < \sum_{i=1}^{N_t} d_{\hat{h}_i}\left(\frac{d_{c_{i*}}}{d_{c_{i*}}+\lambda_*}\right)^2. \tag{57}$$

In contrast, when we choose $\mathbf{U}_c = \mathbf{U}_{\hat{h}}$, we achieve equality in (57). Since λ is a monotonically decreasing function of ϵ, this implies that $\lambda_{**} > \lambda_*$, where λ_{**} denotes the value of λ achieved by the combination of $\mathbf{U}_c = \mathbf{U}_{\hat{h}}$ and the presumed power loading $\mathbf{d}_{c_*}$.

Similarly, let $\bar{\mathbf{c}}_*$ and $\mathbf{d}_{\bar{c}*}$ be the vectors of diagonal elements and eigenvalues, respectively, of the matrix

$$\bar{\mathbf{C}}_* = \mathbf{U}_{\hat{h}c_*}\mathrm{diag}\left\{d_{c_{i*}}\left(\frac{\lambda_*}{d_{c_{i*}}+\lambda_*}\right)^2\right\}\mathbf{U}^{\mathcal{H}}_{\hat{h}c_*} \tag{58}$$

with $d_{\bar{c}_{i*}} = d_{c_{i*}}(\lambda_*/(d_{c_{i*}}+\lambda_*))^2$. The function $g(x) = x(a/(x+a))^2$ has a derivative $(\partial g/\partial x) = (a^2(a-x)/(x+a)^3)$ that is always nonnegative for $0 \leq x \leq a$. Therefore, if $d_{c_{i*}} \leq \lambda_*$ for all i, then the elements of the vector $\mathbf{d}_{\bar{c}_*}$ are in nonincreasing order. In that case, using Lemmas 1–3, we have that

$$\begin{aligned}\mathrm{SNR}_{w_*} &< \sum_{i=1}^{N_t} d_{\hat{h}_i} d_{c_{i*}}\left(\frac{\lambda_*}{d_{c_{i*}}+\lambda_*}\right)^2 \\ &< \sum_{i=1}^{N_t} d_{\hat{h}_i} d_{c_{i*}}\left(\frac{\lambda_{**}}{d_{c_{i*}}+\lambda_{**}}\right)^2\end{aligned} \tag{59}$$

where the last step follows from the fact that $\lambda_{**} > \lambda_*$. However, the right-hand side of (59) is achieved by making the choice $\mathbf{U}_c = \mathbf{U}_{\hat{h}}$. Therefore, when $d_{c_{i*}} \leq \lambda_*$, or, equivalently, when the elements of $\mathbf{d}_{\bar{c}_*}$ are in nonincreasing order, we have a contradiction to our assumption that the combination of $\mathbf{U}_{c_*} \neq \mathbf{U}_{\hat{h}}$ and $\mathbf{d}_{c_*}$ is optimal. In particular, the combination of $\mathbf{U}_c = \mathbf{U}_{\hat{h}}$ and $\mathbf{d}_{c_*}$ provides better performance. Hence, whenever $d_{c_{i*}} \leq \lambda_*$, eigen beamforming is optimal. Unfortunately, the condition that the elements of $\mathbf{d}_{\bar{c}_*}$ be in nonincreasing order does not immediately generate insight, and therefore we will now expand on this condition.

Consider two consecutive elements of $\mathbf{d}_{\bar{c}_*}$, namely $d_{\bar{c}_{i*}}$ and $d_{\bar{c}_{(i+1)*}}$. For these elements to be in nonincreasing order, we require that

$$d_{c_{i*}}\left(\frac{\lambda}{d_{c_{i*}}+\lambda}\right)^2 \geq d_{c_{(i+1)*}}\left(\frac{\lambda}{d_{c_{(i+1)*}}+\lambda}\right)^2. \tag{60}$$

Since $d_{c_{i*}} \geq 0$ and $\lambda_* \geq 0$, this condition is equivalent to

$$\lambda \geq \sqrt{d_{c_{i*}} d_{c_{(i+1)*}}}. \tag{61}$$

Furthermore, since the elements of $\mathbf{d}_{c_*}$ are in nonincreasing order, inequality (61) reduces to

$$\lambda \geq \sqrt{d_{c_{1*}} d_{c_{2*}}}. \tag{62}$$

Now, since λ is a monotonically decreasing function of ϵ, (62) provides a range of values of ϵ for which we can guarantee that the elements of $d_{\bar{c}_{i*}}$ are in nonincreasing order. To be more precise, given a power loading vector $\mathbf{d}_{c_*}$, if $\epsilon \leq \bar{\epsilon}$, where

$$\bar{\epsilon}^2 = \sum_{i=1}^{N_t} d_{\hat{h}_i}\left(\frac{d_{c_{i*}}}{d_{c_{i*}}+\sqrt{d_{c_{1*}} d_{c_{2*}}}}\right)^2 \tag{63}$$

then the elements of $d_{\bar{c}_{i*}}$ will be in nonincreasing order and hence eigen beamforming is optimal.

Unfortunately, the condition in (63) is dependent on the choice of the power loading $\mathbf{d}_{c*}$. While this condition can be checked *a posteriori* (i.e., after having chosen the power loading), it cannot be checked *a priori*. In order to derive an *a priori* condition for eigen beamforming to be optimal, we observe that since λ is a monotonically decreasing function of ϵ, the elements of $\mathbf{d}_{\bar{c}}$ will be in nonincreasing order for *any* power loading that satisfies the power constraint with equality if $\epsilon \leq \epsilon_{\mathrm{th}}$, where

$$\epsilon_{\mathrm{th}} = \min_{\sum_{i=1}^{N_t} d_{c_{i*}}=1} \bar{\epsilon} \tag{64}$$

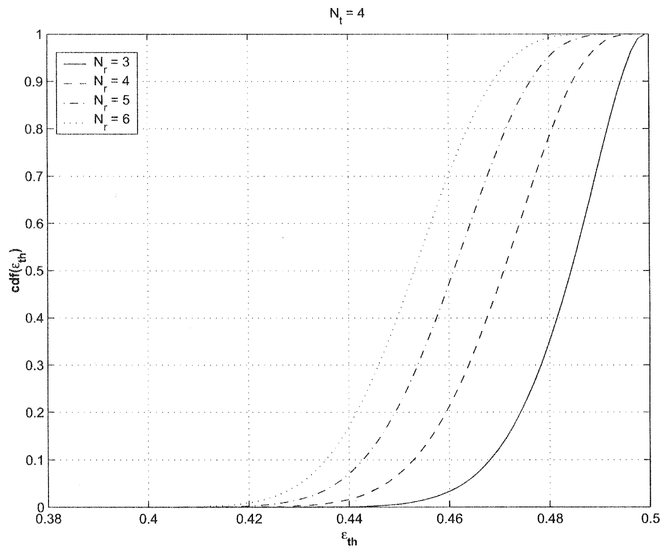


Fig. 2. cdf of $\epsilon_{\rm th}$ for Rayleigh fading channels.

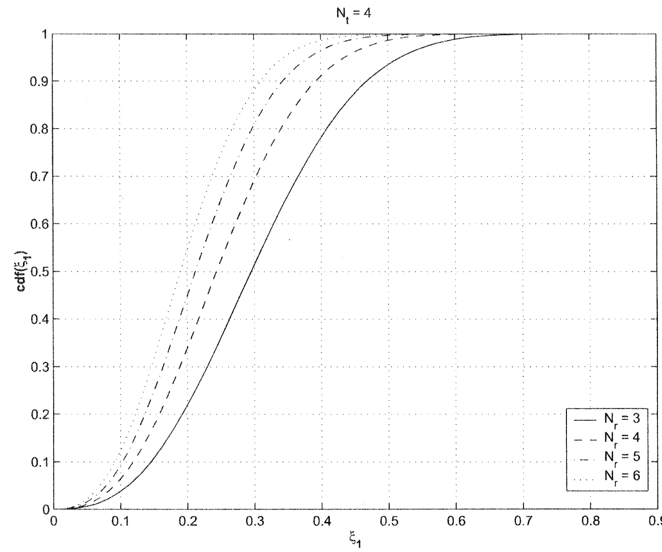


Fig. 3. cdf of ξ_1 for Rayleigh fading channels.

is the threshold value of ϵ. As we show in Appendix IV, the problem in (64) can be solved analytically, and

$$\epsilon_{\rm th} = \frac{\sqrt{d_{\hat{h}_1} + d_{\hat{h}_2}}}{2}. \tag{65}$$

This result adds strength to results in [12] and [13], which suggest that the two strongest eigenvectors play a crucial role in characterizing the channel.

In summary, we have shown in this subsection that for a given power loading $\mathbf{d}_{c_*}$, eigen beamforming is optimal if $\epsilon \leq \bar{\epsilon}$. Moreover, eigen beamforming is optimal for all possible power loadings if $\epsilon \leq \epsilon_{\rm th}$. While this second condition is conservative, it has the advantage that it can be tested before the power loading is chosen.

V. Simulation Results

In this section, we present simulation results that provide insight into the performance of the proposed robust eigen beamformer. Without any loss of generality, we will assume that the parameter ϵ is normalized to $\|\hat{\mathbf{H}}\|_F$ so that $\epsilon \in [0,1]$. Note that in the case of such normalization of this parameter, $\epsilon = 0$ corresponds to perfect CSI at the transmitter, and $\epsilon = 1$ corresponds to an uncertainty that is of the same size as the channel estimate itself.

In Figs. 2–4, we investigate the eigenvalue behavior of the (deterministic) autocorrelation of the spatial channel matrix for Rayleigh flat-fading channels, in which the channel coefficients are independent identically distributed (i.i.d.) zero-mean complex Gaussian random variables. It is obvious that when $\text{rank}(\hat{\mathbf{H}}) = 1$ or 2, we would have $\epsilon_{\rm th} = 1/2$, and it is expected that as the rank increases, the value of $\epsilon_{\rm th}$ will decrease. Fig. 2 provides the cumulative distribution function (cdf) of $\epsilon_{\rm th}$ for $N_t = 4$ transmit antennas and $N_r = 3$, 4, 5 and 6 receive antennas. It is evident that the value of $\epsilon_{\rm th}$ generally decreases with increasing N_r. However, the decrease is rather slow and $\epsilon_{\rm th}$ appears to be lower bounded even as the number of receive antennas is increased. For instance, the chances of having a Rayleigh flat-fading channel with $\epsilon_{\rm th} \leq 0.46$ are rather small for $N_r = 3$, while the probability is approximately 0.2, 0.5, and 0.7 for $N_r = 4$, 5, and 6, respectively. For all the considered values of N_r, the probability of having $\epsilon_{\rm th} \leq 0.42$ is virtually zero.

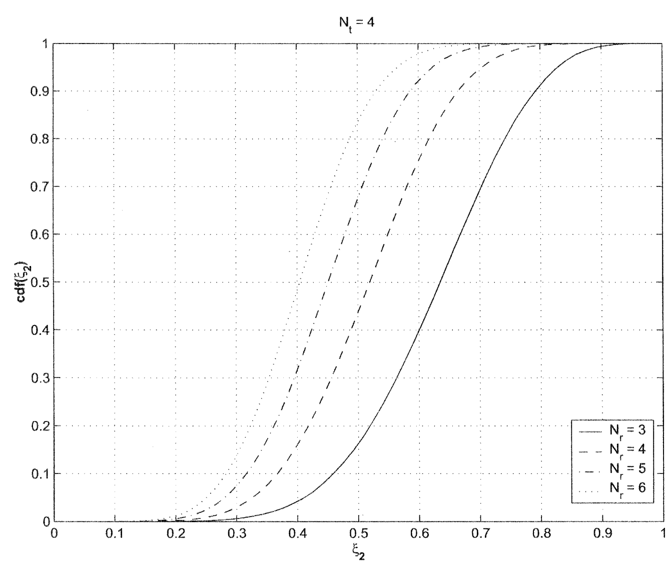


Fig. 4. cdf of ξ_2 for Rayleigh fading channels.

The cdf of ξ_1 is shown in Fig. 3, where ξ_1 is the uncertainty threshold for a channel that signals the transition from the one-directional beamforming to the two-directional one. A one-directional beamformer will offer the best worst-case performance up to an uncertainty of size ξ_1, after which a two-directional beamformer will have to be used. For CSI mismatch with $\epsilon = 0.3$, the probability of using two-directional beamforming is approximately 0.5 for $N_r = 3$ and 0.7, 0.8, and 0.9 for $N_r = 4$, 5, and 6, respectively. On the other hand, Fig. 4 shows the cdf of ξ_2 that signals the transition from two-directional to three-directional beamforming. For CSI mismatch with $\epsilon = 0.5$, the probability of using three-directional beamforming is approximately 0.18 for $N_r = 3$ and 0.45, 0.68, and 0.82 for

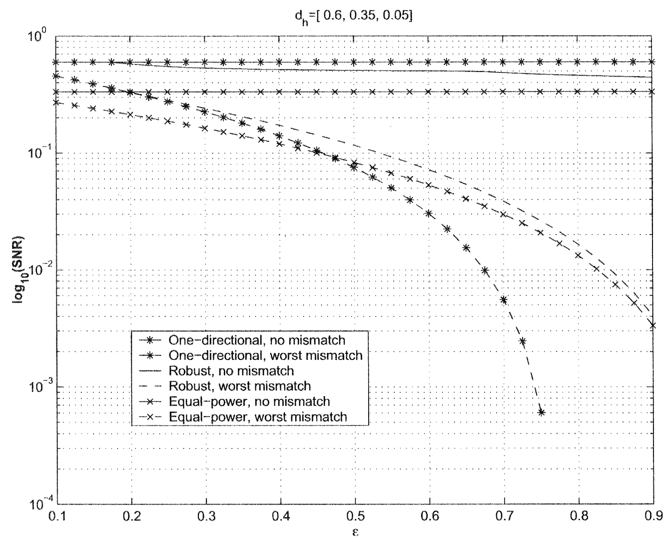


Fig. 5. SNR of conventional, robust, and omnidirectional beamformers versus the norm of error matrix.

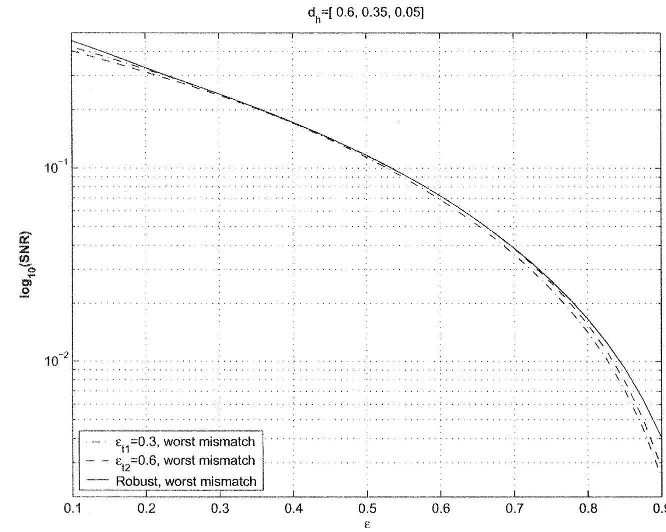


Fig. 6. Worst-case SNR for the proposed robust beamformer with precise design value for ϵ and fixed design values for ϵ.

$N_r = 4, 5$, and 6, respectively. We can conclude that, for a specific N_t, the probability of using k-directional beamforming increases with N_r.

In Fig. 5, we plot the no-mismatch and worst-case receiver SNRs for one-directional, robust, and equal-power eigen beamformers versus the norm of the mismatch error matrix for a sample channel with $\mathbf{d}_{\hat{h}} = [0.65, 0.3, 0.05]^T$. The no-mismatch SNR corresponds to the performance of the beamformer when there is no mismatch in the transmitter's channel model, i.e., $\mathbf{E} = \mathbf{0}$, while the worst-case SNR corresponds to the case of worst-case mismatch $\mathbf{E}_w$ given in (28) which is different for each beamformer. This plot shows that when $\epsilon = 0$ (i.e., perfect CSI), one-directional beamforming offers the highest SNR (since it corresponds to the perfect CSI solution). In contrast, the equal-power beamforming offers the worst SNR of those considered (since it corresponds to the no CSI solution). The proposed robust beamformer represents a tradeoff between the two, and its worst-case performance approaches that of the equal-power beamformer as ϵ increases. Note, however, that the robust beamformer does not reduce to an equal-power beamformer even when $\epsilon = 1$, which is clear from Fig. 1 when $\nu = 0$. On the other hand, it is clear that the robust beamformer offers the best worst-case performance at all values of ϵ among the considered beamformers.

It is also worth noting that equal-power beamforming has the smallest SNR spread, where we use the term "SNR spread" to refer to the difference between the SNRs of the best-case (i.e., perfect CSI) and the worst-case variants of a particular beamformer. In contrast, one-directional beamforming has the largest SNR spread and that of robust beamforming lies in between. These results suggest that while switching from one-directional beamforming to full diversity, we are actually trading the mean of receiver SNR distribution with its variance, as pointed out in [12]. For example, one-directional beamforming has the highest mean and highest variance while equal-power beamforming has the lowest mean and lowest variance (see also Fig. 9).

The results in Fig. 5 assumed that the value of ϵ used in the design, which we will denote by ϵ_t in this paragraph, is equal to the bound on the actual channel uncertainties that arose in that example. Fortunately, the worst-case SNR is reasonably insensitive to ϵ_t. In particular, in Fig. 6, we have plotted the worst-case SNR for our design for the case in which $\epsilon_t = \epsilon$ (as in Fig. 5), and for cases in which ϵ_t is fixed at $\epsilon_{t_1} = 0.3$ and at $\epsilon_{t_2} = 0.6$.

We now consider a multiple-input multiple-output (MIMO) system in which the receiver acquires the channel perfectly and feeds it back to the transmitter. We will consider a delayed feedback scenario that was used in [11], [12], [22], and [26]. In this scenario, it is assumed that the channel coefficients are independent zero-mean complex Gaussian and that they change slowly with time according to Jakes' model. The estimate available at the transmitter $\hat{\mathbf{H}}$ and the actual channel known by the receiver $\mathbf{H}$ are thus drawn from the same distribution and have a correlation coefficient ρ that depends on how fast the channel changes in addition to the feedback delay time. Perfect channel knowledge corresponds to $\rho = 1$, while no channel knowledge corresponds to $\rho = 0$.

The choice of the design value of ϵ in our approach involves a tradeoff between the impact of a large error in the transmitter's channel model on the relevant system performance measure, and the impact of the degradation in performance in the absence of error in the transmitter's channel model. In what follows, the system performance measure will be the symbol error rate (SER), and at low SNRs, the impact on the SER of channel modeling error at the transmitter is dominated by the impact of noise, and hence a small design value of ϵ is sufficient. However, at higher SNRs, the dominant component of the symbol error rate is that due to outages caused by inaccurate channel information at the transmitter. (Recall that we have assumed that the receiver has an accurate model for the channel.) Therefore, a larger design value for ϵ is appropriate for higher SNRs.

In Figs. 7 and 8, we investigate the average SER performance for a four-transmit- and two-receive-antenna system using 16-QAM and quaternary-phase-shift-keying (QPSK) modulations, respectively. We assume a Rayleigh flat-fading channel and consider the case when $\rho = 0.85$ in Fig. 7 and $\rho = 0.75$ in Fig. 8. In the case of the QPSK simulations in

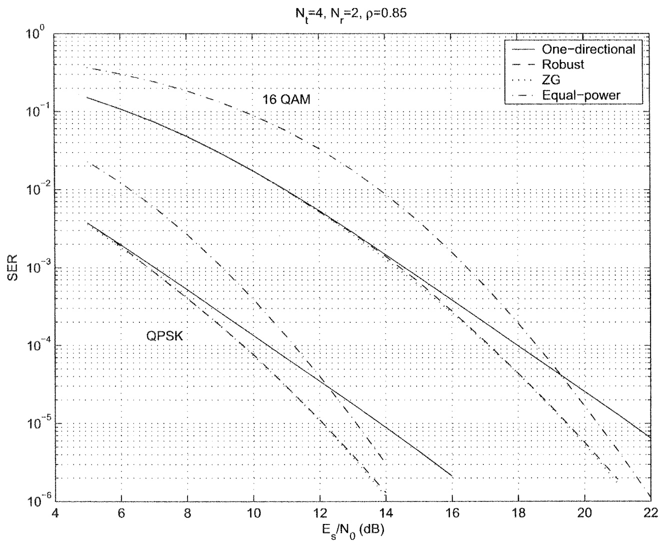


Fig. 7. SER versus E_s/N_0 for a Rayleigh fading channel and delayed feedback with correlation coefficient $\rho = 0.85$.

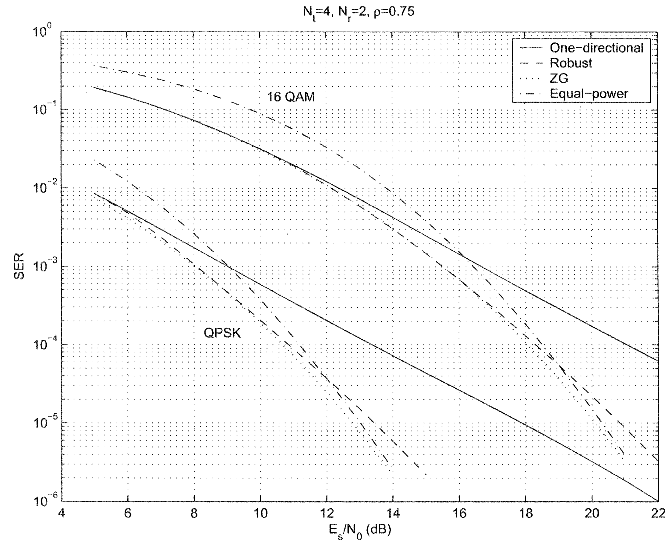


Fig. 8. SER versus E_s/N_0 for a Rayleigh fading channel and delayed feedback with correlation coefficient $\rho = 0.75$.

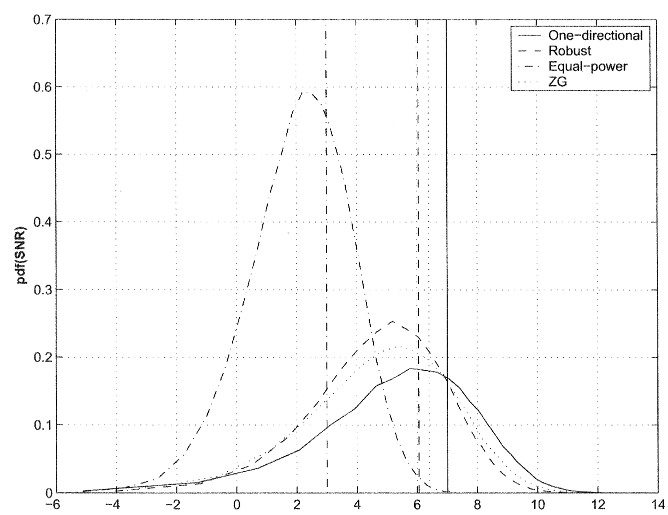


Fig. 9. SNR pdf for a Rayleigh fading channel and delayed feedback with correlation coefficient $\rho = 0.85$.

these figures, the design value of ϵ was increased linearly with $\log(\mathrm{SNR})$ from 0 to 0.95 over the SNR range of 5 to 10 dB, and in the case of the 16-QAM simulations, the design value of ϵ was increased linearly with $\log(\mathrm{SNR})$ from 0 to 0.95 over the SNR range of 10 to 15 dB. Both figures show that the proposed transmit beamformer offers an improved robustness against CSI mismatch while maintaining average error performance that is always better than that of the one-directional eigen beamformer and is better than that of the equal-power beamformer over a fairly wide range of SNR. The fact that the equal-power beamformer performs better at high SNRs is a result of the eigen beamforming structure being potentially suboptimal for $\epsilon > \epsilon_{\mathrm{th}}$. This suboptimality manifests itself as a lower diversity gain than the equal-power beamformer, as is evident from the slope of the SER curves at high SNR. However, we emphasize that the robust beamformer provides the best performance over the range of common operating SERs. In addition, the fact that it almost always uses a lower number of beam directions than the equal-power beamformer makes it a more plausible alternative, especially in multiuser scenarios where multiuser interference has a significant impact on system performance. Figs. 7 and 8 also show the performance of the statistically motivated transmission scheme of Zhou and Giannakis [12], which we will call the ZG scheme. Note that the performance of our approach is nearly identical to that of the ZG scheme in Fig. 7 and that the performance of the ZG scheme is only marginally better in Fig. 8. This is despite the fact that in this scenario the uncertainty in the transmitter's channel model is identical to the statistical model that is assumed in the ZG scheme. In Figs. 10 and 11, we provide scenarios in which mismatch between the actual and presumed statistical models for the uncertainty leads to a degradation in the performance of the ZG scheme.

Fig. 9 shows the SNR pdf of the one-directional, robust, ZG and equal-power beamformers when $\rho = 0.85$ and $\epsilon = 0.95$. The vertical lines show the mean of the corresponding pdf's. As was mentioned above, the tradeoff offered by the robust and ZG beamformers between the mean and variance is clear. Even though the one-directional beamformer has the highest mean, the positive effect of the mean is not enough to mitigate the detrimental effect of the variance on the average SER of a communications system, because the SER is dominated by the worst-case performance. It is clear from this figure that the tails of the SNR pdf of the robust, ZG, and equal-power beamformers decay faster.

In Fig. 10, we investigate a different model for the uncertainty in the transmitter's model of the channel. As in the previous example, we assume a Rayleigh fading model for the channel matrix $\mathbf{H}$; each element of $\mathbf{H}$ is an i.i.d. zero-mean complex Gaussian random variable with variance $\sigma_h^2 = 1$. However, the (i, j)th element of the uncertainty matrix $\mathbf{E}$ is modeled as an i.i.d. zero-mean complex Gaussian random variable with variance, σ_e^2, proportional to the Frobenius norm of $\mathbf{H}$. This could more appropriately model our assumption that $\|\mathbf{E}\|_F \leq \|\hat{\mathbf{H}}\|_F$.

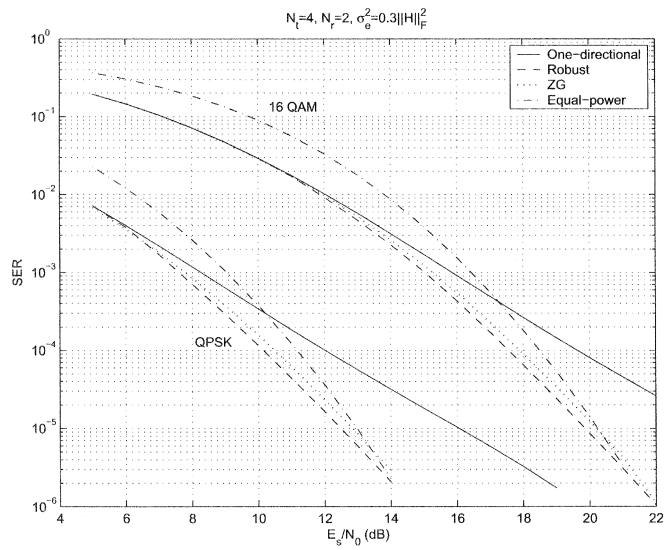


Fig. 10. SER versus E_s/N_0 for a Rayleigh fading channel and Gaussian channel uncertainty with variance $\sigma_e^2 = 0.3\|\mathbf{H}\|_F^2$.

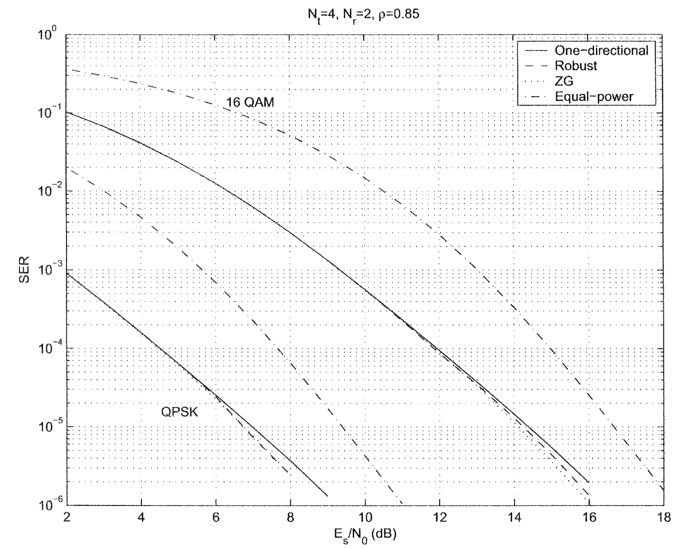


Fig. 12. SER versus E_s/N_0 for a Rician fading channel with $\mathcal{K} = 1$ and delayed feedback with correlation coefficient $\rho = 0.85$.

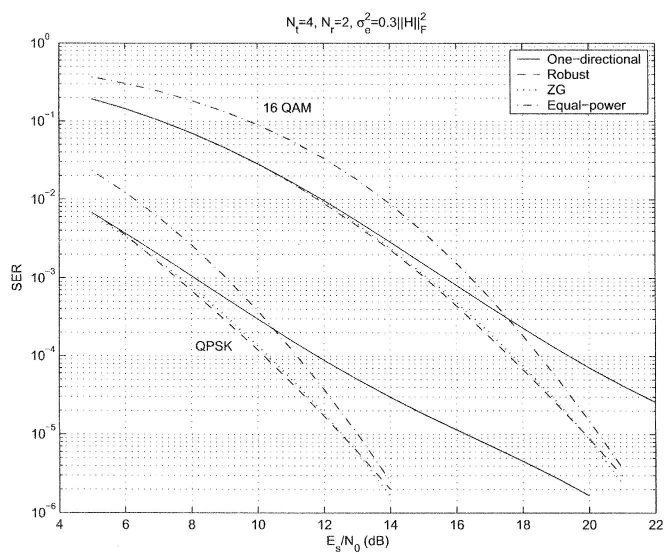


Fig. 11. SER versus E_s/N_0 for a Rayleigh fading channel and uniform channel uncertainty with variance $\sigma_e^2 = 0.3\|\mathbf{H}\|_F^2$.

In Fig. 10, we plot the average SER for $\sigma_e^2 = 0.3\ \|\mathbf{H}\|_F^2$. The same approach is repeated in Fig. 11, but with the elements of $\mathbf{E}$ coming from a uniform distribution that can model quantization errors. In these scenarios, there is a mismatch between the uncertainty model and the model assumed in the ZG approach. This mismatch makes the tuning of the parameter ρ in the ZG approach rather awkward, and we assumed a value of $\rho = 0.85$ for comparison purposes. As can be seen from the figures, the performance of our robust power loading remains essentially unaffected by the model change and slightly outperforms the ZG approach in both cases, although the performance advantage is rather small in Fig. 11.

The previous examples considered the case of Rayleigh fading. In the presence of a strong line-of-sight component, the results have a similar flavor, but the performance of the conventional one-directional beamformer is much closer to that of the robust beamformers. For example, in Fig. 12, we have plotted the SER performance of several transmission strategies for a Rician fading channel with a strong line-of-sight component (the Rician $\mathcal{K}$ factor was 1). The uncertainty in the transmitter's channel model is modeled in the same way as in Fig. 7. The good performance of the one-directional beamformer in this scenario is due to the fact that for line-of-sight channels $\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathcal{H}}$ will typically have a dominant eigenvalue. In that case, the proposed robust beamformer will be a one-directional beamformer for a significant range of values of ϵ. (See the discussion toward the end of Section IV-B.)

It should be pointed out that, with the lack of any statistical assumptions about the channel or the mismatch, the main emphasis of our approach is on the instantaneous transmission performance with an unknown CSI mismatch. Under this assumption, using the worst-case performance as a design criterion is fully justified. On the other hand, other existing approaches that involve capacity or average SER criteria have more emphasis on the performance averaged over longer periods of time and thus have to assume detailed statistical models for this purpose, which could make them vulnerable to model mismatch. We have shown in this section that our approach can indeed deliver good average performance for different mismatch models.

VI. Conclusion

In this paper, a robust framework for the design of transmit beamforming in the presence of partial CSI at the transmitter has been proposed. In this framework, no statistical assumptions about the CSI mismatch or the channel have been made. Our robust transmit beamformer has been designed to have the best performance under the worst-case CSI mismatch. It has been shown that, up to a threshold level on the channel uncertainty, eigen beamforming along the eigenvectors of the (deterministic) autocorrelation of the spatial channel matrix as perceived by the transmitter offers the best worst-case performance among all other beamformers. It has also been shown that this threshold is dependent on the gain along the two strongest eigenvectors. A robust power loading algorithm has also been proposed for eigen beamforming that is implemented via a spatial water-filling-type strategy in which the water level is determined in a simple way.

APPENDIX I
PROOF OF LEMMA 1

Let the maximizing permutation be such that for some $i, j \in \{1, 2 \ldots, n\}$, $\alpha_{\pi^*(i)} > \alpha_{\pi^*(j)}$ and $\beta_i < \beta_j$. We will show that this leads to a contradiction. Specifically, let

$$\begin{aligned} \alpha_{\pi^*(i)}(\beta_j - \beta_i) &> \alpha_{\pi^*(j)}(\beta_j - \beta_i), \\ \alpha_{\pi^*(i)}\beta_j + \alpha_{\pi^*(j)}\beta_i &> \alpha_{\pi^*(i)}\beta_i + \alpha_{\pi^*(j)}\beta_j. \end{aligned} \tag{66}$$

Therefore, the sum $\sum_{i=1}^{n} \alpha_{\pi(i)}\beta_i$ can be increased further by switching the values of $\pi^*(i)$ and $\pi^*(j)$. Hence, π^* cannot be the maximizing permutation, and we have a contradiction.

APPENDIX II
PROOF OF LEMMA 3

We outline the first two steps of the proof. The remaining steps follow in an iterative fashion. By the majorization assumption we have that

$$\beta_1 \geq \alpha_1 \tag{67}$$

$$\beta_1 + \beta_2 \geq \alpha_1 + \alpha_2. \tag{68}$$

Therefore

$$\gamma_2(\beta_1 + \beta_2) + \beta_1(\gamma_1 - \gamma_2) \geq \gamma_2(\alpha_1 + \alpha_2) + \alpha_1(\gamma_1 - \gamma_2)$$

and hence

$$\beta_1\gamma_1 + \beta_2\gamma_2 \geq \alpha_1\gamma_1 + \alpha_2\gamma_2. \tag{69}$$

Proceeding iteratively in this fashion, we arrive at the desired result.

APPENDIX III
KKT SOLUTION FOR ROBUST POWER LOADING

In this Appendix, we derive the expressions for the Karush–Kuhn–Tucker (KKT) solution of the robust power loading problem.

We start with the equation

$$\sum_{j=1}^{N_t} d_{\hat{h}_j} \left(\frac{d_{c_j}}{d_{c_j} + \lambda} \right)^2 = \epsilon^2. \tag{70}$$

Differentiating with respect to d_{c_i}, we have

$$\sum_{j=1}^{N_t} \left(\frac{2d_{\hat{h}_j} d_{c_j}}{d_{c_j} + \lambda} \right) \left[\frac{\frac{\partial d_{c_j}}{\partial d_{c_i}}}{d_{c_j} + \lambda} - \frac{d_{c_j}\left(\frac{\partial d_{c_j}}{\partial d_{c_i}} + \frac{\partial \lambda}{\partial d_{c_i}} \right)}{\left(d_{c_j} + \lambda\right)^2} \right] = 0. \tag{71}$$

Therefore

$$\sum_{j=1}^{N_t} \frac{d_{\hat{h}_j} d_{c_j}}{(d_{c_j} + \lambda)^3} \left[\lambda\delta(i-j) - d_{c_j} \frac{\partial \lambda}{\partial d_{c_i}} \right] = 0 \tag{72}$$

and, finally

$$\frac{\partial \lambda}{\partial d_{c_i}} = \frac{\frac{\lambda d_{\hat{h}_i} d_{c_i}}{(d_{c_i}+\lambda)^3}}{\sum\limits_{j=1}^{N_t} \frac{d_{\hat{h}_j} d_{c_j}^2}{(d_{c_j}+\lambda)^3}}. \tag{73}$$

To prove (39), we use (37). Differentiating it with respect to d_{c_i}, we have

$$\begin{aligned} -\frac{\partial f}{\partial d_{c_i}} &= \sum_{j=1}^{N_t} d_{\hat{h}_j} \\ &\quad \times \left\{ \frac{\partial d_{c_j}}{\partial d_{c_i}} \left(\frac{\lambda}{d_{c_j} + \lambda} \right)^2 + \left(\frac{2\lambda d_{c_j}}{d_{c_j} + \lambda} \right) \right. \\ &\quad \left. \times \left[\frac{\frac{\partial \lambda}{\partial d_{c_i}}}{d_{c_j} + \lambda} - \frac{\lambda\left(\frac{\partial \lambda}{\partial d_{c_i}} + \frac{\partial d_{c_j}}{\partial d_{c_i}} \right)}{(d_{c_j} + \lambda)^2} \right] \right\} \\ &= d_{\hat{h}_i} \left[\left(\frac{\lambda}{d_{c_i} + \lambda} \right)^2 - \frac{2\lambda^2 d_{c_i}}{(d_{c_i} + \lambda)^3} \right] \\ &\quad + \frac{\partial \lambda}{\partial d_{c_i}} \sum_{j=1}^{N_t} \left\{ \frac{2\lambda d_{\hat{h}_j} d_{c_j}}{(d_{c_j} + \lambda)^3} \left[(d_{c_j} + \lambda) - \lambda \right] \right\} \\ &= \frac{\lambda^2 d_{\hat{h}_i}(\lambda - d_{c_i})}{(d_{c_i} + \lambda)^3} + \frac{\partial \lambda}{\partial d_{c_i}} \sum_{j=1}^{N_t} \frac{2\lambda d_{\hat{h}_j} d_{c_j}^2}{(d_{c_j} + \lambda)^3}. \end{aligned} \tag{74}$$

Substituting $(\partial\lambda/\partial d_{c_i})$ from (73), we have

$$\begin{aligned} -\frac{\partial f}{\partial d_{c_i}} &= \frac{\lambda^2 d_{\hat{h}_i}(\lambda - d_{c_i})}{(d_{c_i} + \lambda)^3} + \frac{2\lambda^2 d_{\hat{h}_i} d_{c_i}}{(d_{c_i} + \lambda)^3} \\ &= d_{\hat{h}_i} \left(\frac{\lambda}{d_{c_i} + \lambda} \right)^2. \end{aligned} \tag{75}$$

APPENDIX IV
DERIVATION OF ϵ_{th}

Attempting to solve the problem in (64) in its current form using the Lagrange multiplier method is difficult due to the large number of equations involved. Instead, let $\bar{\epsilon}^2 = \bar{\epsilon}_1^2 + \bar{\epsilon}_2^2$, where

$$\begin{aligned} \bar{\epsilon}_1^2 &= d_{\hat{h}_1} \left(\frac{d_{c_1}}{d_{c_1} + \sqrt{d_{c_1} d_{c_2}}} \right)^2 + d_{\hat{h}_2} \left(\frac{d_{c_2}}{d_{c_2} + \sqrt{d_{c_1} d_{c_2}}} \right)^2 \\ &= \frac{d_{\hat{h}_1} d_{c_1} + d_{\hat{h}_2} d_{c_2}}{\left(\sqrt{d_{c_1}} + \sqrt{d_{c_1}} \right)^2} \end{aligned} \tag{76}$$

and

$$\bar{\epsilon}_2^2 = \sum_{i=3}^{N_t} d_{\hat{h}_i} \left(\frac{d_{c_i}}{d_{c_i} + \sqrt{d_{c_1} d_{c_2}}} \right)^2. \tag{77}$$

We will start by considering the reduced optimization problem over $\bar{\epsilon}_1^2$ only, namely

$$\min_{d_{c_1} \geq d_{c_2}} \frac{d_{\hat{h}_1} d_{c_1} + d_{\hat{h}_2} d_{c_2}}{\left(\sqrt{d_{c_1}} + \sqrt{d_{c_1}} \right)^2}. \tag{78}$$

The Lagrangian can be written as

$$\mathcal{L} = \frac{d_{\hat{h}_1} d_{c_1} + d_{\hat{h}_2} d_{c_2}}{\left(\sqrt{d_{c_1}} + \sqrt{d_{c_1}}\right)^2} + v(d_{c_2} - d_{c_1}) \tag{79}$$

where v is the Lagrange multiplier. Differentiating with respect to d_{c_1} and d_{c_2} we get

$$\sqrt{\frac{d_{c_2}}{d_{c_1}}} \left(\frac{d_{\hat{h}_1} \sqrt{d_{c_1}} - d_{\hat{h}_2} \sqrt{d_{c_2}}}{\left(\sqrt{d_{c_1}} + \sqrt{d_{c_2}}\right)^3} \right) - v = 0 \tag{80}$$

and

$$\sqrt{\frac{d_{c_1}}{d_{c_2}}} \left(\frac{d_{\hat{h}_1} \sqrt{d_{c_1}} - d_{\hat{h}_2} \sqrt{d_{c_2}}}{\left(\sqrt{d_{c_1}} + \sqrt{d_{c_2}}\right)^3} \right) + v = 0 \tag{81}$$

respectively. Adding the two equations to eliminate v, we obtain

$$\left(\frac{d_{c_1} - d_{c_2}}{\sqrt{d_{c_1} d_{c_2}}} \right) \left(\frac{d_{\hat{h}_1} \sqrt{d_{c_1}} - d_{\hat{h}_2} \sqrt{d_{c_2}}}{\left(\sqrt{d_{c_1}} + \sqrt{d_{c_2}}\right)^3} \right) = 0. \tag{82}$$

There are two possible solutions for this equation. The first is $d_{c_2} = (d_{\hat{h}_1}/d_{\hat{h}_2})^2 d_{c_1}$ which obviously violates the constraint $d_{c_1} \geq d_{c_2}$. The second solution is $d_{c_1} = d_{c_2}$ and results in a minimum $\bar{\epsilon}_1^2 = (d_{\hat{h}_1} + d_{\hat{h}_2})/4$ for *any* pair (d_{c_1}, d_{c_2}) that satisfies the power constraint with equality. The minimum of $\bar{\epsilon}_2^2$ is obviously achieved when $d_{c_3} = d_{c_4} \ldots d_{c_{N_t}} = 0$, which leads to the argument of the solution of the optimization in (64) being $\mathbf{d}_c = [1/2, 1/2, 0, \ldots, 0]^T$ and to ϵ_{th} being given by (65).

REFERENCES

[1] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space–time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[2] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space–time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

[3] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1437–1450, Oct. 1998.

[4] E. Visotsky and U. Madhow, "Optimum beamforming using transmit antenna arrays," in *Proc. Vehicular Technology Conf. (VTC)*, vol. 1, Houston, TX, May 1999, pp. 851–856.

[5] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space–time linear precoders and decoders," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1051–1064, May 2002.

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

[7] S. Zhou and G. B. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1285–1294, Jul. 2004.

[8] S. A. Jafar and A. Goldsmith, "On optimality of beamforming for multiple antenna systems with imperfect feedback," in *Proc. Int. Symp. Information Theory*, Jun. 2001, pp. 321–321.

[9] S. A. Jafar, S. Vishwanath, and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *Proc. Int. Conf. Communications*, vol. 7, Jun. 2001, pp. 2266–2270.

[10] E. Visotsky and U. Madhow, "Space–time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2632–2639, Sep. 2001.

[11] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming and space–time block coding," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 611–627, Mar. 2002.

[12] S. Zhou and G. B. Giannakis, "Optimal transmitter eigen-beamforming and space–time block coding based on channel mean feedback," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2599–2613, Oct. 2002.

[13] ——, "Optimal transmitter eigen-beamforming and space–time block coding based on channel correlations," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1673–1690, Jul. 2003.

[14] A. Abdel-Samad, A. B. Gershman, and T. Davidson, "Robust transmit beamforming based on imperfect channel feedback," in *Proc. Vehicular Technology Conf. (VTC)—Fall*, Sep. 2004, pp. 2049–2053.

[15] ——, "Robust transmit eigen-beamforming based on imperfect channel state information," in *Proc. ITG Workshop Smart Antennas*, Apr. 2004, pp. 200–207.

[16] S. A. Vorobyov, A. B. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, Feb. 2003.

[17] S. Shahbazpanahi, A. B. Gershman, Z.-Q. Luo, and K. M. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2257–2269, Sep. 2003.

[18] S. Shahbazpanahi and A. B. Gershman, "Robust blind multiuser detection for synchronous CDMA systems using worst-case performance optimization," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 2232–2245, Nov. 2004.

[19] K. Zarifi, S. Shahbazpanahi, A. B. Gershman, and Z.-Q. Luo, "Robust blind multiuser detection based on the worst-case performance optimization of the MMSE receiver," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 295–305, Jan. 2005.

[20] A. Pascual-Iserte, A. Perez-Neira, and M. A. Lagunas, "A maximin approach for robust MIMO design: Combining OSTBC and beamforming with minimum transmission power requirements," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 2, Mar. 2004, pp. 1–4.

[21] J. K. Cavers, "Single-user and multiuser adaptive maximal ratio transmission for Rayleigh channels," *IEEE Trans. Vehicular Technol.*, vol. 49, no. 6, pp. 2043–2050, Nov. 2000.

[22] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.

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

[24] R. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[25] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Nashua, NH: Athena Scientific, 1999.

[26] E. Visotsky and U. Madhow, "Space–time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2632–2639, Sep. 2001.

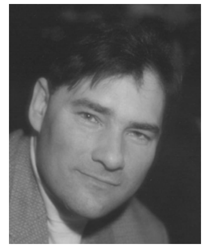
Ayman Abdel-Samad received the B.Sc. and the M.Sc. degrees in electronics and communication engineering from Cairo University, Giza, Egypt, in 1992 and 1995, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, in 2001.

From 2001 to 2002, he was a Member of Technical Staff of BOPS, Inc., Chapel Hill, NC, where he was involved in designing DSP algorithms for broadband wireless local area networks. From 2002 to 2004, he was with the University of Duisburg-Essen, Duisburg, Germany, as a Postdoctoral Researcher, where he was associated with the Smart Antenna Research Team (SmART) and worked on transmit beamforming for multiple-input multiple-output (MIMO) communication systems. Since 2004, he has been a Postdoctoral Fellow with the IC Design Department, MIMOS Berhad, Kuala, Lumpur, Malaysia. His research interests are in signal processing for communications.

Timothy N. Davidson (M'96) 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, Oxford, U.K., in 1995.

He is currently an Associate Professor in the Department of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada, where he holds the (Tier II) Canada Research Chair in Communication Systems. His research interests are in signal processing, communications, and control, with current activity focused on signal processing for digital communication systems. 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.



Alex B. Gershman (M'97–SM'98–F'06) received the Diploma and Ph.D. degrees in radiophysics from the Nizhny Novgorod State University, Russia, in 1984 and 1990, respectively.

From 1984 to 1999, he held several full-time and visiting research appointments in Russia, Switzerland, and Germany. In 1999, he joined the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he became a Full Professor in 2002. From 2003 to 2005, he held a visiting professorship at the Department of Communication Systems, University of Duisburg-Essen, Duisburg, Germany. Since April 2005, he has been with the Darmstadt University of Technology, Darmstadt, Germany, as a Full Professor and Head of Communication Systems Group. His research interests are in the area of signal processing and communications with primary emphasis on statistical signal and array processing, adaptive beamforming, spatial diversity in MIMO communications, multiuser communications, and parameter estimation and detection. He has (co)authored several book chapters and more than 90 journal and 150 conference papers on these subjects. He has also coedited two books: *High-Resolution and Robust Signal Processing* (New York: Marcel Dekker, 2003, with Y. Hua and Q. Cheng) and *Space-Time Processing for MIMO Communications* (New York: Wiley, 2005, with N. D. Sidiropoulos).

Dr. Gershman received the 2004 IEEE Signal Processing Society Best Paper Award for his paper on robust minimum variance beamforming. He coauthored the paper that received the 2005 IEEE Signal Processing Society Young Author Best Paper Award. He was also a recipient of the 2002 Young Explorers Prize from the Canadian Institute for Advanced Research (CIAR) which has honored Canada's top 20 researchers aged 40 or under; the 2001 Wolfgang Paul Award from the Alexander von Humboldt Foundation, Germany; the 2000 Premier's Research Excellence Award, ON, Canada; the 1995–1996 Alexander von Humboldt Fellowship (Germany); the 1994 Swiss Academy of Engineering Science and Branco Weiss Fellowship (Switzerland); the 1994 Outstanding Young Scientist Presidential Fellowship (Russia); and the 1993 International Union of Radio Science (URSI) Young Scientist Award. He has been Editor-in-Chief of the IEEE SIGNAL PROCESSING LETTERS since 2006. He was Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING (1999–2005) and Guest Editor of the *EURASIP Journal of Applied Signal Processing* in 2004 and has been on Editorial Boards of *EURASIP Journal on Wireless Communications and Networking* and *EURASIP Signal Processing Journal* from 2003 and 2005, respectively. He is Vice-Chair of the Sensor Array and Multichannel (SAM) Technical Committee (TC) and a Member of the Signal Processing Theory and Methods (SPTM) TC of the IEEE Signal Processing Society. He was Technical Co-Chair of the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Darmstadt, Germany, December 2003; Co-Chair of the International ITG/IEEE Workshop on Smart Antennas, Duisburg, Germany, April 2005; and also General Co-Chair of the First IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Puerto Vallarta, Mexico, December 2005. He is the Technical Co-Chair of the Fourth IEEE Sensor Array and Multichannel Signal Processing Workshop, Waltham, MA, June 2006.