Vector Perturbation Precoding Revisited

Johannes Maurer, Joakim Jaldén, *Member, IEEE*, Dominik Seethaler, and Gerald Matz, *Senior Member, IEEE*

***Abstract*—We consider the downlink of a multiuser system with multiple antennas at the base station. Vector perturbation (VP) precoding is a promising variant of transmit-side channel inversion allowing the users to detect their data in a simple, noncooperative manner. VP precoding has so far been developed and analyzed under the assumptions that the transmitter has perfect channel state information (CSI) and that the receivers know perfectly a channel-dependent transmit power normalization factor and have infinite dynamic range. We demonstrate that the violation of any of these idealizing assumptions degrades the performance of VP significantly and almost always results in an error floor. Motivated by this observation, we propose a novel scheme which we term transmit outage precoding (TOP). With TOP, the transmitter uses a prearranged power scaling known by the receivers and refrains from transmitting when channel conditions are poor. We further show how to augment TOP and conventional VP to deal with a finite dynamic range at the receiver. The performance of the proposed schemes under various levels of transmit CSI is studied in terms of a theoretical diversity analysis and illustrated by numerical results.**

***Index Terms*—Diversity, imperfect channel state information, precoding.**

I. Introduction

We consider a wireless communication scenario in which a base station is in possession of channel state information (CSI) and uses multiple antennas to transmit simultaneously to multiple users with one antenna each. In such a multiple-input multiple-output (MIMO) broadcast scenario, dirty-paper coding [1] is a capacity-achieving strategy [2]–[4]. However, our interest here is in vector perturbation (VP) precoding [5]–[7] since it is a scheme with much lower complexity. VP precoding is a promising technique that shifts most of the signal processing to the transmit side and enables the users to detect their data in a simple and noncooperative manner. The basic idea is to perform channel inversion at the base station, preceded by a perturbation of the transmit vector in order to reduce the transmit power. To meet the transmit power constraint, the transmit signal undergoes a channel-dependent power normalization [6] which the receivers have to know. The users recover the transmitted data symbols by scaling the received signal, performing a modulo operation to compensate for the vector perturbation, and quantizing the result to to the nearest constellation point. Throughout the paper, we will refer to this scheme as *conventional VP precoding* or CVP. Our goal here is to identify and resolve several problems that are specific to CVP and hamper its practical implementation. These problems are briefly outlined in the next three subsections. We note that there are other impairments (e.g., timing and frequency offsets, I/Q imbalance) which we do not address here since they affect any communication system and are not specific to CVP.

Manuscript received January 06, 2010; accepted September 12, 2010. Date of publication October 07, 2010; date of current version December 17, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Anna Scaglione. This work was supported by the STREP project MASCOT (IST-026905) within the Sixth Framework Programme of the European Commission and by the FWF project "Information Networks" (S10606) within the National Research Network SISE. The material in this paper was presented at the IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2008), at the IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP 2009), and at the Asilomar Conference on Signals, Systems, and Computers, 2008.

J. Maurer and G. Matz are with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, A-1040 Wien, Austria (e-mail: jmaurer@nt.tuwien.ac.at; gmatz@nt.tuwien.ac.at).

J. Jaldén was with the Vienna University of Technology, A-1040 Wien, Austria. He is now with the ACCESS Linnaeus Center, KTH Signal Processing Lab, Royal Institute of Technology, Stockhom, Sweden (e-mail: jalden@kth.se).

D. Seethaler was with the Communication Technology Lab, ETH Zurich, Zurich, Switzerland and is now with RobArt, 4020 Linz, Austria (e-mail: dominik.seethaler@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2084083

A. Power Normalization

In order to satisfy the transmit power constraint, the transmitter in CVP must apply a channel- and data-dependent transmit power normalization. The corresponding power scaling factor must be known to the receivers in order to be able to recover the transmitted data. Under a long-term average transmit power constraint in the ergodic regime this is no issue: with increasing transmission block length the power scaling factor here converges to a fixed limit [6] which can be assumed to be known by the receivers. However, in a quasi-static fading regime or under a short-term power constraint, the power scaling factor must be communicated to the receivers prior to data transmission. The transmission of the power scaling factor obviously cannot be performed with CVP itself and hence necessitates the implementation of a second transmission mode which is undesirable.

We address this problem by proposing in Section III a novel VP precoding scheme referred to as *transmit outage precoding* (TOP). TOP is a modification of CVP that uses a transmit power scaling factor which is agreed upon in advance and hence known to the receivers. In some time slots, the prearranged power scaling would lead to a violation of the power constraint, i.e., the transmit power is not sufficient to invert the channel; this situation is referred to as *transmit outage*. Since for a transmit outage it is likely that the data cannot be recovered anyways, TOP refrains from transmitting in such cases. We will show that in spite of intentionally discarding some data, TOP is able to achieve a performance similar to that of genie-aided (i.e., unrealizable) CVP (see Sections III, IV, and VI).

B. Dynamic Range

The vector perturbation underlying CVP and TOP results in an infinitely large symbol constellation in which each data vector is represented by an equivalence class. The actual constellation point from the equivalence class to be transmitted is chosen to minimize the unnormalized transmit power. The use of an infinitely large constellation imposes unrealistic requirements on the dynamic range of the transceiver chain (a similar problem is encountered in Tomlinson-Harashima precoding [8]). Specifically, a finite dynamic range causes an error floor in CVP and TOP. We propose to resolve this problem by using a restricted set of perturbation vectors (see Section V). This modification, referred to as *restricted vector perturbation* (RVP) in what follows, is applicable to CVP and TOP.

C. Diversity Order and Imperfect CSI

The accuracy of the CSI at the transmitter is crucial for the performance of any precoding scheme. Understanding its impact analytically is thus important for system design [9]. For the MIMO broadcast channel, it has been shown that the sum capacity at high signal-to-noise ratio (SNR) deteriorates significantly if the transmitter has imperfect CSI that does not improve with SNR [10]. Robust transceiver designs [11], [12] are an attempt to maintain a certain performance even if the CSI is affected by errors.

As an analytical contribution, we completely characterize the diversity order of CVP and TOP in quasi-static fading under a variety of CSI conditions, both for unrestricted perturbations (Section IV) and for RVP (Section V-A). These results show that with increasing SNR the CSI mismatch must vanish at least as fast as the reciprocal SNR in order for any of the VP schemes to achieve the same diversity as under perfect CSI. A result in a similar spirit showed that in order to achieve the full multiplexing gain with zero-forcing precoding in an ergodic setting, the channel estimation error must scale as the inverse of SNR [13]. The effect of imperfect CSI on CVP and on spatial Tomlinson-Harashima precoding (which can be viewed as approximation to CVP) was investigated in [14] and [15], [16], respectively. Different from those works, our primary aim herein is to obtain closed form expressions for the diversity of CVP.

D. Paper Outline

The rest of the paper is organized as follows. In Section II we present the system model, discuss the CSI model, and review CVP. TOP is introduced in Section III and the performance of TOP and CVP is discussed in Section IV. RVP is proposed in Section V. In Sections VI and VII, we present simulation results and provide conclusions, respectively. Three appendices offer proofs of the main analytical results.

II. Preliminaries

A. System Model

We consider a multiuser communication system operating in the downlink (cf. [5] and [7]). The base station is equipped with M transmit antennas and each of the $K \leq M$ noncooperative users has a single receive antenna. Transmission happens in blocks of duration N. The symbol vectors transmitted in the time slots $n = 1, \ldots, N$ are denoted as $\mathbf{x}_n \triangleq (x_{n,1} \ldots x_{n,M})^{\mathrm{T}}$ where $x_{n,m}$ is the signal transmitted from the mth antenna in the nth time slot. The multiple-input single-output channels $\mathbf{h}_k$, $k = 1, \ldots, K$, from the base station to the individual users are assumed quasi-static, i.e., $\mathbf{h}_k$ is random but remains constant within a block [17]. In the nth time slot, the kth user receives $y_{n,k} = \mathbf{h}_k^{\mathrm{T}} \mathbf{x}_n + w_{n,k}$; here, $w_{n,k} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_w^2)$ denotes spatially and temporally white complex Gaussian noise. By collecting the receive values of all users in a receive vector $\mathbf{y}_n \triangleq (y_{n,1} \ldots y_{n,K})^{\mathrm{T}}$, the overall channel input-output relation is given by

$$\mathbf{y}_n = \mathbf{H}\mathbf{x}_n + \mathbf{w}_n \tag{1}$$

where $\mathbf{H} \triangleq (\mathbf{h}_1 \ldots \mathbf{h}_K)^{\mathrm{T}}$ and $\mathbf{w}_n \triangleq (w_{n,1} \ldots w_{n,K})^{\mathrm{T}}$. For the performance analysis and numerical simulations, we will assume the elements of the $K \times M$ channel matrix $\mathbf{H}$ to be i.i.d. Rayleigh distributed, i.e., $\mathrm{vec}(\mathbf{H}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I})$.

We impose a per-block *short-term* power constraint (cf. [18] and [19]) that requires

$$\frac{1}{N} \sum_{n=1}^{N} \|\mathbf{x}_n\|^2 \leq P. \tag{2}$$

This is different from the less restrictive *long-term* average power constraint[1] $\mathrm{E}\left\{\|\mathbf{x}_n\|^2\right\} \leq P$ (see [5]–[7]), which can be viewed as limiting case of (2) obtained for $N \to \infty$. The *instantaneous* (peak) power constraint $\|\mathbf{x}_n\|^2 \leq P$ is obtained as a special case of (2) with $N = 1$. This instantaneous power constraint is different from the per-antenna power constraint $|x_{n,k}|^2 \leq P/M$ which is addressed in [20] (see also the conclusions section). Based on the power constraint (2), we define the nominal SNR as

$$\rho \triangleq \frac{P}{\sigma_w^2}. \tag{3}$$

B. CSI Model

We assume that the transmitter (base station) is in possession of imperfect CSI, given by the "noisy" channel matrix

$$\hat{\mathbf{H}} = \mathbf{H} + \mathbf{E}. \tag{4}$$

The CSI accuracy is characterized by the error matrix $\mathbf{E}$ which is assumed to be independent of $\mathbf{H}$. We model $\mathbf{E}$ as complex Gaussian with i.i.d. elements, i.e.

$$\mathrm{vec}(\mathbf{E}) \sim \mathcal{N}_{\mathbb{C}}(0, \eta\mathbf{I}) \quad \text{with} \quad \eta \triangleq \beta\rho^{-\alpha}. \tag{5}$$

With this model, the error variance η depends on the nominal SNR ρ unless $\alpha = 0$. The parameters $\alpha \geq 0$ and $\beta > 0$ are used to capture a variety of scenarios. In particular, perfect CSI

[1]Here, $\mathrm{E}\{\cdot\}$ denotes expectation.

is reobtained in the limit $\alpha \longrightarrow \infty$. Two special cases of (5) which are of particular interest are described next.

1) Reciprocal Channels: In reciprocal systems (e.g., time division duplex), uplink and downlink channel are identical [21]. Here, the downlink channel can be estimated using pilots sent over the uplink channel and the CSI error matrix $\mathbf{E}$ models the channel estimation error which depends on the noise level at the base station and on the pilot power. If the pilot power increases proportionally to P, the channel estimation error scales inversely proportional with increasing nominal SNR; this case is captured by (5) with $\alpha = 1$.

2) CSI Feedback: For nonreciprocal systems, the channel matrix is usually estimated using pilot transmissions in the downlink; a quantized version of the channel estimate is then sent back to the base station via a dedicated feedback channel. In such a scenario, the CSI mismatch will be dominated by errors resulting from quantization and from feedback delay, which causes the CSI at the base station to be outdated if the channel coherence time is smaller than the feedback delay [22]. The channel coherence time and often also the quantizer resolution do not depend on the nominal SNR. This situation is modeled by (5) with $\alpha = 0$.

C. Review of CVP

We next rephrase CVP as introduced in [5]–[7] for the case of imperfect CSI.

1) Base-Station Processing: At each time instant n, the base station transmits a length-K symbol vector $\mathbf{s}_n \triangleq (s_{n,1} \dots s_{n,K})^{\mathrm{T}}$ whose elements are independent, uniformly distributed over the symbol alphabet $\mathcal{A}$, and normalized such that $\mathrm{E}\left\{|s_{n,k}|^2\right\} = 1$. The transmit vector in (1) is then obtained as [6]

$$\mathbf{x}_n = \sqrt{\frac{P}{\Gamma}}\, \hat{\mathbf{H}}^{\dagger}(\mathbf{s}_n + \tau \mathbf{z}_n^{\star}). \tag{6}$$

Here, Γ is a real-valued scalar factor used for transmit power normalization; $\hat{\mathbf{H}}^{\dagger} = \hat{\mathbf{H}}^{\mathrm{H}}(\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathrm{H}})^{-1}$ denotes the right pseudo inverse of $\hat{\mathbf{H}}$ used to preequalize the channel; $\mathbf{z}_n^{\star} \in \mathbb{G}^K$ is a perturbation vector whose elements are Gaussian integers[2]; and τ is a real-valued translation parameter chosen such that $\mathcal{A} + \tau\mathbb{G}$ forms an extended symbol alphabet consisting of nonoverlapping copies of $\mathcal{A}$ (cf. [6]). The idea underlying (6) is to preequalize the channel by precoding with the pseudo inverse and to perturb the symbol vector so that $\mathbf{s}_n + \tau\mathbf{z}_n^{\star}$ is better matched to the channel. This can be viewed as communicating $\mathbf{s}_n$ by some other vector in the equivalence class $\mathbf{s}_n + \tau\mathbb{G}^K$. We note that zero-forcing (ZF) precoding [5] corresponds to (6) without perturbation, i.e., $\mathbf{z}_n^{\star} = \mathbf{0}$.

The perturbation vector $\mathbf{z}_n^{\star}$ is obtained according to [6]

$$\mathbf{z}_n^{\star} = \arg \min_{\mathbf{z} \in \mathbb{G}^K} \left\| \hat{\mathbf{H}}^{\dagger}(\mathbf{s}_n + \tau\mathbf{z}) \right\|^2. \tag{7}$$

This particular choice of $\mathbf{z}_n^{\star}$ is further explained in Section II-D. The minimization problem in (7) can be efficiently solved using the sphere decoding algorithm [6], [23] (in this context also referred to as sphere encoding [6]). Even though the complexity of the sphere decoder is exponential in K [24], it performs favorably for moderate K.

The power normalization factor Γ is chosen according to

$$\Gamma = \frac{1}{N}\sum_{n=1}^{N} \Gamma_n, \quad \text{with} \quad \Gamma_n \triangleq \|\hat{\mathbf{H}}^{\dagger}(\mathbf{s}_n + \tau\mathbf{z}_n^{\star})\|^2. \tag{8}$$

With this choice of Γ, (2) is satisfied with equality.

2) Receiver Processing: Under the assumption of perfect CSI ($\hat{\mathbf{H}} = \mathbf{H}$) the receive signal of user k in the nth time-slot is given by

$$y_{n,k} = \sqrt{\frac{P}{\Gamma}}(s_{n,k} + \tau z_{n,k}^{\star}) + w_{n,k} \tag{9}$$

where $z_{n,k}^{\star}$ denotes the kth component of $\mathbf{z}_n^{\star}$. The receiver can thus detect $s_{n,k}$ by multiplying the received signal by $\sqrt{\Gamma/P}$, followed by a modulo-τ operation to remove $z_{n,k}^{\star}$ and quantization to the symbol alphabet $\mathcal{A}$ (cf. [6]).

We assume throughout that the receiver acts similarly in the case of imperfect CSI. In particular, the kth receiver scales the received signals according to

$$r_{n,k} \triangleq \mu\, y_{n,k} \tag{10}$$

(the factor μ will be made explicit in Section II-D) and obtains an estimate of $s_{n,k}$ according to

$$\hat{s}_{n,k} = \mathcal{Q}\left\{\mathcal{M}_{\tau}\left\{r_{n,k}\right\}\right\}. \tag{11}$$

Here, $\mathcal{M}_{\tau} : \mathbb{C} \mapsto [-\tau/2, \tau/2] + j[-\tau/2, \tau/2]$ denotes the modulo operation and $\mathcal{Q} : \mathbb{C} \mapsto \mathcal{A}$ denotes quantization to the closest symbol. The detector in (11) has very low complexity and does not require user cooperation, both of which are major advantages of VP precoding.

3) Regularized Precoding: Regularizing the pseudo inverse in (6) provides a slight performance improvement in CVP [5], [25]. However, as is apparent from [25, eq. (23)], such a regularization will not eliminate the need to communicate a channel and data dependent power scaling factor nor will it limit the dynamic range of the perturbed transmit symbols. For these reasons, and in order not to obscure our main arguments, we only consider nonregularized precoding as described earlier.

D. Performance of CVP

In order to assess the performance of CVP under imperfect CSI, we consider $r_{n,k}$ in (10) conditioned on $\hat{\mathbf{H}}$ and $\mathbf{s}_n$. Since $\hat{\mathbf{H}} = \mathbf{H} + \mathbf{E}$ where $\mathbf{H}$ and $\mathbf{E}$ are statistically independent Gaussian matrices, $\mathbf{H}$ and $\hat{\mathbf{H}}$ are jointly Gaussian. Conditioned on $\hat{\mathbf{H}}$, $\mathbf{H}$ has a Gaussian distribution with mean $\hat{\mathbf{H}}/(1+\eta)$ and statistically independent elements of variance $\eta/(1+\eta)$ (cf. [26, Th. 10.2]). We can therefore write[3]

$$\mathbf{H} = \frac{1}{1+\eta}\hat{\mathbf{H}} + \mathbf{\Psi} \tag{12}$$

where $\mathrm{vec}(\mathbf{\Psi}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \eta\mathbf{I}/(1+\eta))$ is independent of $\hat{\mathbf{H}}$.

[2]The set of Gaussian integers $\mathbb{G} = \mathbb{Z} + j\,\mathbb{Z}$ comprises all complex numbers with integer real and imaginary parts.

[3]In [27] the expression $\mathbf{H} = \hat{\mathbf{H}} - \mathbf{E}$ was used in place of (12) based on the incorrect assumption that $\mathbf{E}$ and $\hat{\mathbf{H}}$ are independent. However, this mistake does not affect the results of the diversity analysis presented in [27].

Since $\mathbf{r}_n = \mu\,\mathbf{y}_n$ (cf. (10)), inserting (6) into (1) and using (12) results in

$$\begin{aligned}\mathbf{r}_n &= \mu\sqrt{\frac{P}{\Gamma}}\mathbf{H}\hat{\mathbf{H}}^{\dagger}(\mathbf{s}_n+\tau\mathbf{z}_n^{\star})+\mu\,\mathbf{w}_n\\ &= \mu\sqrt{\frac{P}{\Gamma}}\left(\frac{1}{1+\eta}\hat{\mathbf{H}}+\mathbf{\Psi}\right)\hat{\mathbf{H}}^{\dagger}(\mathbf{s}_n+\tau\mathbf{z}_n^{\star})+\mu\,\mathbf{w}_n\\ &= \frac{\mu}{1+\eta}\sqrt{\frac{P}{\Gamma}}(\mathbf{s}_n+\tau\mathbf{z}_n^{\star})\\ &\quad+\mu\sqrt{\frac{P}{\Gamma}}\mathbf{\Psi}\hat{\mathbf{H}}^{\dagger}(\mathbf{s}_n+\tau\mathbf{z}_n^{\star})+\mu\,\mathbf{w}_n.\end{aligned}$$

By choosing the scaling factor as

$$\mu = (1+\eta)\sqrt{\frac{\Gamma}{P}}, \tag{13}$$

the multiplicative factor in the signal part $\mathbf{s}_n+\tau\mathbf{z}_n^{\star}$ is removed and we obtain

$$\mathbf{r}_n = \mathbf{s}_n+\tau\mathbf{z}_n^{\star}+\mathbf{v}_n. \tag{14}$$

Here, the interference-plus-noise term is given by

$$\mathbf{v}_n \triangleq \underbrace{(1+\eta)\mathbf{\Psi}\hat{\mathbf{H}}^{\dagger}(\mathbf{s}_n+\tau\mathbf{z}_n^{\star})}_{\text{multiuser interference}}+\underbrace{(1+\eta)\sqrt{\frac{\Gamma}{P}}\,\mathbf{w}_n}_{\text{noise}}. \tag{15}$$

Throughout the rest of the paper, we assume that μ is chosen according to (13)[4]. We note that in the case of perfect CSI (i.e., $\eta=0$) the normalization factor becomes $\mu=\sqrt{\Gamma/P}$ [cf. (9)] and the multiuser interference term vanishes.

Based on (15), it can be shown that the distribution of $\mathbf{v}_n$ conditioned on $\hat{\mathbf{H}}$ and $\mathbf{s}_n$ is i.i.d. zero mean Gaussian, i.e.

$$f(\mathbf{v}_n|\hat{\mathbf{H}},\mathbf{s}_n) = \frac{1}{(\pi\nu)^K}\exp\left(\nu^{-1}\|\mathbf{v}_n\|^2\right)$$

with the interference-plus-noise power ν given by

$$\nu = \nu_{\mathrm{CVP}} \triangleq \underbrace{(1+\eta)\eta\,\Gamma_n}_{\text{interference}}+\underbrace{(1+\eta)^2\,\Gamma\rho^{-1}}_{\text{noise}} \tag{16}$$

where Γ_n and Γ are defined in (8) and ρ is the nominal SNR in (3).

A common approach (see [28]) to assess the high SNR behavior and the diversity order of a system is to consider the probability that the noise-plus-interference power at the detector input exceed a given threshold. The argument[5] is that detection errors are unlikely when $\nu \ll 1$ and likely when $\nu \gg 1$. From (16) it follows that $\nu \gg 1$ either if $\Gamma_n \gg \eta^{-1}$ (in which case the system is interference limited) or if $\Gamma \gg \rho$ (in which case the system is noise limited). When $\Gamma_n \ll \eta^{-1}$ and $\Gamma \ll \rho$, noise and interference are small and the detector is likely to produce correct symbol decisions.

The above argument shows that CVP becomes sensitive to CSI imperfections whenever the unnormalized instantaneous transmit power Γ_n is large. Similarly, the rescaling of the receive signal amplifies the noise if Γ is large. These observations serve as a basis for our proposed precoding strategy. However, before discussing this in detail we reconsider the critical issues of CVP mentioned in the introduction.

E. Critical Issues With CVP

1) Power Normalization: As made apparent by (13), the receive signal scaling (10) requires knowledge of the power normalization factor Γ. Since Γ depends on the channel realization and on the data, it must be communicated to the receivers by the base station via an auxiliary transmission. Since CVP presupposes knowledge of Γ at the receivers, it cannot be used for broadcasting Γ to the receivers. Therefore, the feedforward of Γ requires an additional transmission scheme, which is undesirable. In addition, the feedforward transmission requires a quantization of Γ which—regardless of the quantization codebook—introduces an error floor [29]. In case of large block lengths, pilot-aided or blind estimation of the power scaling is a conceivable alternative to feedforward transmission but suffers from the same limitations (error floor). To alleviate these problems, the precoding scheme proposed in Section III uses a power normalization factor that is agreed upon in advance, thereby obviating the need for auxiliary feedforward transmissions or estimation of the power normalization.

2) Dynamic Range: Another problem with CVP is its large dynamic range at the receivers. Specifically, the desired part $s_{n,k}+\tau z_{n,k}^{\star}$ of the scaled receive signal $r_{n,k}$ [cf. (10) and (14)] can have arbitrarily large magnitude due to the perturbation vector $z_{n,k}\in\mathbb{G}$. This is clearly a problem for practical implementations of CVP, where amplifiers, A/D converters, DSPs, etc. have a limited dynamic range. This problem is not unique to CVP and is seen in other precoding schemes such as Tomlinson-Harashima precoding [8].

We will show in Section V-B that CVP indeed occasionally uses arbitrarily large perturbations $z_{n,k}^{\star}$. Since CVP implementations with finite dynamic range incur errors when $|s_{n,k}+\tau z_{n,k}^{\star}|$ is large (due to clipping etc.), an undesired error floor results again. In Section V, we propose a VP precoder that uses a limited set of perturbation vectors and avoids an error floor.

III. Transmit Outage Precoding

A. Basic Idea

In order to avoid the need for (error-prone) feedforward transmission or estimation of the channel- and data-dependent normalization factor Γ, we propose a new precoding strategy termed transmit outage precoding (TOP) (see also [30]). TOP is a modification of CVP where the transmit signal is given by

$$\mathbf{x}_n = \begin{cases}\sqrt{\frac{P}{\gamma}}\,\hat{\mathbf{H}}^{\dagger}(\mathbf{s}_n+\tau\mathbf{z}_n^{\star}) & \text{for } \Gamma\le\gamma,\\ \mathbf{0} & \text{for } \Gamma>\gamma,\end{cases} \tag{17}$$

with Γ as defined in (8); here, γ is a design parameter that is agreed upon in advance and hence independent of channel and data and known in advance by the receivers. The precoding strategy in (17) is straightforwardly shown to satisfy the short-term power constraint (2). In particular, TOP will always use

[4]The inclusion of $(1+\eta)$ in (13) removes a bias which would otherwise be present in (14) due to the imperfect CSI.

[5]For the diversity order, the argument can be made mathematically rigorous as outlined in [28, Ch. 3]. More details are given in Section IV.

less transmit power than CVP. The receiver processing is assumed to be the same as with CVP, with the important difference that Γ is replaced by γ, resulting in a receive signal scaling factor of

$$\mu = (1+\eta)\sqrt{\frac{\gamma}{P}}.$$

We next explain the rationale underlying (17).

1) Outage Case: The case $\Gamma > \gamma$, i.e., large average unnormalized transmit power, corresponds to unfavorable channel realizations (large channel condition number). Here, TOP declares a *transmit outage* and discards the data by sending nothing (i.e., $\mathbf{x}_n = \mathbf{0}$). Since the receivers are unaware of the transmit outage, they attempt to detect that data, which is highly likely to result in incorrect symbol decisions (the receive signal consists of noise only). While seemingly detrimental to the performance of the proposed scheme, this event must be viewed in relation to what will happen with CVP. In particular, when Γ is large either $\Gamma \gg \rho$ or $\Gamma_n \gg \eta^{-1}$ [note that $\Gamma_n \geq \Gamma$ for some n, cf. (8)] so that according to Section II-D detection errors are highly probable with CVP as well. Therefore, the fact that TOP refrains from transmitting whenever Γ is large does not seriously affect the system performance.

2) Non-Outage Case: Under favorable channel conditions we have $\Gamma \leq \gamma$. Here, the VP-precoded data is transmitted using the prearranged power normalization γ. The average transmit power within a block in this case is $1/N\sum_{n=1}^{N}\|\mathbf{x}_n\|^2 = P\Gamma/\gamma \leq P$, thus satisfying the transmit power constraint (2). A similar derivation as in Section II-D shows that the noise-plus-interference power at the input of the detectors in this case equals

$$\nu = \nu_{\mathrm{TOP}} \triangleq \underbrace{(1+\eta)\eta\,\Gamma_n}_{\text{interference}} + \underbrace{(1+\eta)^2\gamma\rho^{-1}}_{\text{noise}}. \tag{18}$$

The inference part equals that of CVP which is natural as the signal-to-self-interference power ratio is independent of the transmit power. The power of the noise part is however larger than for CVP. Still, as long as $\gamma \gg \rho$, the noise component will contribute only marginally to the overall interference-plus-noise power.

The arguments above will be made rigorous in Section IV by showing that TOP has the same diversity order as CVP, provided that γ is chosen appropriately. We thus next discuss the optimal choice of the power threshold γ.

B. Threshold Optimization

Let $\mathcal{O}_\gamma$ denote the outage event that $\Gamma > \gamma$ and $\bar{\mathcal{O}}_\gamma$ denote its complement. The choice of γ influences the transmit outage probability $\Pr\{\mathcal{O}_\gamma\}$, but it also influence the noise-plus-interference power ν_{TOP} [cf. (18)] and thus the error probability in the non-outage case. A small outage probability requires large γ, whereas small non-outage error probability requires small γ. Thus, the overall performance of TOP is a tradeoff between two opposite effects.

The optimal choice of γ in the sense of minimizing the bit error probability is

$$\gamma^\star = \arg\min_{\gamma\in\mathbb{R}^+} \Pr\{\mathcal{E}_b, \gamma\} \tag{19}$$

where $\Pr\{\mathcal{E}_b,\gamma\}$ denotes the uncoded bit error probability for a particular γ. The bit error probability can be written as

$$\begin{aligned}\Pr\{\mathcal{E}_b,\gamma\} &= \Pr\{\mathcal{E}_b|\mathcal{O}_\gamma\}\Pr\{\mathcal{O}_\gamma\} \\ &\quad + \Pr\{\mathcal{E}_b|\bar{\mathcal{O}}_\gamma\}\Pr\{\bar{\mathcal{O}}_\gamma\} \qquad (20)\\ &\leq \frac{1}{2}\Pr\{\mathcal{O}_\gamma\} + \Pr\{\mathcal{E}_b|\bar{\mathcal{O}}_\gamma\} \qquad (21)\end{aligned}$$

where the second line follows from $\Pr\{\bar{\mathcal{O}}_\gamma\} \leq 1$ and the fact that in the outage case where $\mathbf{y}_n = \mathbf{w}_n$ the receivers will guess only half of the bits correctly, i.e., $\Pr\{\mathcal{E}_b|\mathcal{O}_\gamma\} = 1/2$. Since in practice we are interested in small bit error rates, it follows that $\Pr\{\mathcal{O}_\gamma\}$ must be small or, equivalently, $\Pr\{\bar{\mathcal{O}}_\gamma\} \approx 1$, thus implying that the upper bound (21) is tight (at least for γ close to $\gamma^\star$). The two terms in (21) are depicted in Fig. 1 along with the bit error probability, highlighting the tradeoff between $\Pr\{\mathcal{O}_\gamma\}$ and $\Pr\{\mathcal{E}_b|\bar{\mathcal{O}}\}$.

As part of the system design, finding $\gamma^\star$ requires Monte-Carlo simulation to determine $\Pr\{\mathcal{E}_b,\gamma\}$. Note that $\gamma^\star$ is a function of the nominal SNR ρ and grows with increasing ρ (see Fig. 1). In the low-SNR regime, detection is likely to fail, and therefore a larger number of outages can be tolerated, corresponding to small $\gamma^\star$. Since detection is likely to be correct in the high-SNR regime, only few outages are acceptable and hence $\gamma^\star$ is large. As an alternative, in Appendix A we provide an approximation of $\Pr\{\mathcal{E}_b,\gamma\}$ that is valid for any threshold γ, CSI accuracy η, and SNR ρ. This approximation allows for predicting and optimizing the performance of TOP quickly, without the need to perform Monte-Carlo simulations repeatedly anew.

C. Complexity Reduction

One of the advantages of VP precoding is that most of the computational complexity is shifted to the base station, whereas the receiver complexity is extremely low. The base station complexity in turn is dominated by the search for the optimal perturbation vector [cf. (7)] that minimizes the unnormalized transmit power Γ_n. As noted previously, the most promising algorithm for this purpose is the sphere decoding algorithm [23], in this context also referred to as *sphere encoding* [6].

In the case of an instantaneous power constraint ($N = 1$), there is an interesting complexity reduction for TOP. A promising search strategy of the transmitter is to find an arbitrary vector $\mathbf{z}_n \in \mathbb{G}$ for which

$$\left\|\hat{\mathbf{H}}^\dagger(\mathbf{s}_n + \tau\mathbf{z}_n)\right\|^2 \leq \gamma$$

or to establish that no such vector exists (the outage case). The sphere decoder solves (7) by enumerating all perturbation vectors that satisfy [23]

$$\left\|\hat{\mathbf{H}}^\dagger(\mathbf{s}_n + \tau\mathbf{z}_n)\right\|^2 \leq R^2 \tag{22}$$

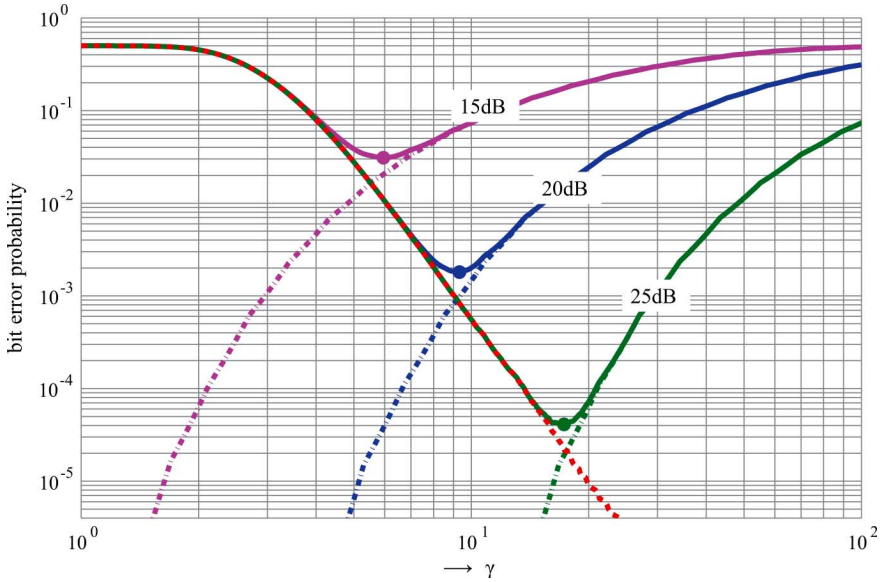


Fig. 1. Bit error probability $\Pr\{\mathcal{E}_b, \gamma\}$ (solid), outage term $\Pr\{\mathcal{O}_\gamma\}/2$ (dashed), and non-outage error probability $\Pr\{\mathcal{E}_b|\bar{\mathcal{O}}_\gamma\}$ (dash-dotted) versus threshold γ for a 6×6 system with 4-QAM in i.i.d. Rayleigh fading with perfect CSI and SNR $\rho \in \{15, 20, 25\}$ dB. The minima of $\Pr\{\mathcal{E}_b, \gamma\}$ are marked with filled circles.

where R is the initial search radius that needs to be chosen suitably. In the context of TOP, the optimal initial search radius is *a priori* known to be $R = \sqrt{\gamma}$. If the sphere decoder doesn't find a perturbation vector satisfying (22) with this choice of R, a transmit outage is declared. In addition, the sphere decoder may terminate early as soon as the first admissible perturbation vector satisfying (22) is found. If the sphere decoder is implemented according to the Schnorr-Euchner strategy [23], [31], early termination and the choice $R = \sqrt{\gamma}$ provide a significant complexity reduction without noticeably increasing the probability of error. This is particularly interesting for perfect transmit-side CSI ($\eta = 0$), where the proposed search strategy does not imply any performance penalty at all. This may be seen by noting that the noise power equals $\nu_{\mathrm{TOP}} = \gamma\rho^{-1}$, independently of the perturbation vector $\mathbf{z}_n^\star$ in the case of perfect CSI and no outage. We note that for imperfect CSI the transmit power resulting with this reduced-complexity implementation is in general larger than for TOP using $\mathbf{z}_n^\star$ according to (7), but still smaller than for CVP. Therefore, the increased interference power in (18) typically leads to a small performance degradation.

Fig. 2 confirms that the complexity of the reduced-complexity implementation is indeed much smaller than for the standard sphere decoder (the graph depicts the mean and the 99% percentile of the number of nodes visited by the sphere decoder versus the SNR ρ). This example uses $\gamma = \gamma^\star$ [cf. (19)] and $\eta = 0$. In contrast to the standard sphere decoder, the number of nodes visited by the reduced-complexity implementation depends on the SNR ρ since $\gamma^\star$ depends on ρ. The complexity reduction tends to be particularly pronounced for low SNR (here, the initial search radius $\gamma^\star$ is small which allows the sphere decoder to aggressively prune the search tree) and for high SNR (here, an admissible perturbation vector is quickly found since the search radius is large).

For imperfect CSI ($\eta > 0$), simulation results in Section VI show that the performance degradation due to increased interference power is often small. Similar strategies for reducing the complexity are possible for block lengths $N > 1$, although the complexity savings are not as large in these cases.

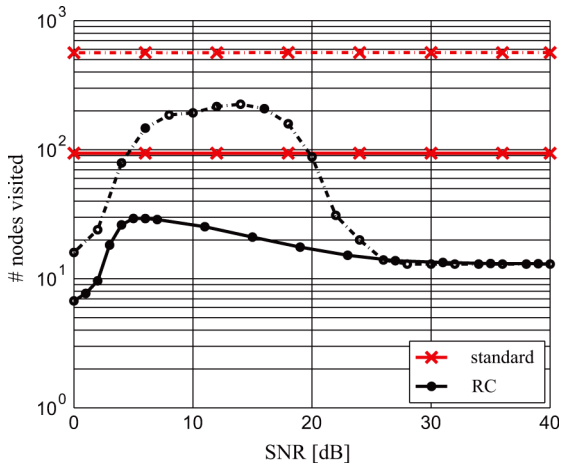


Fig. 2. Mean (solid) and 99% percentile (dash-dotted) of the number of nodes visited by the standard sphere decoder ('×') and by the reduced-complexity implementation ('•') versus the SNR ρ for a 6×6 system with 4-QAM in i.i.d. Rayleigh fading.

IV. Diversity Analysis

In this section we characterize the diversity order of TOP and CVP under the short-term power constraint. In particular, we demonstrate that TOP achieves the same diversity as CVP. For technical reasons, we impose the condition that $0 \notin \mathcal{A}$, which is satisfied for most common symbol alphabets.

A. Main Diversity Result

The diversity order is defined as

$$d \triangleq -\lim_{\rho \to \infty} \frac{\log \Pr\{\mathcal{E}_b\}}{\log \rho} \tag{23}$$

where $\Pr\{\mathcal{E}_b\}$ denotes the bit error probability. We note that the diversity order d remains unchanged if the bit error probability in (23) is replaced by the symbol or block error probability. For reasons of symmetry, the same is true if the diversity order is defined on a per-user basis (all users have identical individual error probabilities). We recall that the variance of the elements of the CSI error matrix $\mathbf{E} = \hat{\mathbf{H}} - \mathbf{H}$ is given by $\eta = \beta\rho^{-\alpha}$ [cf. (5)].

Theorem 1: The diversity orders of CVP and of TOP with γ chosen according to (19) are given by

$$d_{\mathrm{CVP}} = d_{\mathrm{TOP}} = \min(\alpha, 1)M$$

where α characterizes the CSI accuracy.

This result reveals that the critical parameter determining diversity order of CVP and TOP is the exponent α in the CSI error variance. Specifically, imperfect CSI degrades the diversity order unless $\alpha \geq 1$.

Proof of Theorem 1

To prove Theorem 1, note that in the outage case TOP cannot perform better than CVP and in the non-outage case (i.e., when $\Gamma \leq \gamma$) (16) and (18) imply $\nu_{\mathrm{CVP}} \leq \nu_{\mathrm{TOP}}$. Thus, the error probability of CVP must be smaller than that of TOP, regardless of how γ is chosen. This in turn implies that

$$d_{\mathrm{CVP}} \geq d_{\mathrm{TOP}}. \tag{24}$$

In order to establish Theorem 1, we thus only need to show that $\min(\alpha, 1)M$ constitutes a lower bound on the diversity of TOP and an upper bound on the diversity of CVP. The average unnormalized transmit power Γ will turn out to be a key quantity in our proof (see also Section II-D). We hence provide the following lemma whose proof is given in Appendix A.

Lemma 1: The tail probability of the average unnormalized transmit power Γ defined in (8) decays according to

$$-\lim_{\rho \longrightarrow \infty} \frac{\log \Pr\{\Gamma > c\rho^{\kappa}\}}{\log \rho} = \kappa M$$

for any constants $\kappa \geq 0$ and $c > 0$.

For notational convenience, we will in the following use the $\doteq$ notation (cf. [28]) where $a(\rho) \doteq b(\rho)$ is short for

$$-\lim_{\rho \longrightarrow \infty} \frac{\log a(\rho)}{\log \rho} = -\lim_{\rho \longrightarrow \infty} \frac{\log b(\rho)}{\log \rho}. \tag{25}$$

The symbols $\dot{\geq}$ and $\dot{\leq}$ are defined similarly. In this notation, the statement of Lemma 1 becomes $\Pr\{\Gamma > c\,\rho^{\kappa}\} \doteq \rho^{-\kappa M}$. We use the notation $a(\rho) \doteq \rho^{-\infty}$ for the case where $a(\rho)$ vanishes exponentially fast with increasing ρ.

1) Top Diversity Lower Bound: To show that $\min(\alpha, 1)M$ is a lower bound for the diversity of TOP, we need to establish a corresponding upper bound on the error probability $\Pr\{\mathcal{E}_b, \gamma^{\star}\}$. We note that for $\alpha = 0$, the statement $d_{\mathrm{TOP}} \geq \min(\alpha, 1)M = 0$ is trivially true and we can thus restrict to $\alpha > 0$. From the definition of $\gamma^{\star}$ in (19) we have

$$\Pr\{\mathcal{E}_b, \gamma^{\star}\} \leq \Pr\{\mathcal{E}_b, \gamma\} \tag{26}$$

for arbitrary $\gamma > 0$. Furthermore

$$\begin{aligned}\Pr\{\mathcal{E}_b, \gamma\} &= \Pr\{\mathcal{E}_b|\mathcal{O}_\gamma\}\Pr\{\mathcal{O}_\gamma\} \\ &\quad + \Pr\{\mathcal{E}_b|\bar{\mathcal{O}}_\gamma\}\Pr\{\bar{\mathcal{O}}_\gamma\} \\ &\leq \Pr\{\mathcal{O}_\gamma\} + \Pr\{\mathcal{E}_b|\bar{\mathcal{O}}_\gamma\}\end{aligned} \tag{27}$$

where we used $\Pr\{\mathcal{E}_b|\mathcal{O}_\gamma\} \leq 1$ and $\Pr\{\bar{\mathcal{O}}_\gamma\} \leq 1$. By setting $\gamma = \rho^{\kappa}$ with $\kappa \geq 0$, Lemma 1 implies

$$\Pr\{\mathcal{O}_\gamma\} = \Pr\{\Gamma \geq \rho^{\kappa}\} \doteq \rho^{-\kappa M}. \tag{28}$$

In order to bound the second term in (27) we consider the noise-plus-interference power of TOP [cf. (18)] under the condition imposed by $\bar{\mathcal{O}}_\gamma$, i.e., $\Gamma \leq \gamma = \rho^{\kappa}$. Since $\Gamma_n \leq N\Gamma$ and $\eta = \beta\rho^{-\alpha}$, it follows for $\rho \geq 1$ that

$$\begin{aligned}\nu_{\mathrm{TOP}} &= (1+\eta)\eta\Gamma_n + (1+\eta)^2\gamma\rho^{-1} \\ &\leq c_1\rho^{\kappa-\alpha} + c_2\rho^{\kappa-1} \\ &\leq c_3\rho^{\kappa-\min(1,\alpha)} \triangleq \bar{\nu}_{\mathrm{TOP}}\end{aligned}$$

for constants c_1, c_2, and c_3. Due to the Gaussianity of the interference and noise we may upper bound the second term in (27) according to

$$\Pr\{\mathcal{E}_b|\bar{\mathcal{O}}_\gamma\} \leq a_0\Pr\left\{\mathcal{E}_s|\bar{\mathcal{O}}_\gamma\right\} \leq a_1 Q\left(\frac{a_2}{\sqrt{\bar{\nu}_{\mathrm{TOP}}}}\right)$$

where $Q(\cdot)$ denotes the Q-function, and a_0, a_1, and a_2 denote positive constellation-dependent constants. Here, we used the fact that the bit error probability can be bounded by the symbol error probability, e.g., $\Pr\{\mathcal{E}_b|\bar{\mathcal{O}}_\gamma\} \leq a_0\Pr\left\{\mathcal{E}_s|\bar{\mathcal{O}}_\gamma\right\}$. Since the tails of the Gaussian distribution decay exponentially, it follows that

$$\Pr\{\mathcal{E}_b|\bar{\mathcal{O}}\} \doteq \rho^{-\infty} \tag{29}$$

whenever $\bar{\nu}_{\mathrm{TOP}} \longrightarrow 0$ as $\rho \longrightarrow \infty$, i.e., when $\kappa < \min(\alpha, 1)$.

Combining (26)–(29), it follows that for any $\kappa < \min(\alpha, 1)$

$$\Pr\{\mathcal{E}_b, \gamma^{\star}\} \dot{\leq} \rho^{-\kappa M}.$$

Considering the limit $\kappa \longrightarrow \min(\alpha, 1)$, we obtain

$$\Pr\{\mathcal{E}_b, \gamma^{\star}\} \dot{\leq} \rho^{-\min(\alpha,1)M}$$

which confirms the lower bound

$$d_{\mathrm{TOP}} \geq \min(\alpha, 1)M. \tag{30}$$

2) CVP Diversity Upper Bound: To establish $\min(\alpha, 1)M$ as upper bound for the CVP diversity, we need a lower bound for the error probability. Specifically, we have for any γ

$$\Pr\{\mathcal{E}_b\} \geq \Pr\{\mathcal{E}_b|\mathcal{O}_\gamma\}\Pr\{\mathcal{O}_\gamma\}. \tag{31}$$

With $\gamma = \rho^{\kappa}$ where $\kappa \geq 0$, the event $\mathcal{O}_\gamma$ corresponds to $\Gamma > \gamma = \rho^{\kappa}$, in which case the noise-plus-interference power of CVP in (16) satisfies

$$\begin{aligned}\nu_{\mathrm{CVP}} &= (1+\eta)\eta\Gamma_n + (1+\eta)^2\Gamma\rho^{-1} \\ &\geq \eta\Gamma + \Gamma\rho^{-1} > \beta\rho^{-\alpha}\rho^{\kappa} + \rho^{\kappa}\rho^{-1} \\ &\geq c_4\rho^{\kappa-\min(\alpha,1)} \triangleq \underline{\nu}_{\mathrm{CVP}}\end{aligned} \tag{32}$$

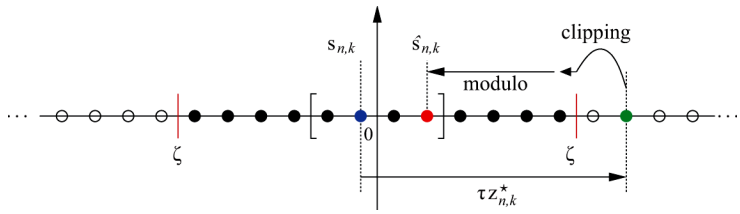


Fig. 3. Illustration of error mechanism resulting from receive-side clipping.

for some n, where $c_4 = \min(1, \beta)$. The first inequality follows from $\Gamma_n \geq \Gamma$ for some n and $\eta \geq 0$. For any $\kappa > \min(1, \alpha)$, (32) shows that $\underline{\nu}_{\mathrm{CVP}} \longrightarrow \infty$ as $\rho \longrightarrow \infty$, which in turn implies that $\Pr\{\mathcal{E}_b | \mathcal{O}_\gamma\}$ is bounded away from zero. As a consequence, we conclude with (31) and Lemma 1 that

$$\Pr\{\mathcal{E}_b\} \dot{\geq} \Pr\{\mathcal{O}_\gamma\} \doteq \rho^{-\kappa M}$$

for any $\kappa > \min(1, \alpha)$. Considering the limit as $\kappa \longrightarrow \min(\alpha, 1)$, it follows that

$$\Pr\{\mathcal{E}_b\} \dot{\geq} \rho^{-\min(\alpha,1)M}$$

and consequently

$$d_{\mathrm{CVP}} \leq \min(\alpha, 1) M. \tag{33}$$

Combining (24), (30), and (33) establishes Theorem 1.

V. Restricted VP Precoding

As noted in Section II-E, CVP and TOP involve unbounded signal magnitudes at the receivers. We address this problem in the following.

A. Restricted Vector Perturbation

We first propose a straightforward method to limit the dynamic range that we will refer to as restricted vector perturbation (RVP). RVP is similar in spirit to the method developed in [8] in the context of Tomlinson-Harashima precoding. The idea is to consider a predetermined *finite* set $\mathcal{Z}^K \subset \mathbb{G}^K$ of perturbation vectors at the transmitter [cf. (7)]

$$\bar{\mathbf{z}}_n^{\star} = \arg \min_{\mathbf{z} \in \mathcal{Z}^K} \left\| \hat{\mathbf{H}}^{\dagger} (\mathbf{s}_n + \tau \mathbf{z}) \right\|^2. \tag{34}$$

We note that (34) can again be solved efficiently using sphere encoding; the finiteness of $\mathcal{Z}^K$ can be additionally exploited to reduce the complexity of the algorithm. The set $\mathcal{Z}^K$ is chosen in advance so that any combination $s_{n,k} + \tau \bar{z}^{\star}_{n,k}$ of transmit symbol and perturbation remains well within the dynamic range of the receivers. We emphasize that both CVP and TOP are straightforwardly augmented using RVP. For CVP, only the power scaling needs to be modified accordingly as

$$\bar{\Gamma} \triangleq \frac{1}{N} \sum_{n=1}^{N} \bar{\Gamma}_n \quad \text{where} \quad \bar{\Gamma}_n \triangleq \|\hat{\mathbf{H}}^{\dagger} (\mathbf{s}_n + \tau \bar{\mathbf{z}}_n^{\star})\|^2. \tag{35}$$

With TOP, an outage is declared for $\bar{\Gamma} \leq \bar{\gamma}$, with suitably chosen $\bar{\gamma}$. For both CVP and TOP, the receive side processing remains unchanged.

The reduced dynamic range of RVP is bought by an increase of the noise-plus-interference power that results from $\bar{\Gamma} \geq \Gamma$. This leads to an increased error probability, especially when $|\mathcal{Z}|$ is small. In fact, as stated by the next theorem, the restriction of the perturbation set results in a loss of the entire diversity advantage of VP precoding over ZF precoding.

Theorem 2: The diversity order of CVP and TOP using RVP with arbitrary finite $\mathcal{Z}$ according to (34) equals

$$d_{\mathrm{RVP}} = \min(\alpha, 1)(M - K + 1)$$

where α characterizes the CSI accuracy as in (5).

The proof of this theorem is completely analogous to Section IV-B, with the only difference being that Lemma 1 is replaced with the following result, whose proof can be found in Appendix C.

Lemma 2: The tail probability of the average unnormalized transmit power $\bar{\Gamma}$ defined in (35) decays according to

$$-\lim_{\rho \longrightarrow \infty} \frac{\log \Pr\{\bar{\Gamma} > c\, \rho^{\kappa}\}}{\log \rho} = \kappa(M - K + 1)$$

for any constants $\kappa \geq 0$ and $c > 0$.

Since Theorem 2 holds for arbitrary finite $\mathcal{Z}$, the diversity order of RVP is the same as that of ZF precoding, which corresponds to $\mathcal{Z} = \{0\}$. Although this result may seem very pessimistic, our simulation results reveal that for practical SNR the performance degradation often is small and the diversity result kicks in only at quite high SNR (see Section VI). Furthermore, we note that RVP has the potential to improve the peak-to-average power ratio (cf. [8]).

B. Impact of Limited Dynamic Range on CVP and TOP

We next argue that with limited dynamic range both CVP and TOP suffer from an error floor if RVP is not used. Let $\mathcal{D} = [-\zeta, \zeta] + j[-\zeta, \zeta]$ with $\zeta > 0$ model the finite dynamic range at the receivers. If the receive signal $r_{n,k}$ exceeds the receiver dynamic range in the sense that $r_{n,k} \notin \mathcal{D}$, a detection error will occur with nonzero probability even without noise. We provide a simple and intuitive example illustrating why this is the case.

Example: For simplicity, we assume a real-valued signal $r_{n,k}$, a 4-PAM alphabet $\mathcal{A} = \{-3, -1, 1, 3\}$, and $\zeta = \tau(2k + 1)/2$ with $\tau = 8$. The receiver truncates the signal $r_{n,k}$ in (10) to the range $[-\zeta, \zeta]$ prior to detection (see Fig. 3). Whenever $r_{n,k} \notin [-\zeta, \zeta]$, the signal is clipped to $\pm\zeta$, depending on the sign of $r_{n,k}$, leading to the decisions $\hat{s}_{n,k} = \pm 3$. Hence, if the transmitted symbol is in the interior of the constellation (i.e., $s_{n,k} \in \{-1, 1\}$), the clipping will always cause in an incorrect decision. A similar argument that identifies constellation points for which clipping always causes a detection error can be made for arbitrary constellations and clipping thresholds ζ.

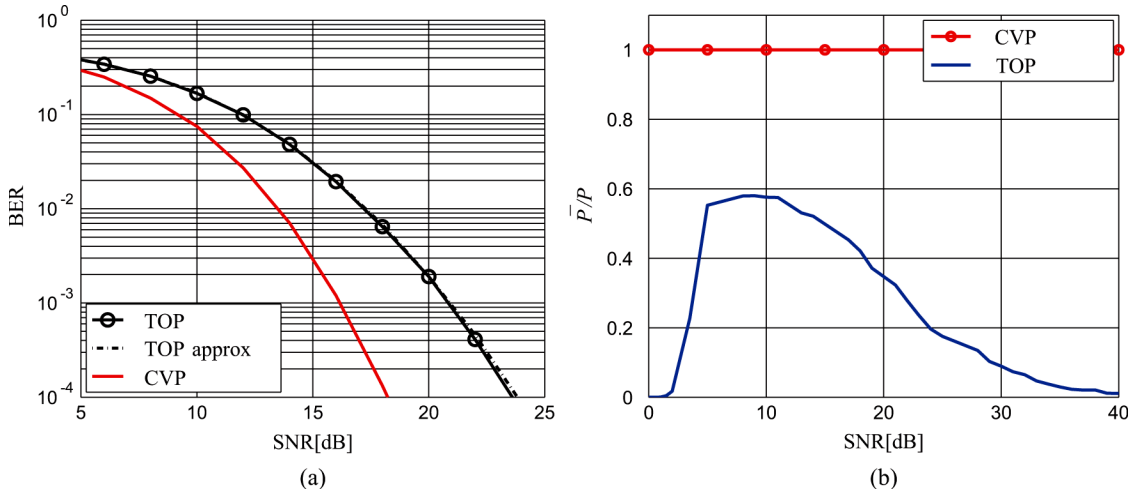


Fig. 4. (a) BER versus SNR for TOP, TOP with approximately optimal threshold, and CVP. (b) Average transmit power normalized by the available power P versus SNR for TOP and CVP ($M = K = 6$, 4-QAM, and perfect CSI).

To establish the existence of an error floor, we show by way of contradiction that $s_{n,k}+\tau z^{\star}_{n,k} \notin \mathcal{D}$ with nonzero probability, i.e., the receive signal will occasionally exceed the dynamic range even in the absence of noise and thus a certain fraction of errors will occur even when SNR tends to infinity. To this end, assume that the statement is false, i.e., $s_{n,k} + \tau z^{\star}_{n,k} \in \mathcal{D}$ for all $\mathbf{z}^{\star}_n$ obtained according to (7). This means, that the perturbation vectors $\mathbf{z}^{\star}_n$ have to come from a finite set $\mathcal{Z}$ and we would have obtained the same results with RVP based on this specific $\mathcal{Z}$ so that $\bar{\Gamma} = \Gamma$. In this case, both Lemma 1 and Lemma 2 would simultaneously apply which is impossible. Thus, the assumption $s_{n,k}+\tau z^{\star}_{n,k} \in \mathcal{D}$ cannot be true, i.e., we have $s_{n,k}+\tau z^{\star}_{n,k} \notin \mathcal{D}$ with nonzero probability for any $\zeta < \infty$. In other words, the unrestricted VP precoders will occasionally use arbitrarily large perturbations.

VI. Simulation Results

In order to illustrate the theoretical results obtained in the previous sections and to assess the performance of TOP and RVP, we next present numerical simulation results. TOP results are based on Monte Carlo evaluations of the optimal threshold $\gamma^{\star}$ (cf. Section III-B). Unless stated otherwise, we consider $M = K = 6$, a 4-QAM symbol alphabet, and block length $N = 12$. Furthermore, all CVP results were obtained with the unrealistic assumption that the power normalization factor Γ is perfectly known at the receivers.

A. Perfect CSI

Fig. 4(a) shows the bit error rate (BER) versus nominal SNR $\rho = P/\sigma_w^2$ for TOP under the assumption of perfect CSI and infinite dynamic range. The BER of CVP with perfect knowledge of Γ at the receivers is plotted as an ultimate (but unrealizable) benchmark. We observe that TOP has a 5 dB performance loss compared to CVP. However, we stress that TOP uses a prearranged power normalization factor and uses only a fraction of the transmit power P, all of which is in stark contrast to CVP. For comparison, we also show TOP with the approximately optimal threshold $\tilde{\gamma}^{\star}$ (labeled 'TOP approx') that is obtained according to Appendix A. It is seen that using $\tilde{\gamma}^{\star}$ instead of $\gamma^{\star}$ induces virtually no performance penalty. The power savings achieved with TOP are illustrated in Fig. 4(b) which depicts the average transmit power normalized by the available power P versus the SNR. It is seen that here TOP consumes at most 60% (and for a wide SNR range significantly less) of the average transmit power spent by CVP.

B. Imperfect CSI

Fig. 5(a) depicts BER versus SNR for TOP and CVP for the case of imperfect CSI ($\alpha = 1$, $\beta = 10$) and a block length of $N = 12$. Results for TOP using the reduced-complexity sphere encoder (labeled 'TOP-RC') for $N = 1$ are also included (here, the power constraint is more stringent). Both versions of TOP are shown with optimal threshold and approximately optimal threshold. In this imperfect CSI scenario, all implementations of TOP achieve a performance very close to that of CVP. This improvement is due to the fact that the overall system now is to a large extent interference limited and the interference power in TOP and CVP is identical [see (16) and (18)]. The figure further confirms that $\tilde{\gamma}^{\star}$ performs as well as $\gamma^{\star}$ and that the reduced-complexity implementation of TOP results only in a small performance loss.

The behavior of both CVP and TOP changes dramatically for imperfect CSI with $\alpha = 0$ and $\beta = 0.03$, see Fig. 5(b). Here, we observe an error floor for both schemes in accordance with Theorem 1 ($d_{\mathrm{TOP}} = d_{\mathrm{CVP}} = 0$). This confirms that poor CSI accuracy can have a disastrous effect on the performance of any VP precoding scheme.

C. TOP/CVP With RVP

We next illustrate the performance of VP precoding for finite dynamic range. The clipping threshold was $\zeta = 3\tau/2$ which entails the restricted perturbation set $\mathcal{Z} = \{-1, 0, 1\}+j\{-1, 0, 1\}$ of size $|\mathcal{Z}| = 9$.

For $M = K = 2$ and perfect CSI, Fig. 6(a) shows BER versus nominal SNR for plain CVP and CVP with RVP (here, $|\mathcal{Z}^K| = 81$). For comparison, we show CVP with infinite dynamic range (labeled 'CVP inf') as ultimate benchmark and ZF precoding as opposite extreme with no perturbation.

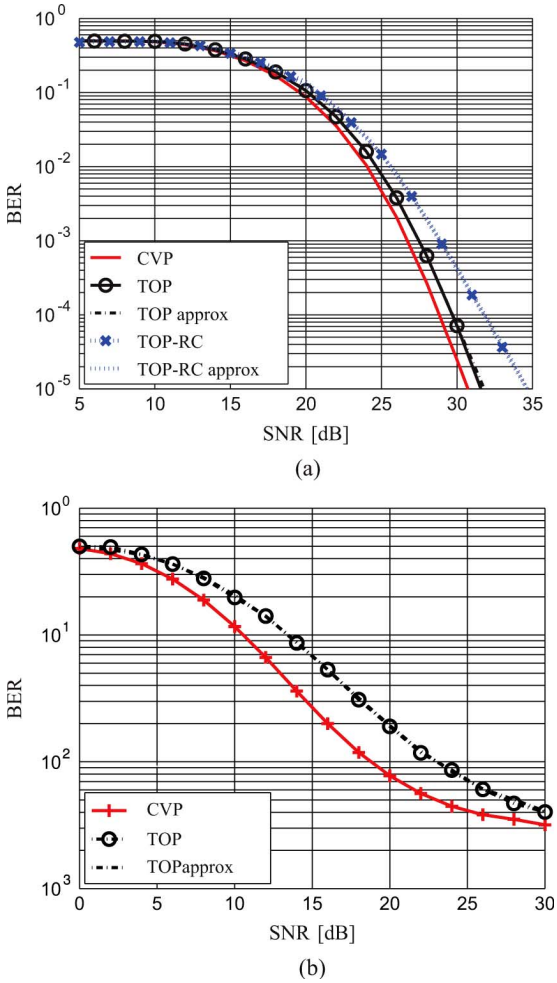


Fig. 5. BER versus SNR for TOP, TOP with approximately optimal threshold, and CVP for imperfect CSI and block length $N = 12$ ($M = K = 6$, 4-QAM). (a) $\alpha = 1$ and $\beta = 10$. (b) $\alpha = 0$ and $\beta = 0.03$. For comparison, (a) also shows reduced-complexity TOP ('TOP-RC') with optimal and approximately optimal threshold and block length $N = 1$.

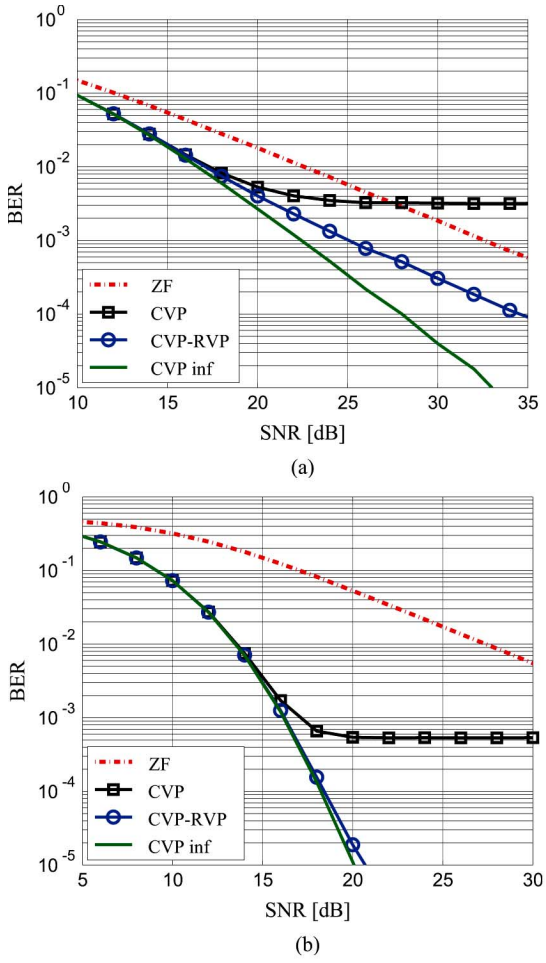


Fig. 6. BER versus SNR for CVP, CVP with RVP, and ZF precoding, all with finite dynamic range, and for CVP with infinite dynamic range (4-QAM, perfect CSI). (a) $M = K = 2$. (b) $M = K = 6$.

All CVP schemes are based on perfect knowledge of Γ at the receivers. It is seen that CVP with RVP (diversity order 1) clearly outperforms plain CVP which—in agreement with Section V-B—features an error floor (diversity order 0) and eventually performs even worse than ZF precoding.

Similar observations apply to the case $M = K = 6$, shown in Fig. 6(b). The main difference to the case $M = K = 2$ is the fact that for the BER range shown, CVP with RVP now performs within 0.5 dB of CVP with infinite dynamic range, i.e., the diversity order $d_{\mathrm{RVP}} = 1$ kicks in only at much higher SNR. This can be partly attributed to the fact that for $K = 6$ the size of the perturbation set is significantly larger (i.e., $|\mathcal{Z}^K| = 531\,441$).

Finally, Fig. 7 illustrates the performance of TOP using RVP and of CVP under finite dynamic range and imperfect CSI ($\alpha = 1$, $\beta = 10$) for $M = K = 6$. For comparison, CVP with infinite dynamic range is also shown. At small SNR, TOP performs close to CVP. At high SNR, TOP clearly outperforms CVP under finite dynamic range (which again shows an error floor) and approaches the performance achieved by CVP with infinite dynamic range. We conclude that for practically relevant SNRs, TOP-RVP performs almost as well as CVP without requiring the unrealistic assumptions of perfectly known power normalization and infinite dynamic range.

VII. Conclusion

We considered vector perturbation precoding for a quasi-static MIMO downlink multiuser system. We identified a number of practical problems of vector perturbation schemes presented so far in the literature and proposed corresponding solutions. In particular, we introduced a novel scheme termed transmit outage precoding (TOP) which does not require the receivers to know a channel-dependent transmit power normalization factor. This was accomplished by allowing transmit outages in case of bad channel realizations, where power is saved by discarding the data to be transmitted. We further modified TOP and conventional vector perturbation precoding (CVP) by restricting the perturbation set; this avoids that the precoding schemes break down in practical implementations with finite dynamic range at the receivers. A theoretical analysis of the diversity order was provided and numerical results obtained by simulations illustrated the performance of our precoding scheme under a variety of CSI models.

Determining whether there is a straight-line approximation for the error probability of CVP and TOP at high SNR and, if

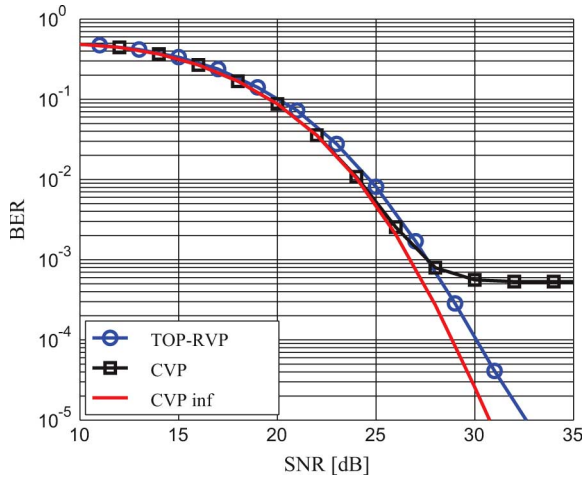


Fig. 7. BER versus SNR for TOP using RVP and for CVP with limited dynamic range and for CVP with infinite dynamic range ($M = K = 6$, 4-QAM, imperfect CSI with $\alpha = 1$, $\beta = 10$).

yes, obtaining the corresponding SNR shift (or coding gain) in addition to the diversity order is a challenging topic for future research since current techniques cannot be straightforwardly applied. For example, the approach from [32] fails here since it requires a polynomial approximation of the distribution of an SNR-independent fading parameter. Unfortunately, with CVP and TOP under imperfect CSI the fading parameter ν in (16) and (18) is SNR-dependent and there is no apparent polynomial approximation for its distribution.

Another interesting research avenue is the modification of our results for the case of a per-antenna power constraint where the ℓ_2 norm in the definition of Γ in (8) is replaced by the ℓ_∞ norm. The optimization of the perturbation vector can here be achieved using the p-sphere decoder [20] instead of the standard ℓ_2 sphere decoder. The diversity results of Theorem 1 hold true also for this case. This follows as a consequence of the following facts: 1) the ℓ_2 norm and the ℓ_∞ norm are equivalent (i.e., their ratios are upper and lower bounded by fixed nonzero constants); 2) the diversity is invariant under constant (i.e., SNR-independent) multiplicative changes in the transmit, noise, and interference powers. However, under a per-antenna power constraint the performance difference between TOP and CVP at finite SNR may not be the same as for the total power constraint.

Appendix A
Approximating the Optimal TOP Threshold

The optimization of the TOP threshold γ with respect to error probability according to (19) in general requires Monte Carlo simulations to obtain $\Pr\{\mathcal{E}_b, \gamma\}$. These simulations need to be repeated for any desired parameter configuration (SNR, CSI accuracy, etc.). In this Appendix, we present an approximation of $\Pr\{\mathcal{E}, \gamma\}$ which is valid for any γ, channel accuracy η, and SNR ρ. This approximation induces a corresponding approximation for the optimal TOP threshold $\gamma^\star$ in (19), which can be easily determined without repeated Monte Carlo simulations. Our starting point is the expression for the bit error probability in (20), i.e., $\Pr\{\mathcal{E}_b, \gamma\} = 1/2 \Pr\{\mathcal{O}_\gamma\} + \Pr\{\mathcal{E}_b|\bar{\mathcal{O}}_\gamma\} \Pr\{\bar{\mathcal{O}}_\gamma\}$. We will next show how to (approximately) obtain the outage probability $\Pr\{\mathcal{O}_\gamma\}$ and the non-outage error probability $\Pr\{\mathcal{E}_b|\bar{\mathcal{O}}_\gamma\}$.

Outage Probability: We first consider the outage probability (equivalently, the complementary cumulative distribution function of the average unnormalized transmit power Γ) for the case of perfect CSI, i.e.,

$$g(\gamma) \triangleq \Pr\{\mathcal{O}_\gamma|\hat{\mathbf{H}} = \mathbf{H}\} = \Pr\{\Gamma > \gamma|\hat{\mathbf{H}} = \mathbf{H}\} = \Pr\left\{\frac{1}{N}\sum_{n=1}^{N} \|\mathbf{H}^\dagger(\mathbf{s}_n + \tau\mathbf{z}_n)\|^2 > \gamma\right\}. \quad (36)$$

The important observation now is that $\hat{\mathbf{H}}$ and $\mathbf{H}$ are identical in distribution up to a factor, i.e., $\hat{\mathbf{H}} \sim \sqrt{1+\eta}\,\mathbf{H}$. This means that the outage probability for the case of imperfect CSI can be simply obtained as [cf. (36) with $\mathbf{H}$ replaced with $1/\sqrt{1+\eta}\hat{\mathbf{H}}$]

$$\Pr\{\mathcal{O}_\gamma\} = g\big((1+\eta)\gamma\big). \quad (37)$$

This means that $g(\gamma)$ needs to be determined only once by Monte Carlo simulations (for $\eta = 0$) and computing the outage probability for arbitrary CSI accuracy requires only one function evaluation according to (37).

We note that the perturbation vector in (36) depends on which precoding scheme is considered, i.e., for TOP we use $\mathbf{z}_n = \mathbf{z}_n^\star$ and for RVP we use $\mathbf{z}_n = \bar{\mathbf{z}}_n^\star$.

Non-Outage Error Probability: The error probability conditioned on no outage can be written as

$$\Pr\{\mathcal{E}_b|\bar{\mathcal{O}}_\gamma\} = \int_0^\infty \Pr\{\mathcal{E}_b|\bar{\mathcal{O}}_\gamma, \nu = x\}\, f_{\nu|\bar{\mathcal{O}}_\gamma}(x|\bar{\mathcal{O}}_\gamma)\, dx \quad (38)$$

Here, $f(\nu|\bar{\mathcal{O}}_\gamma)$ denotes the conditional (conditioned on no outage) probability density function of the interference-plus-noise power ν in a randomly chosen time slot, given by

$$f_{\nu|\bar{\mathcal{O}}_\gamma}(x|\bar{\mathcal{O}}_\gamma) = \frac{d}{dx}\frac{\Pr\{(\nu \le x) \cap \bar{\mathcal{O}}_\gamma\}}{\Pr\{\bar{\mathcal{O}}_\gamma\}}. \quad (39)$$

The probability $\Pr\{(\nu \le x) \cap \bar{\mathcal{O}}_\gamma\}$ is difficult to evaluate since ν and the random variables Γ_n are statistically dependent, which implies that also ν and the no-outage event $\bar{\mathcal{O}}_\gamma = \{\Gamma \le \gamma\}$ are statistically dependent. Therefore, we propose to use the approximation $\Gamma_n \approx \Gamma$ (which is exact iff $N = 1$), i.e.

$$\nu \approx \tilde{\nu} \triangleq (1+\eta)\eta\,\Gamma + (1+\eta)^2\gamma\rho^{-1} \quad (40)$$

which in the non-outage case is bounded as $\tilde{\nu}_L \le \tilde{\nu} \le \tilde{\nu}_\mathrm{U}$ with

$$\tilde{\nu}_L \triangleq (1+\eta)^2\gamma\rho^{-1}, \quad \tilde{\nu}_\mathrm{U} \triangleq (1+\eta)\eta\,\gamma + (1+\eta)^2\gamma\rho^{-1}. \quad (41)$$

With (40), the event $\nu \le x$ is approximately equal to the event $\Gamma \le (x - \tilde{\nu}_L)/((1+\eta)\eta)$ and, hence, (39) can be approximated for $x \in [\tilde{\nu}_L, \tilde{\nu}_\mathrm{U}]$ as

$$f_{\nu|\bar{\mathcal{O}}_\gamma}(x|\bar{\mathcal{O}}_\gamma) \approx \frac{1}{\Pr\{\bar{\mathcal{O}}_\gamma\}}\frac{d}{dx}\Pr\left\{\Gamma \le \frac{x - \tilde{\nu}_L}{(1+\eta)\eta}\right\} \quad (42)$$

$$= \frac{1}{\Pr\{\bar{\mathcal{O}}_\gamma\}}\frac{d}{dx}\left[1 - g\left(\frac{x - \tilde{\nu}_L}{\eta}\right)\right] \quad (43)$$

where in the second line we again use the function $g(\xi)$ defined in (36).

A further benefit of using the approximation (40) is that $\Pr\{\mathcal{E}_b|\bar{\mathcal{O}}_\gamma, \tilde{\nu} = x\} = \Pr\{\mathcal{E}_b|\tilde{\nu} = x\}$ since the non-outage event $\bar{\mathcal{O}}_\gamma = \{\Gamma < \gamma\}$ follows necessarily from $\tilde{\nu} = x \in [\tilde{\nu}_L, \tilde{\nu}_U]$ (see (41)). The function $h(x) \triangleq \Pr\{\mathcal{E}_b|\tilde{\nu} = x\}$ can again be easily determined using a single Monte Carlo simulation, and is independent of the parameters of interest (e.g., CSI accuracy or operating SNR).

Inserting the above intermediate results into (38) leads to the approximation

$$\Pr\{\mathcal{E}_b|\bar{\mathcal{O}}_\gamma\} \approx \frac{1}{\Pr\{\bar{\mathcal{O}}_\gamma\}} m(\gamma, \eta, \rho) \tag{44}$$

with

$$\begin{aligned} m(\gamma, \eta, \rho) &\triangleq \int_{\tilde{\nu}_L}^{\tilde{\nu}_U} h(x) \left[-\frac{d}{dx} g\left(\frac{x - \tilde{\nu}_L}{\eta} \right) \right] dx \\ &= -\int_0^{(1+\eta)\gamma} h(x\eta + \tilde{\nu}_L)\, g'(x)\, dx \end{aligned} \tag{45}$$

which can be obtained by numerical integration. For perfect CSI it can be shown that $m(\gamma, \eta, \rho) = h(\gamma\rho^{-1}) g((1+\eta)\gamma)$ and hence the integration (45) is not required in this case.

Approximate Threshold Optimization: Inserting (37), (44), and (45) into (20) leads to the bit error probability approximation

$$\Pr\{\mathcal{E}_b, \gamma\} \approx \tilde{P}(\gamma) \triangleq \frac{1}{2}\, g\big((1+\eta)\gamma\big) + m(\gamma, \eta, \rho) \tag{46}$$

and the corresponding approximately optimal threshold

$$\tilde{\gamma}^\star = \arg \min_{\gamma \in \mathbb{R}^+} \tilde{P}(\gamma). \tag{47}$$

Once the functions $g(x)$ and $h(x)$ have been obtained by one Monte Carlo simulation each, (46) and (47) can thus be easily computed for arbitrary parameter choices.

Appendix B
Proof of Lemma 1

To prove $\Pr\{\Gamma > c\rho^\kappa\} \doteq \rho^{-\kappa M}$, we show that $\Pr\{\Gamma > c\rho^\kappa\} \dot{\geq} \rho^{-\kappa M}$, and $\Pr\{\Gamma > c\rho^\kappa\} \dot{\leq} \rho^{-\kappa M}$.

Lower Bound: We define $c_{\mathcal{A}} \triangleq \min_{a \in \mathcal{A}} |a|^2$, which is strictly positive ($c_{\mathcal{A}} > 0$) due to our assumption that[6] $0 \notin \mathcal{A}$. By the definition of the pseudo-inverse, it follows that $\hat{\mathbf{h}}_i^{\mathrm{H}} \hat{\mathbf{H}}^\dagger = \mathbf{e}_i^{\mathrm{H}}$, where $\hat{\mathbf{h}}_i^{\mathrm{H}}$ denotes the ith row of $\hat{\mathbf{H}}$ and $\mathbf{e}_i$ is the ith natural basis vector. Thus, for any $i = 1, \ldots, K$, we have

$$c_{\mathcal{A}} \leq |\tilde{s}_i|^2 = |\mathbf{e}_i^{\mathrm{H}} \tilde{\mathbf{s}}|^2 = |\hat{\mathbf{h}}_i^{\mathrm{H}} \hat{\mathbf{H}}^\dagger \tilde{\mathbf{s}}|^2 \leq \|\hat{\mathbf{h}}_i\|^2 \|\hat{\mathbf{H}}^\dagger \tilde{\mathbf{s}}\|^2$$

where $\tilde{s}_i$ denotes the ith element of $\tilde{\mathbf{s}} = \mathbf{s} + \tau\mathbf{z}$. It follows that

$$\Gamma_n = \|\hat{\mathbf{H}}^\dagger(\mathbf{s}_n + \tau \mathbf{z}_n^\star)\|^2 \geq \frac{c_{\mathcal{A}}}{\|\hat{\mathbf{h}}_i\|^2}. \tag{48}$$

[6]We note that a similar but more tedious proof can be given for the case where $0 \in \mathcal{A}$.

Since the lower bound is independent of n, we also have

$$\Gamma \geq \frac{c_{\mathcal{A}}}{\|\hat{\mathbf{h}}_i\|^2}$$

and thus a corresponding lower bound for $\Pr\{\Gamma > c\rho^\kappa\}$:

$$\begin{aligned} \Pr\{\Gamma > c\rho^\kappa\} &\geq \Pr\left\{ \|\hat{\mathbf{h}}_i\|^2 \leq c^{-1} c_{\mathcal{A}} \rho^{-\kappa} \right\} \\ &= \Pr\left\{ \|\mathbf{h}_i\|^2 < \frac{c_{\mathcal{A}} \rho^{-\kappa}}{c(1 + \beta\rho^{-\alpha})} \right\}, \\ &\geq \Pr\left\{ \|\mathbf{h}_i\|^2 < c_{\mathrm{LB}} \rho^{-\kappa} \right\} \end{aligned} \tag{49}$$

for $\rho \geq 1$, where $c_{\mathrm{LB}} = c_{\mathcal{A}} c^{-1} (1+\beta)^{-1}$ and where we used that $\hat{\mathbf{h}} \sim \sqrt{1+\eta}\, \mathbf{h}$ with $\eta = \beta\rho^{-\alpha}$. Since $\|\mathbf{h}_i\|^2$ is χ^2-distributed with $2M$ degrees of freedom, we have (cf. [28])

$$\Pr\{\Gamma > c\rho^\kappa\} \geq \Pr\left\{ \|\mathbf{h}_i\|^2 < c_{\mathrm{LB}} \rho^{-\kappa} \right\} \dot{\geq} \rho^{-\kappa M}. \tag{50}$$

Upper Bound: Inspired by [32], we view $\hat{\mathbf{H}}^\dagger$ as the generator matrix of a K-dimensional lattice in $\mathbb{C}^M$. The *covering radius* of this lattice is defined as [33]

$$\zeta(\hat{\mathbf{H}}^\dagger) \triangleq \max_{\mathbf{a} \in \mathbb{C}^M} \min_{\mathbf{z} \in \mathbb{G}^M} \|\hat{\mathbf{H}}^\dagger(\mathbf{a} - \mathbf{z})\|.$$

The dual lattice is generated by the matrix $\hat{\mathbf{H}}^{\mathrm{H}}$; the length of its shortest lattice vector, also called the *first successive minimum*, equals [33]

$$m_1(\hat{\mathbf{H}}^{\mathrm{H}}) \triangleq \min_{\mathbf{z} \in \mathbb{G}^K \setminus \{\mathbf{0}\}} \|\hat{\mathbf{H}}^{\mathrm{H}} \mathbf{z}\|.$$

Using [34, Th. 2.2], it follows that

$$m_1(\hat{\mathbf{H}}^{\mathrm{H}})\, \zeta(\hat{\mathbf{H}}^\dagger) \leq K. \tag{51}$$

We note that [34, Th. 2.2] applies to real-valued lattices, but that (51) follows directly by considering one complex dimension as two real valued dimensions.

The definition of the covering radius provides the following upper bound on Γ_n:

$$\Gamma_n = \min_{\mathbf{z} \in \mathbb{G}^M} \|\hat{\mathbf{H}}^\dagger(\mathbf{s}_n + \tau\mathbf{z})\|^2 \leq \tau^2 \zeta^2(\hat{\mathbf{H}}^\dagger). \tag{52}$$

This upper bound is again independent of n and hence also applies to Γ; in view of (51), we thus obtain

$$\Gamma \leq \frac{(\tau K)^2}{m_1^2(\hat{\mathbf{H}}^{\mathrm{H}})}. \tag{53}$$

Using this upper bound on Γ, we have

$$\begin{aligned} \Pr\{\Gamma > c\rho^\kappa\} &\leq \Pr\left\{ m_1^2(\mathbf{H}^{\mathrm{H}}) < \frac{c^{-1}}{(\tau K)^2} \frac{\rho^{-\kappa}}{1 + \beta\rho^{-\alpha}} \right\} \\ &\leq \Pr\left\{ m_1^2(\mathbf{H}^{\mathrm{H}}) < c_{\mathrm{UB}} \rho^{-\kappa} \right\} \end{aligned} \tag{54}$$

where $c_{\mathrm{UB}} \triangleq c^{-1}(\tau K)^{-2}$ and where we again used that $\hat{\mathbf{H}} \sim \sqrt{1+\eta}\, \mathbf{H}$. Applying [32, Lemma 3], to (54) shows that

$$\Pr\{\Gamma > c\rho^\kappa\} \leq \Pr\left\{ m_1^2(\mathbf{U}^{\mathrm{H}}) < c_{\mathrm{UB}} \rho^{-\kappa} \right\} \dot{\leq} \rho^{-\kappa M}. \tag{55}$$

APPENDIX C
PROOF OF LEMMA 2

We establish that $\Pr\left\{\bar{\Gamma} > c\rho^{\kappa}\right\} \dot{\geq} \rho^{-\kappa(M-K+1)}$ and $\Pr\left\{\bar{\Gamma} > c\rho^{\kappa}\right\} \dot{\leq} \rho^{-\kappa(M-K+1)}$ which implies the statement in Lemma 2.

Upper Bound: An upper bound on $\bar{\Gamma}$ is obtained as

$$\begin{aligned}\bar{\Gamma} &= \frac{1}{N}\sum_{n=1}^{N} ||\hat{\mathbf{H}}^{\dagger}(\mathbf{s}_n + \tau\bar{\mathbf{z}}_n^{\star})||^2 \\ &\leq \frac{1}{N}\sum_{n=1}^{N} \lambda_{\min}^{-1}||\mathbf{s}_n + \tau\bar{\mathbf{z}}_n^{\star}||^2 \leq \lambda_{\min}^{-1}\bar{c}_1\end{aligned}$$

where in the first inequality we used the smallest eigenvalue of $\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathrm{H}}$, denoted $\lambda_{\min}$, and the second inequality exploits the finiteness of $\mathcal{Z}$ and $\mathcal{A}$. It follows that

$$\begin{aligned}\Pr\{\bar{\Gamma} > c\,\rho^{\kappa}\} &\leq \Pr\left\{\lambda_{\min} \leq c^{-1}\bar{c}_1\rho^{-\kappa}\right\} \\ &\dot{\leq} \rho^{-\kappa(M-K+1)}\end{aligned}$$

where in the second step we used [28, eq. (15)], and the observation that for $\rho \geq 1$ the variance of $\dot{\mathbf{H}}$ is $1+\eta \leq 1+\beta$ where β is independent of the nominal SNR.

Lower Bound: A lower bound on $\bar{\Gamma}$ is given by

$$\begin{aligned}\bar{\Gamma} &= \frac{1}{N}\sum_{n=1}^{N} ||\hat{\mathbf{H}}^{\dagger}(\mathbf{s}_n + \tau\bar{\mathbf{z}}_n^{\star})||^2 \\ &\geq \frac{1}{N}\sum_{n=1}^{N} \lambda_{\min}^{-1}|\mathbf{q}^{\mathrm{H}}(\mathbf{s}_n + \tau\bar{\mathbf{z}}_n^{\star})|^2 \\ &\geq \lambda_{\min}^{-1}\, f(\mathbf{q}) \qquad (56)\end{aligned}$$

where $\mathbf{q}$ denotes the eigenvector of $\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathrm{H}}$ associated to the eigenvalue $\lambda_{\min}$. Furthermore, the SNR-independent random variable $f(\mathbf{q})$ is defined as

$$f(\mathbf{q}) \triangleq \min_{\mathbf{p}\in\mathcal{P}^K} |\mathbf{q}^{\mathrm{H}}\mathbf{p}|^2$$

where $\mathcal{P} \triangleq \mathcal{A} + \tau\mathcal{Z}$. Since $\mathcal{P}$ is finite, the minimum of $f(\mathbf{q})$ is guaranteed to exist and furthermore $f(\mathbf{q}) = 0$ if and only if there is a $\mathbf{p} \in \mathcal{P}$ such that $\mathbf{q}^{\mathrm{H}}\mathbf{p} = 0$. For any given $\mathbf{p}$, this is a zero probability event due to the uniform distribution of $\mathbf{q}$ over the unit sphere [35]. It follows that there is some $\delta > 0$ for which $\Pr\{f(\mathbf{q}) \geq \delta\} > 0$. With (56) we hence obtain

$$\Pr\left\{\bar{\Gamma} > c\,\rho^{\kappa}\right\} \geq \Pr\{f(\mathbf{q}) \geq \delta\}\Pr\left\{\lambda_{\min} < \delta c^{-1}\rho^{-\kappa}\right\} \qquad (57)$$

where we used the statistical independence of $\lambda_{\min}$ and $\mathbf{q}$ [35]. Applying [28, eq. (15)], to (57) finally yields

$$\Pr\{\bar{\Gamma} > c\,\rho^{\kappa}\} \dot{\geq} \rho^{-\kappa(M-K+1)}.$$

ACKNOWLEDGMENT

The authors thank A. Burg for drawing their attention to the dynamic range problem of vector perturbation precoding.

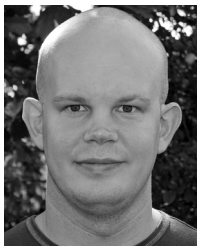
REFERENCES

[1] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
[2] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.
[3] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.
[4] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.
[5] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.
[6] B. M. Hochwald, C. B. Peel, and A. L. 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.
[7] 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.
[8] R. F. H. Fischer, W. H. Gerstacker, and J. B. Huber, "Dynamics limited precoding, shaping, and blind equalization for fast digital transmission over twisted pair lines," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 9, pp. 1622–1633, Dec. 1995.
[9] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, "An introduction to the multiuser MIMO downlink," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 60–67, Oct. 2004.
[10] A. Lapidoth, S. Shamai (Shitz), and M. A. Wigger, "On the capacity of fading MIMO broadcast channels with imperfect transmitter side-information," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Sep. 2005.
[11] N. Vucic and H. Boche, "Robust QoS-constrained optimization of downlink multiuser MISO systems," *IEEE Trans. Signal Process.*, vol. 57, pp. 714–725, Feb. 2009.
[12] F. A. Dietrich and W. Utschick, "Robust Tomlinson-Harashima precoding," in *Proc. IEEE Symp. Pers., Indoor, and Mobile Radio Commun.*, Berlin, Germany, Sep. 2005, pp. 136–140.
[13] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5059, Nov. 2006.
[14] D. J. Ryan, I. B. Collings, V. L. Clarkson, and R. W. Heath, "Performance of vector perturbation multiuser MIMO systems with limited feedback," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2633–2644, Sep. 2009.
[15] S. Bahng, J. Liu, A. Host-Madsen, and X. Wang, "The effects of channel estimation on Tomlinson-Harashima precoding in TDD MIMO systems," in *Proc. IEEE SPAWC-2005*, New York, Jun. 2005.
[16] P. Amihood, E. Masry, L. B. Milstein, and J. G. Proakis, "The effects of channel estimation errors on a nonlinear precoder for multiple antenna downlink channels," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3307–3315, Nov. 2009.
[17] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[18] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1468–1589, Jul. 1999.
[19] B. Hochwald and S. Vishwanath, "Space-time multiple access: Linear growth in sum rate," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Oct. 2002.
[20] F. Boccardi and G. Caire, "The p-sphere encoder: Peak-power reduction by lattice precoding for the MIMO Gaussian broadcast channel," *IEEE Trans. Commun.*, vol. 54, no. 11, pp. 2085–2091, Nov. 2006.
[21] T. L. Marzetta and B. M. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. Signal Process.*, vol. 54, pp. 1268–1278, Apr. 2006.
[22] D. L. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 844–855, Jun. 1999.
[23] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.
[24] J. Jaldén and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Process.*, vol. 53, pp. 1474–1484, Apr. 2005.

[25] D. A. Schmidt, M. Joham, and W. Utschick, "Minimum mean square error vector precoding," *Eur. Trans. Telecommun.*, vol. 19, no. 3, pp. 219–231, Mar. 2007.
[26] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[27] J. Jaldén, J. Maurer, and G. Matz, "On the diversity order of vector perturbation precoding with imperfect channel state information," in *Proc. IEEE SPAWC-2008*, Recife, Brazil, Jul. 2008, pp. 211–215.
[28] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
[29] J. Maurer, J. Jaldén, and G. Matz, "Multithreshold TOP—full-diversity vector perturbation precoding with finite-rate feedforward," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Oct. 2008, pp. 428–432.
[30] J. Maurer, J. Jaldén, and G. Matz, "Transmit outage precoding with imperfect channel state information under an instantaneous power constraint," in *Proc. IEEE SPAWC-2008*, Recife, Brazil, Jul. 2008, pp. 66–70.
[31] C. P. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," *Math. Program.*, vol. 66, no. 1–3, pp. 181–199, Aug. 1994.
[32] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "Communication over MIMO broadcast channels using lattice-basis reduction," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4567–4582, Dec. 2007.
[33] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. Berlin, Heidelberg, New York: Springer, 1988.
[34] W. Banaszczyk, "New bounds in some transference theorems in the geometry of numbers," *Mathematische Annalen*, vol. 296, no. 1, pp. 625–635, Dec. 1993.
[35] A. M. Tulino and S. Verdú, "Random matrices and wireless communications," *Found. Trends in Commun. Inf. Theory*, vol. 1, no. 1, Jun. 2004.

Johannes Maurer received the Dipl.-Ing. and Dr. techn. degrees in electrical/communication engineering from the Vienna University of Technology, Vienna, Austria, in 2005 and 2009, respectively.

From 2005 to 2009, he was a Research and Teaching Assistant with the Institute of Communications and Radio Frequency Engineering, Vienna University of Technology. His research interests are in wireless communications with emphasis on MIMO multiuser signal processing.



Joakim Jaldén (S'02–M'08) received the M.Sc. and Ph.D. degrees in electrical engineering from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2002 and 2007, respectively.

From July 2007 to June 2009, he held a Postdoctoral research position with the Vienna University of Technology, Vienna, Austria. He also studied at Stanford University, CA, from September 2000 to May 2002, and was with ETH, Zürich, Switzerland, as a Visiting Researcher, from August to September 2008. In July 2009, he joined the Signal Processing Lab within the School of Electrical Engineering at KTH, Stockholm, Sweden, as an Assistant Professor.

Dr. Jaldén was awarded the IEEE Signal Processing Society's 2006 Young Author Best Paper Award for his work on MIMO communications and the first prize in the Student Paper Contest at the 2007 International Conference on Acoustics, Speech and Signal Processing (ICASSP). He is also a recipient of the Ingvar Carlsson Award issued in 2009 by the Swedish Foundation for Strategic Research.

Dominik Seethaler received the Dipl.-Ing. and Dr. techn. degrees in electrical/communication engineering from Vienna University of Technology, Vienna, Austria, in 2002 and 2006, respectively.

From 2002 to 2007, he was a Research and Teaching Assistant with the Institute of Communications and Radio Frequency Engineering, Vienna University of Technology. From September 2007 to August 2009, he was a Postdoctoral Researcher with the Communication Technology Laboratory, ETH Zurich, Switzerland. From September 2009 to September 2010, he was freelancing in the area of home robotics. Since October 2010, he has been with RobArt, Linz, Austria.

Gerald Matz (S'95-M'01-SM'07) received the Dipl.-Ing. and Dr. techn. degrees in electrical engineering in 1994 and 2000, respectively, and the Habilitation degree for communication systems in 2004, all from the Vienna University of Technology, Vienna, Austria.

Since 1995, he has been with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, where he currently holds a tenured position as Associate Professor. From March 2004 to February 2005, he was on leave as an Erwin Schrödinger Fellow with the Laboratoire des Signaux et Systèmes, Ecole Supérieure d'Electricité, France. During summer 2007, he was a Guest Researcher with the Communication Theory Lab, ETH Zurich, Switzerland. He has directed or actively participated in several research projects funded by the Austrian Science Fund (FWF), the Vienna Science and Technology Fund (WWTF), and the European Union. He has published more than 130 papers in international journals, conference proceedings, and edited books. He is a coeditor of the book *Wireless Communications over Rapidly Time-Varying Channels* (New York: Academic, 2011). His research interests include wireless communication and sensor networks, statistical signal processing, and information theory.

Prof. Matz serves on the IEEE Signal Processing Society (SPS) Technical Committee on Signal Processing for Communications and Networking and on the IEEE SPS Technical Committee on Signal Processing Theory and Methods. He is an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS (2004–2008) and for the EURASIP journal SIGNAL PROCESSING (2007–2010). He was a Technical Program Co-Chair of EUSIPCO 2004 and has been on the Technical Program Committee of numerous international conferences. In 2006, he received the Kardinal Innitzer Most Promising Young Investigator Award.