Hermitian Precoding for Distributed MIMO Systems

Jianwen Zhang
Dept. of Electronic Engineering
City University of Hong Kong
Hong Kong
jzhang35@student.cityu.edu.hk

Xiaojun Yuan
Institute of Network Coding
Dept. of Information Engineering
Chinese University of Hong Kong
Hong Kong
xjyuan@inc.cuhk.edu.hk

Li Ping
Dept. of Electronic Engineering
City University of Hong Kong
Hong Kong
eeliping@cityu.edu.hk

Abstract—**In this paper, we consider a distributed MIMO communication network in which multiple transmitters cooperatively send common messages to a single receiver. In this scenario, it is usually costly to acquire full channel state information at the transmitters (CSIT), i.e., every transmitter perfectly knows the overall channel state information (CSI) of the network. Hence, we assume individual CSIT (I-CSIT), i.e., each transmitter only knows its own CSI. We propose a novel precoding technique, named *Hermitian* precoding, to enhance the system performance under the constraint of I-CSIT. We show that the proposed scheme can perform close to the system capacity with full CSIT. This reveals that the amount of CSI required at the transmitters can be significantly reduced without considerably compromising performance.**

I. Introduction

Consider a distributive communication system in which multiple transmitters located at different places cooperatively send common messages to a single receiver. This scenario arises, e.g., in the following two applications.

- In a cooperative cellular system, several adjacent base stations (BSs) simultaneously serve a mobile terminal (MT) in the downlink transmission [1]. Such a scheme is referred to as coordinated multi-point (CoMP) [2]-[4] in long term evolution (LTE) standardization documents [5].
- In a parallel relaying system, multiple relays decode and forward (DF) messages from a common source to a common destination [6]-[8].

Intensive work has been focused on this scenario under various assumptions on the availability of channel state information at the transmitters (CSIT). In [7]-[8] no CSIT is assumed and distributed space-time coding is proposed for efficient transmission. In [1]-[5] full CSIT is assumed, i.e., every transmitter perfectly knows all the channel state information (CSI). With full CSIT, beamforming is performed jointly among the transmitters to enhance the system performance significantly.

However, in practice, it is very costly to acquire full CSI at all transmitters. This problem is especially serious in multiple-input multiple-output (MIMO) systems in which multiple antennas are employed in transmission. The resulting overall system may have a very large dimension. Distributing all these coefficients to all transmitters can be a daunting task.

In this paper, we study the situation of individual CSIT (I-CSIT), in which each transmitter has the CSI of its own link but not other's. This setting can significantly decrease the overhead for distributing CSI among transmitters. We develop a *Hermitian* precoding technique for efficient transmission with I-CSIT. We show analytically that the proposed precoding technique is optimal in the sense of the rate maximization under the I-CSIT assumption, provided that the other transmitters are *a priori* fixed to the same precoding structure. Numerical results demonstrate that the proposed scheme with I-CSIT performs close to the system capacity with full CSIT.

II. System Model

A. System Model

We describe the system model for the distributed MIMO system, in which K transmitters send common messages to a single receiver. Each transmitter k, $k = 1, \ldots, K$, is equipped with N antennas and the receiver is equipped with M antennas. The received signal at the receiver can be expressed as

$$\boldsymbol{r} = \sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{r}$ is an M-by-1 received signal vector, $\boldsymbol{H}_k$ is an M-by-N channel transfer matrix for the link between the transmitter k and the receiver, $\boldsymbol{x}_k$ is an N-by-1 signal vector sent by the transmitter k, and $\boldsymbol{n} \sim \mathcal{CN}(0, \sigma^2\boldsymbol{I})$ is an additive white Gaussian noise vector. Moreover, each transmitter k has an individual power constraint of

$$\mathrm{E}\left[\|\boldsymbol{x}_k\|_2^2\right] \le P_k,\ k = 1, \ldots, K, \tag{2}$$

where P_k is the maximum transmission power of transmitter k, $\|\cdot\|_2$ stands for the Euclidean norm of a vector. Perfect CSI at the receiver is always assumed, i.e., the receiver knows all $\boldsymbol{H}_k$.

B. Haar Matrix

Definition 1: A random square matrix $\boldsymbol{U}$ is called a *Haar* matrix if it is uniformly distributed on the set of all the unitary matrices (of the same size as $\boldsymbol{U}$) [9][1].

Property 1: [p.25, 9] A *Haar* matrix $\boldsymbol{U}$ is unitarily invariant, i.e., for any fixed unitary matrix $\boldsymbol{T}$ independent of $\boldsymbol{U}$, the statistical behavior of $\boldsymbol{TU}$ or $\boldsymbol{UT}$ is the same as $\boldsymbol{U}$.

Property 2: [p.25, 9] Consider a random Hermitian matrix $\boldsymbol{B}$ given by $\boldsymbol{B} = \boldsymbol{U\Lambda U}^{\mathrm{H}}$, with $\boldsymbol{U}$ being a *Haar* matrix and $\boldsymbol{\Lambda}$ being a random diagonal matrix. Assume that $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ are independent of each other. Then $\boldsymbol{B}$ is unitarily invariant, i.e., for any unitary matrix $\boldsymbol{T}$ independent of $\boldsymbol{B}$, $\boldsymbol{TBT}^{\mathrm{H}}$ has the same distribution as $\boldsymbol{B}$. Also, $\boldsymbol{B}^*$ has the same distribution as $\boldsymbol{B}$.

[1] A *Haar* matrix is also called an *isotropic* matrix in [10]. More detail about *Haar* matrices can be found in [9], [11].

978-1-4673-2579-0/12/$31.00 ©2012 IEEE

C. Properties of $\{\boldsymbol{H}_k\}$

Now we describe some useful properties of $\{\boldsymbol{H}_k\}$. Throughout this paper, the entries of $\boldsymbol{H}_k$, $k = 1, \ldots, K$, are assumed to be i.i.d. random variables drawn from $\mathcal{CN}(0, \sigma_k^2)$.[2] Let the singular value decomposition (SVD) of $\boldsymbol{H}_k$ be

$$\boldsymbol{H}_k = \boldsymbol{U}_k \boldsymbol{D}_k \boldsymbol{V}_k^{\mathrm{H}}, \quad (3)$$

where $\boldsymbol{U}_k$ and $\boldsymbol{V}_k$ are unitary matrices, and $\boldsymbol{D}_k$ is an M-by-N diagonal matrix[3] with non-negative diagonal elements arranged in the descending order. Then

Property 3: The matrices $\boldsymbol{U}_k$ and $\boldsymbol{D}_k$ are independent.

Property 4: $\boldsymbol{U}_k$ is a *Haar* matrix [9] [11].

Based on Property 4, the properties of *Haar* matrix in Subsection B hold for $\boldsymbol{U}_k$.

III. PROBLEM FORMULATION WITH I-CSIT

In this section, we consider the situation of I-CSIT in which each transmitter k only knows its own channel $\boldsymbol{H}_k$.

A. Modeling of the Transmit Signals

Recall that all the transmitters share the same message to be sent to a single receiver. Yet, the transmitters still have the freedom to use the same codebook or different codebooks in channel coding. Hence the transmitted signals $\{\boldsymbol{x}_k\}$ may be either correlated or uncorrelated. We use $\boldsymbol{c}_0$ to represent the correlated signal component shared by all the transmitters, and $\boldsymbol{c}_k$ to represent the uncorrelated signal component for transmitter k (for $k = 1, \ldots, K$), where $\{\boldsymbol{c}_k\}_{k=0}^{K}$ are M-by-1 random vectors with the entries independently drawn from $\mathcal{CN}(0, 1)$. By definition, $\mathrm{E}[\boldsymbol{c}_k\boldsymbol{c}_k^{\mathrm{H}}] = \boldsymbol{I}$ and $\mathrm{E}[\boldsymbol{c}_k\boldsymbol{c}_j^{\mathrm{H}}] = \boldsymbol{0}$, $\forall k, j = 0, 1, \ldots, K, k \neq j$. Then, the transmitted signal of transmitter k can be expressed as

$$\boldsymbol{x}_k = \boldsymbol{F}_k \boldsymbol{c}_0 + \boldsymbol{G}_k \boldsymbol{c}_k, \quad k = 1, \ldots, K, \quad (4)$$

where $\boldsymbol{F}_k$ and $\boldsymbol{G}_k$ are N-by-M precoding matrices that is adaptive to the channel. With (4), we rewrite (1) as

$$\boldsymbol{r} = \sum_{k=1}^{K} \boldsymbol{H}_k (\boldsymbol{F}_k \boldsymbol{c}_0 + \boldsymbol{G}_k \boldsymbol{c}_k) + \boldsymbol{n}. \quad (5)$$

We focus on the design of $\{\boldsymbol{F}_k\}$ and $\{\boldsymbol{G}_k\}$ in the following.

B. Problem Formulation

We are interested in the following distributive precoder:

$$\boldsymbol{F}_k = f_k(\boldsymbol{H}_k), \text{ and } \boldsymbol{G}_k = g_k(\boldsymbol{H}_k), \ k = 1, \ldots, K \quad (6)$$

where "distributive" means that the precoders of each transmitter k only depends on the local CSI of transmitter k, i.e., $f_k(\cdot)$ and $g_k(\cdot)$ are functions of $\boldsymbol{H}_k$ only (but not functions of $\{\boldsymbol{H}_{k'}, k' \neq k\}$). This ensures that the precoding scheme in (6) can be realized under the I-CSIT assumption.

The distributive precoding functions $\{f_k(\cdot)\}$ and $\{g_k(\cdot)\}$ should be optimized to maximize the average achievable rate of the system in (5) under the individual transmitter power constraints in (2). This problem can be formulated as

$$\max_{\substack{\{f_k(\cdot)\},\\ \{g_k(\cdot)\}}} \mathrm{E}\left[\log\det\left(\boldsymbol{I} + \frac{1}{\sigma^2}\left(\left|\sum_{k=1}^{K}\boldsymbol{H}_k f_k(\boldsymbol{H}_k)\right|^2 + \sum_{k=1}^{K}\left|\boldsymbol{H}_k g_k(\boldsymbol{H}_k)\right|^2\right)\right)\right]$$

$$\text{s.t. } \mathrm{tr}\left\{\left|f_k(\boldsymbol{H}_k)\right|^2 + \left|g_k(\boldsymbol{H}_k)\right|^2\right\} \le P_k, \forall \boldsymbol{H}_k, \ k = 1, 2, \ldots, K \quad (7)$$

where $|\boldsymbol{A}|^2$ is a shorthand of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ for any matrix $\boldsymbol{A}$, the expectation in (7) is taken over the joint distribution of $\{\boldsymbol{H}_k \mid k = 1, \ldots, K\}$, and the power constraint is imposed on every possible realization of $\boldsymbol{H}_k$, $k = 1, \ldots, K$.

C. Locally Optimal Solution

The problem in (7) is difficult to handle in general since jointly optimizing $\{f_k(\cdot)\}$ and $\{g_k(\cdot)\}$ is required. In what follows, we focus on a *locally optimal* solution defined below.

Definition 2: A set of precoding functions $\{f_k(\cdot), g_k(\cdot)\}_{k=1}^{K}$ is *locally optimal* if, $\forall k = 1, \ldots, K$, $\{f_k(\boldsymbol{H}_k), g_k(\boldsymbol{H}_k)\} =$

$$\underset{\{\boldsymbol{F}_k,\boldsymbol{G}_k\}:\,\mathrm{tr}\{\boldsymbol{F}_k\boldsymbol{F}_k^{\mathrm{H}}+\boldsymbol{G}_k\boldsymbol{G}_k^{\mathrm{H}}\}\le P_k}{\arg\max} \mathrm{E}\left[\log\det\left(\boldsymbol{I} + \frac{1}{\sigma^2}\left(\left|\boldsymbol{H}_k\boldsymbol{F}_k + \sum_{k'\neq k}^{K}\boldsymbol{H}_{k'} f_{k'}(\boldsymbol{H}_{k'})\right|^2 + \left|\boldsymbol{H}_k\boldsymbol{G}_k\right|^2 + \sum_{k'\neq k}^{K}\left|\boldsymbol{H}_{k'} g_{k'}(\boldsymbol{H}_{k'})\right|^2\right)\right)\middle|\boldsymbol{H}_k\right] \quad (8)$$

where the expectation is taken over $\{\boldsymbol{H}_{k'} \mid k' = 1, \ldots, K, k' \neq k\}$. In other words, any particular transmitter k deviating from its own precoding strategy leads to performance degradation.

IV. HERMITIAN PRECODING

A. Hermitian Precoding

Recall the SVD of $\boldsymbol{H}_k$ in (3) as $\boldsymbol{H}_k = \boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{V}_k^{\mathrm{H}}$. As $\boldsymbol{U}_k$ and $\boldsymbol{V}_k$ are unitary, the precoders in (6) can always be written in the following form: for any $k = 1, \ldots, K$,

$$\boldsymbol{F}_k = f(\boldsymbol{H}_k) = \boldsymbol{V}_k \boldsymbol{W}_k \boldsymbol{U}_k^{\mathrm{H}} \text{ and } \boldsymbol{G}_k = g(\boldsymbol{H}_k) = \boldsymbol{V}_k \boldsymbol{\Sigma}_k \boldsymbol{U}_k^{\mathrm{H}}, \quad (9)$$

where $\boldsymbol{W}_k$ and $\boldsymbol{\Sigma}_k$ can be arbitrary N-by-M matrices except that they only depend on $\boldsymbol{H}_k$ (due to the I-CSIT assumption). We have the following result with the proof given in Appendix.

Theorem 1: The optimal $\{\boldsymbol{F}_k\}$ and $\{\boldsymbol{G}_k\}$ to (8) are given by (9) with $\{\boldsymbol{W}_k\}$ and $\{\boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_k^{\mathrm{H}}\}$ being real and diagonal matrices.

Theorem 1 gives the optimal structure of the precoders for the problem defined in (8).

B. Further Discussions

With Theorem 1, we still need to optimize $\{\boldsymbol{W}_k\}$ and $\{\boldsymbol{\Sigma}_k\}$ for power allocation based on available CSI. Due to space limitation, we only consider a simple choice of $\boldsymbol{W}_k$ and $\boldsymbol{\Sigma}_k$ as

$$\boldsymbol{W}_k = \sqrt{P_k/M}\,\boldsymbol{I} \text{ and } \boldsymbol{\Sigma}_k = \boldsymbol{0}. \quad (10)$$

More advanced power allocation strategies are discussed in [12]. The reason for $\boldsymbol{\Sigma}_k = \boldsymbol{0}$ is intuitively explained as follows.

Consider the multiple-input single-output (MISO) case (i.e., $N \ge M = 1$). The channels $\{\boldsymbol{H}_k\}$ reduce to vectors $\{\boldsymbol{h}_k\}$ and allow the expression of $\boldsymbol{h}_k = \|\boldsymbol{h}_k\|_2 \boldsymbol{v}_k^{\mathrm{H}}$, where $\boldsymbol{v}_k = \boldsymbol{h}_k^{\mathrm{H}} / \|\boldsymbol{h}_k\|_2$, for $k = 1, \ldots, K$. Then, it is easy to see that the optimal precoding strategy to the problem in (7) is given by

$$\boldsymbol{F}_k = f_k(\boldsymbol{h}_k) = \sqrt{P_k}\,\boldsymbol{v}_k, \text{ and } \boldsymbol{G}_k = g_k(\boldsymbol{h}_k) = \boldsymbol{0}, \ k = 1, \ldots, K. \quad (11)$$

Correspondingly, the received signal can be expressed as

[2] The results in this paper directly hold for other fading channels, provided that the channel directions $\{\boldsymbol{U}_k\}$ are *Haar* matrices.

[3] Here the term diagonal matrix refers to rectangular diagonal matrix with only the $(i, i)^{\text{th}}$ $i = 1, \ldots, \min\{M, N\}$ entries possibly non-zero.

$$r=\left(\sum_{k=1}^{K}\sqrt{P_k}\left\|\boldsymbol{h}_k\right\|_2\right)c_0+n \qquad (12)$$

where r, c_0, and n are respectively the scalar versions of $\boldsymbol{r}$, $\boldsymbol{c}_0$, and $\boldsymbol{n}$. In (12), we see that the signals from different transmitters are coherently superimposed on each other. Based on the above observation, it is reasonable to expect that $\boldsymbol{G}_k = \boldsymbol{0}$ (or $\boldsymbol{\Sigma}_k = \boldsymbol{0}$) is also a good choice for the MIMO cases.

From the above, our proposed precoding strategy is

$$\boldsymbol{F}_k = f_k(\boldsymbol{H}_k) = \boldsymbol{V}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}} \text{ and } \boldsymbol{G}_k = g_k(\boldsymbol{H}_k) = \boldsymbol{0}. \qquad (13)$$

With (13), the received signal in (5) can be rewritten as

$$\boldsymbol{r}=\left(\sum_{k=1}^{K}\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}}\right)\boldsymbol{c}_0+\boldsymbol{n}. \qquad (14)$$

Clearly, each term $\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}}$ in (14) is a positive semi-definite Hermitian matrix. From Weyl Theorem [13], for any positive semi-definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, we have $\boldsymbol{A}$ (or $\boldsymbol{B}$) $\preceq \boldsymbol{A}+\boldsymbol{B}$. Hence, our proposed precoding strategy in (13) increases the eigen-values of the overall equivalent channel, which provides an intuitive explanation of the related gain. This can be seen as the effect of *coherent transmission*.

V. Numerical Results

In this section, numerical results are provided to evaluate the performance of the proposed Hermitian precoding technique in distributed MIMO channels. Some alternative schemes are listed below for comparison. For simplicity of discussion, we always assume $N = M$.

(i) Hermitian Precoding with Equal Power Allocation (HP-EPA): $\boldsymbol{F}_k = \sqrt{P_k/M}\boldsymbol{V}_k\boldsymbol{U}_k^{\mathrm{H}}$ and $\boldsymbol{G}_k = \boldsymbol{0}$.

This is obtained by substituting $\boldsymbol{W}_k = \sqrt{P_k/M}\boldsymbol{I}$ in (13).

(ii) Hermitian Precoding with Channel Inverse (HP-CI):

$$\boldsymbol{F}_k = \beta_k\boldsymbol{V}_k\boldsymbol{D}_k^{-1}\boldsymbol{U}_k^{\mathrm{H}} \text{ and } \boldsymbol{G}_k = \boldsymbol{0}.$$

This is obtained by substituting $\boldsymbol{W}_k = \beta_k\boldsymbol{D}_k^{-1}$ in (13) with β_k a scalar chosen to meet the power constraint.

(iii) No CSIT with fully Correlated Signaling (No-CSIT-CS):

$$\boldsymbol{F}_k = \sqrt{P_k/M}\boldsymbol{I} \text{ and } \boldsymbol{G}_k = \boldsymbol{0}.$$

(iv) No CSIT with Independent Signaling (No-CSIT-IS):

$$\boldsymbol{F}_k = \boldsymbol{0} \text{ and } \boldsymbol{G}_k = \sqrt{P_k/M}\boldsymbol{I}.$$

(v) Individual Water-Filling (IWF):

$$\boldsymbol{F}_k = \boldsymbol{V}_k\boldsymbol{W}_k \text{ and } \boldsymbol{G}_k = \boldsymbol{0}.$$

Here $\boldsymbol{V}_k$ is obtained from the SVD $\boldsymbol{H}_k = \boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{V}_k^{\mathrm{H}}$ and $\boldsymbol{W}_k$ (that is diagonal) is obtained by water-filling over $\boldsymbol{D}_k$. In other words, IWF applies standard water-filling [14] at each individual transmitter.

Note that the only difference between (i) and (ii) is the choice of $\boldsymbol{W}_k$. They both have the Hermitian precoding structure. On the other hand, the choices (iii)–(v) do not involve the Hermitian precoding structure, and thus lack the coherent transmission effect discussed in Section IV.C.

Fig.1 illustrates the performance of various precoding schemes with $K = 3$, $N = 16$, and $M = 8$. It is seen that HP-EPA performs very close to the upper bound obtained by assuming full CSIT (as studied in [3]). This implies that the potential performance loss due to the I-CSIT assumption is marginal. We also see that HP-EPA significantly outperforms other alternatives. Note that the performance of No-CSIT-CS,

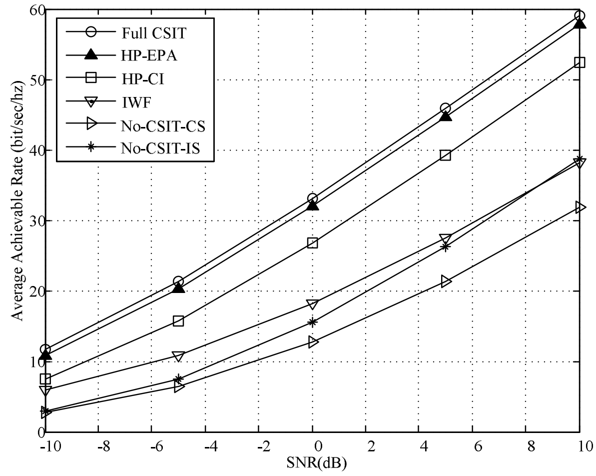

Fig.1 Performance comparison among different precoding schemes with K = 3, M = 8, and N = 16 in Rayleigh-fading distributed MIMO Gaussian channels. SNR = P/σ^2, with $P_k = P$, k = 1, …, 3.

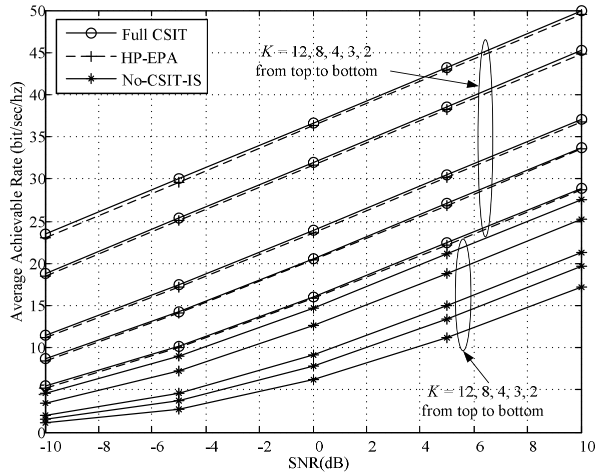

Fig.2 Comparisons of average achievable rate between the Hermitian precoding scheme with EPA and the channel capacity with full-CSIT and No-CSIT-IS. All transmitters have the same power constraint P, and SNR = P/σ^2. M = 4, N = 16. The values of K are marked on the curves.

No-CSIT-IS and IWF is relatively poor in Fig. 1. The reason is that they cannot achieve the coherent effect provided by the Hermitian precoding structure.

Fig.2 studies the impact of the numbers of transmitters on the performance of Hermitian precoding. The numbers of antennas are set to M = 4 and N = 16. The performance curves with full-CSIT and No-CSIT-IS are also included for comparison. Again, the gap between the Hermitian precoding scheme and the full-CSIT capacity is very small. Compared with the curve of No-CSIT-IS, the power gain of the proposed precoder is about 9 dB for K = 2, 12 dB for K = 4, 15 dB for K = 8, and 17 dB for K = 12. Roughly speaking, this power gain increases linearly with K.

VI. Conclusions

In this paper, we consider efficient transmission over distributed MIMO channels with I-CSIT. A Hermitian precoding technique is proposed to exploit the available CSIT. We show analytically that the proposed Hermitian precoding technique is optimal for each transmitter under the I-CSIT assumption when the other transmitters are *a priori* fixed to the same precoding structure. Numerical results show that the proposed Hermitian precoding scheme with I-CSIT suffers marginal performance loss compared with the system with full CSIT in various settings. This indicates that the amount of CSI required at the transmitter sides can be significantly reduced without considerably compromising the system performance.

APPENDIX: PROOF OF THEOREM 1

For notational simplicity, we only consider the case of $K = 2$. The extension of the proof to the case of $K > 2$ is trivial. Since all transmitters experience the same channel distribution, these two transmitters are symmetric. Then, the problem in (8) reduces to

$$\max_{\substack{\{F_1,G_1\}:\\ \mathrm{tr}\{F_1F_1^{\mathrm{H}}+G_1G_1^{\mathrm{H}}\}\le P_1}} \mathrm{E}\left[\log\det\left(\boldsymbol{I}+\frac{1}{\sigma^2}\left(\left|\boldsymbol{H}_1\boldsymbol{F}_1+\boldsymbol{H}_2\boldsymbol{F}_2\right|^2+\left|\boldsymbol{H}_1\boldsymbol{G}_1\right|^2+\left|\boldsymbol{H}_2\boldsymbol{G}_2\right|^2\right)\right)\middle|\boldsymbol{H}_1\right] \quad (15)$$

where $\boldsymbol{F}_2 = f_2(\boldsymbol{H}_2) = \boldsymbol{V}_2\boldsymbol{W}_2\boldsymbol{U}_2^{\mathrm{H}}$, and $\boldsymbol{G}_2 = g_2(\boldsymbol{H}_2) = \boldsymbol{V}_2\boldsymbol{\Sigma}_2\boldsymbol{U}_2^{\mathrm{H}}$. The expectation is taken over $\boldsymbol{H}_2$. Due to the local optimality defined in Section III.C that other transmitters are *a priori* fixed to the desired structure, $\boldsymbol{W}_2$ and $\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_2^{\mathrm{H}}$ here are real diagonal matrices and independent of $\boldsymbol{V}_2$ and $\boldsymbol{U}_2$. Furthermore, note that $\boldsymbol{H}_2\boldsymbol{F}_2 = \boldsymbol{U}_2\boldsymbol{D}_2\boldsymbol{W}_2\boldsymbol{U}_2^{\mathrm{H}}$ and $|\boldsymbol{H}_2\boldsymbol{G}_2|^2 = \boldsymbol{U}_2\boldsymbol{D}_2\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_2^{\mathrm{H}}\boldsymbol{D}_2^{\mathrm{H}}\boldsymbol{U}_2^{\mathrm{H}}$ are Hermitian, and that $\boldsymbol{U}_2$ is a *Haar* matrix. From Property 2 in Section II.B, $\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{H}_2\boldsymbol{F}_2\boldsymbol{U}_1$ and $\boldsymbol{U}_1^{\mathrm{H}}|\boldsymbol{H}_2\boldsymbol{G}_2|^2\boldsymbol{U}_1$ have the same distribution as $\boldsymbol{H}_2\boldsymbol{F}_2$ and $|\boldsymbol{H}_2\boldsymbol{G}_2|^2$, respectively. Thus, with the individual CSI (i.e., $\boldsymbol{H}_1 = \boldsymbol{U}_1\boldsymbol{D}_1\boldsymbol{V}_1^{\mathrm{H}}$) of the transmitter 1, we rewrite the objective function in (15) as

$$\mathrm{E}\left[\log\det\left(\boldsymbol{I}+\frac{1}{\sigma^2}\left(\left|\boldsymbol{D}_1\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{F}_1\boldsymbol{U}_1+\boldsymbol{H}_2\boldsymbol{F}_2\right|^2+\left|\boldsymbol{D}_1\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{G}_1\right|^2+\left|\boldsymbol{H}_2\boldsymbol{G}_2\right|^2\right)\right)\middle|\boldsymbol{H}_1\right]$$
$$=\mathrm{E}\left[\log\det\left(\boldsymbol{I}+\frac{1}{\sigma^2}\begin{pmatrix}\boldsymbol{D}_1 & \boldsymbol{H}_2\boldsymbol{F}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{\Lambda}_1 & \boldsymbol{W}_1\\ \boldsymbol{W}_1^{\mathrm{H}} & \boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2\end{pmatrix}\begin{pmatrix}\boldsymbol{D}_1^{\mathrm{H}}\\ \boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}\end{pmatrix}\right)\middle|\boldsymbol{H}_1\right] \quad (16)$$

where $\boldsymbol{W}_1 = \boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{F}_1\boldsymbol{U}_1$ from (9) and $\boldsymbol{\Lambda}_1$ is defined as

$$\boldsymbol{\Lambda}_1 = \boldsymbol{V}_1^{\mathrm{H}}\left(\boldsymbol{F}_1\boldsymbol{F}_1^{\mathrm{H}}+\boldsymbol{G}_1\boldsymbol{G}_1^{\mathrm{H}}\right)\boldsymbol{V}_1. \quad (17)$$

From (16) and (17), the problem in (15) can be rewritten as an optimization problem for $\boldsymbol{\Lambda}_1$ and $\boldsymbol{W}_1$ as

$$\max_{\{\Lambda_1,W_1\}:\ \mathrm{tr}\{\Lambda_1\}\le P_1,\ W_1W_1^{\mathrm{H}}\preceq\Lambda_1} \varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1) \quad (18)$$

where $\boldsymbol{X}\preceq\boldsymbol{Y}$ means that $\boldsymbol{Y}-\boldsymbol{X}$ is positive semi-definite, and $\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)\triangleq$

$$\mathrm{E}\left[\log\det\left(\boldsymbol{I}+\frac{1}{\sigma^2}\begin{pmatrix}\boldsymbol{D}_1 & \boldsymbol{H}_2\boldsymbol{F}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{\Lambda}_1 & \boldsymbol{W}_1\\ \boldsymbol{W}_1^{\mathrm{H}} & \boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2\end{pmatrix}\begin{pmatrix}\boldsymbol{D}_1^{\mathrm{H}}\\ \boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}\end{pmatrix}\right)\middle|\boldsymbol{H}_1\right] \quad (19)$$

The following lemmas are proven in Subsection A-C. Let $(\boldsymbol{A})_{\mathrm{diag}}$ be the diagonal matrix formed by the diagonal of $\boldsymbol{A}$.

Lemma 1: The maximum in (18) is upper-bounded as

$$\max_{\substack{\Lambda_1:\ \mathrm{tr}\{\Lambda_1\}\le P_1\\ W_1:\ W_1W_1^{\mathrm{H}}\preceq\Lambda_1}} \varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)\le \max_{\substack{\Lambda_1:\mathrm{tr}\{\Lambda_1\}\le P_1,\\ W_1:(W_1)_{\mathrm{diag}}(W_1)_{\mathrm{diag}}^{\mathrm{H}}\preceq(\Lambda_1)_{\mathrm{diag}}}} \varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1). \quad (20)$$

Lemma 2: For any positive semi-definite matrix $\boldsymbol{\Lambda}_1$ and any matrix $\boldsymbol{W}_1$, we have

$$\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)\le\varphi\left((\boldsymbol{\Lambda}_1)_{\mathrm{diag}},(\boldsymbol{W}_1)_{\mathrm{diag}}\right). \quad (21)$$

Lemma 3: The optimal solution to the problem below is achieved at real-valued $\boldsymbol{\Lambda}_1$ and $\boldsymbol{W}_1$.

$$\max_{\substack{\{\Lambda_1,W_1\}:\mathrm{tr}\{\Lambda_1\}\le P_1,\ W_1W_1^{\mathrm{H}}\preceq\Lambda_1\\ \Lambda_1 \text{ and } W_1 \text{ are diagonal}}} \varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1). \quad (22)$$

From Lemma 1-3, we have

$$\max_{\{\Lambda_1,W_1\}:\ \mathrm{tr}\{\Lambda_1\}\le P_1, W_1W_1^{\mathrm{H}}\preceq\Lambda_1} \varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1) \quad (23a)$$
$$\overset{\text{using Lemma 1}}{\le} \max_{\{\Lambda_1,W_1\}:\mathrm{tr}\{\Lambda_1\}\le P_1,\ (W_1)_{\mathrm{diag}}(W_1)_{\mathrm{diag}}^{\mathrm{H}}\preceq(\Lambda_1)_{\mathrm{diag}}} \varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)$$
$$\overset{\text{using Lemma 2}}{\le} \max_{\{\Lambda_1,W_1\}:\mathrm{tr}\{(\Lambda_1)_{\mathrm{diag}}\}\le P_1,(W_1)_{\mathrm{diag}}(W_1)_{\mathrm{diag}}^{\mathrm{H}}\preceq(\Lambda_1)_{\mathrm{diag}}} \varphi\left((\boldsymbol{\Lambda}_1)_{\mathrm{diag}},(\boldsymbol{W}_1)_{\mathrm{diag}}\right)$$
$$\overset{\text{using Lemma 3}}{=} \max_{\substack{\{\Lambda_1,W_1\}:\mathrm{tr}\{\Lambda_1\}\le P_1,\ W_1W_1^{\mathrm{H}}\preceq\Lambda_1\\ \Lambda_1 \text{ and } W_1 \text{ are real diagonal}}} \varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1) \quad (23b)$$

Comparing (23a) with (23b), we conclude that the optimal $\boldsymbol{\Lambda}_1$ and $\boldsymbol{W}_1$ to the problem in (18) should be real diagonal matrices. Finally, from (9) and (17), we have

$$\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_1^{\mathrm{H}} = \boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{G}_1\boldsymbol{U}_1\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{G}_1^{\mathrm{H}}\boldsymbol{V}_1 = \boldsymbol{\Lambda}_1-\boldsymbol{V}_1^{\mathrm{H}}\boldsymbol{F}_1\boldsymbol{F}_1^{\mathrm{H}}\boldsymbol{V}_1 = \boldsymbol{\Lambda}_1-\boldsymbol{W}_1\boldsymbol{W}_1^{\mathrm{H}}$$

which is real diagonal provided that $\boldsymbol{\Lambda}_1$ and $\boldsymbol{W}_1$ are. Hence Theorem 1 holds.

A. *Proof of Lemma 1*

Let $\boldsymbol{B}$ be a Hermitian matrix. Consider any matrix $\boldsymbol{A}$ satisfying $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\preceq\boldsymbol{B}$, i.e., $\boldsymbol{B}-\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ is positive semi-definite. The diagonal elements of $\boldsymbol{B}-\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ are non-negative, i.e., $B_{i,i}-(|A_{i,i}|^2+\sum_{j\ne i}|A_{i,j}|^2)\ge 0$, for $i = 1, \ldots, N$. Thus, $B_{i,i}-|A_{i,i}|^2 \ge 0$, for $i = 1, \ldots, N$, or equivalently, $(\boldsymbol{A})_{\mathrm{diag}}(\boldsymbol{A})_{\mathrm{diag}}^{\mathrm{H}}\preceq(\boldsymbol{B})_{\mathrm{diag}}$. The above reasoning implies that

$$\left\{\boldsymbol{A}\mid\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\preceq\boldsymbol{B}\right\}\subseteq\left\{\boldsymbol{A}\mid(\boldsymbol{A})_{\mathrm{diag}}(\boldsymbol{A})_{\mathrm{diag}}^{\mathrm{H}}\preceq(\boldsymbol{B})_{\mathrm{diag}}\right\}. \quad (24)$$

Letting $\boldsymbol{A} = \boldsymbol{W}_1$ and $\boldsymbol{B} = \boldsymbol{\Lambda}_1$, we see that the region of $\boldsymbol{W}_1$ for $\boldsymbol{W}_1\boldsymbol{W}_1^{\mathrm{H}}\preceq\boldsymbol{\Lambda}_1$ is contained in that for $(W_1)_{\mathrm{diag}}(W_1)_{\mathrm{diag}}^{\mathrm{H}}\preceq(\Lambda_1)_{\mathrm{diag}}$. Therefore, (20) holds.

B. *Proof of Lemma 2*

We first consider the case of $M = N$. Let $\bar{\boldsymbol{U}}_l$ be an M-by-M diagonal matrix with the diagonal elements being ±1. There are in total 2^M different such matrices indexed from $l = 1$ to 2^M. Then (cf., [15])

$$\frac{1}{2^M}\sum_{l=1}^{2^M}\bar{\boldsymbol{U}}_l\boldsymbol{A}\bar{\boldsymbol{U}}_l=(\boldsymbol{A})_{\mathrm{diag}} \quad (25)$$

where $\boldsymbol{A}$ be an arbitrary M-by-M matrix.

For the function defined in (19), we note that both $\boldsymbol{H}_2\boldsymbol{F}_2$ and $\boldsymbol{I}+|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2|^2$ are Hermitian matrices, and that $\boldsymbol{U}_2$ is a *Haar* matrix. From Property 2 in Section II.B, $\bar{\boldsymbol{U}}_l\boldsymbol{H}_2\boldsymbol{F}_2\bar{\boldsymbol{U}}_l$ and $\bar{\boldsymbol{U}}_l(\boldsymbol{I}+|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2|^2)\bar{\boldsymbol{U}}_l$ have the same distribution as $\boldsymbol{H}_2\boldsymbol{F}_2$ and $\boldsymbol{I}+|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2|^2$, respectively. Thus, (19) can be written as:

$$\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)=\mathrm{E}\left[\log\det\left(\boldsymbol{I}+\frac{1}{\sigma^2}\begin{pmatrix}\bar{\boldsymbol{U}}_l\boldsymbol{D}_1\bar{\boldsymbol{U}}_l & \bar{\boldsymbol{U}}_l\boldsymbol{H}_2\boldsymbol{F}_2\bar{\boldsymbol{U}}_l\end{pmatrix}\begin{pmatrix}\boldsymbol{\Lambda}_1 & \boldsymbol{W}_1\\ \boldsymbol{W}_1^{\mathrm{H}} & \bar{\boldsymbol{U}}_l\left(\boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2\right)\bar{\boldsymbol{U}}_l\end{pmatrix}\begin{pmatrix}\bar{\boldsymbol{U}}_l\boldsymbol{D}_1^{\mathrm{H}}\bar{\boldsymbol{U}}_l\\ \bar{\boldsymbol{U}}_l\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}\bar{\boldsymbol{U}}_l\end{pmatrix}\right)\middle|\boldsymbol{H}_1\right]$$
$$=\mathrm{E}\left[\log\det\left(\boldsymbol{I}+\frac{1}{\sigma^2}\begin{pmatrix}\boldsymbol{D}_1 & \boldsymbol{H}_2\boldsymbol{F}_2\end{pmatrix}\right.\right.$$

$$\left.\left.\begin{pmatrix}\bar{U}_l\Lambda_1\bar{U}_l & \bar{U}_l W_1 \bar{U}_l \\ \bar{U}_l W_1^{\mathrm{H}}\bar{U}_l & I+\left|F_2^{-1}G_2\right|^2\end{pmatrix}\begin{pmatrix}D_1^{\mathrm{H}} \\ F_2^{\mathrm{H}}H_2^{\mathrm{H}}\end{pmatrix}\right)\right|H_1\Bigg]$$

$$=\varphi\left(\bar{U}_l\Lambda_1\bar{U}_l,\bar{U}_l W_1\bar{U}_l\right) \quad (26)$$

Define

$$\psi(\Lambda_1,W_1)=I+\frac{1}{\sigma^2}\begin{pmatrix}D_1 & H_2F_2\end{pmatrix}\begin{pmatrix}\bar{U}_l\Lambda_1\bar{U}_l & \bar{U}_l W_1 \bar{U}_l \\ \bar{U}_l W_1^{\mathrm{H}}\bar{U}_l & I+\left|F_2^{-1}G_2\right|^2\end{pmatrix}\begin{pmatrix}D_1^{\mathrm{H}} \\ F_2^{\mathrm{H}}H_2^{\mathrm{H}}\end{pmatrix}$$

Note that logdet(·) is concave and $\psi(\Lambda_1,W_1)$ is an affine function of Λ_1 and W_1. As composition with affine mapping preserves concavity [16], $\varphi(\Lambda_1,W_1)=\mathrm{E}\left[\log\det\left(\psi(\Lambda_1,W_1)\right)\right]$ is concave in (Λ_1, W_1). Thus, we have

$$\varphi(\Lambda_1,W_1)=\frac{1}{2^M}\sum_{l=1}^{2^M}\varphi(\Lambda_1,W_1)\overset{(a)}{=}\frac{1}{2^M}\sum_{l=1}^{2^M}\varphi\left(\bar{U}_l\Lambda_1\bar{U}_l,\bar{U}_l W_1\bar{U}_l\right)$$

$$\overset{(b)}{\le}\varphi\left(\frac{1}{2^M}\sum_{l=1}^{2^M}\bar{U}_l\Lambda_1\bar{U}_l,\frac{1}{2^M}\sum_{l=1}^{2^M}\bar{U}_l W_1\bar{U}_l\right)\overset{(c)}{=}\varphi\left((\Lambda_1)_{\mathrm{diag}},(W_1)_{\mathrm{diag}}\right) \quad (27)$$

where step (a) follows from (26), step (b) follows from the Jensen's inequality, and step (c) utilizes (25). Thus, (21) holds for the case of $M = N$.

We now discuss the case of $M > N$. We replace (25) with

$$\frac{1}{2^M}\sum_{l=1}^{2^M}\bar{U}_l^{\mathrm{left}}A\bar{U}_l^{\mathrm{right}}=(A)_{\mathrm{diag}}$$

where A is an arbitrary M-by-N matrix, $\bar{U}_l^{\mathrm{left}}=\bar{U}_l$, and $\bar{U}_l^{\mathrm{right}}$ is the N-by-N principle submatrix of $\bar{U}_l^{\mathrm{left}}$ with index set {1, …, N}. The other reasoning literally follows the case of $M = N$, except for some minor modifications. The treatment for $M < N$ is similar. This completes the proof of Lemma 2.

C. *Proof of Lemma 3*

By definition in (17), Λ_1 is Hermitian. This implies that, any diagonal Λ_1 is real-valued. Thus, we only need to consider W_1. We first show that, for any diagonal matrix Λ_1 and W_1,

$$\varphi(\Lambda_1,W_1)=\varphi\left(\Lambda_1,W_1^*\right), \quad (28)$$

where $(\cdot)^*$ represents the conjugate operation.

Since $\det(A) = \det(A^*)$ for any Hermitian matrix A, we have

$$\varphi(\Lambda_1,W_1)=\mathrm{E}\left[\log\det\left(I+\frac{1}{\sigma^2}\begin{pmatrix}D_1 & (H_2F_2)^*\end{pmatrix}\right.\right.$$

$$\left.\left.\begin{pmatrix}\Lambda_1 & W_1^* \\ W_1^{\mathrm{T}} & \left(I+\left|F_2^{-1}G_2\right|^2\right)^*\end{pmatrix}\begin{pmatrix}D_1^{\mathrm{H}} \\ \left(F_2^{\mathrm{H}}H_2^{\mathrm{H}}\right)^*\end{pmatrix}\right)\right|H_1\Bigg]. \quad (29)$$

From Property 2 in Section II.B, $(H_2F_2)^*$ and $(I+|F_2^{-1}G_2|^2)^*$ have the same distribution as H_2F_2 and $I+|F_2^{-1}G_2|^2$, respectively. Thus, we obtain

$$\varphi(\Lambda_1,W_1)=\mathrm{E}\left[\log\det\left(I+\frac{1}{\sigma^2}\begin{pmatrix}D_1 & H_2F_2\end{pmatrix}\right.\right.$$

$$\left.\left.\begin{pmatrix}\Lambda_1 & W_1^* \\ W_1^{\mathrm{T}} & I+\left|F_2^{-1}G_2\right|^2\end{pmatrix}\begin{pmatrix}D_1^{\mathrm{H}} \\ F_2^{\mathrm{H}}H_2^{\mathrm{H}}\end{pmatrix}\right)\right|H_1\Bigg]$$

$$=\varphi\left(\Lambda_1,W_1^*\right) \quad (30)$$

Then $\varphi(\Lambda_1,W_1)\overset{(a)}{=}\frac{1}{2}\left(\varphi(\Lambda_1,W_1)+\varphi\left(\Lambda_1,W_1^*\right)\right)$

$$\overset{(b)}{\le}\varphi\left(\Lambda_1,\frac{1}{2}\left(W_1+W_1^*\right)\right)=\varphi\left(\Lambda_1,\mathrm{Re}\{W_1\}\right) \quad (31)$$

where step (a) follows from (30), and (b) from the Jensen's inequality (as $\varphi(\Lambda_1, W_1)$ is concave). The equality in (31) is achieved when W_1 is a real-valued matrix.

Consider the optimization problem in (22). For any diagonal matrix W_1 satisfying $W_1W_1^{\mathrm{H}}\preceq\Lambda_1$, we obtain $\mathrm{Re}\{W_1\}\mathrm{Re}\{W_1\}^{\mathrm{H}}\preceq\Lambda_1$. This implies that $\mathrm{Re}\{W_1\}$ falls into the feasible region of (22) if W_1 does. Together with (31), we conclude that the optimal W_1 to (22) is real-valued.

ACKNOWLEDGMENT

This work was partially supported by a grant from the University Grants Committee (Project No. AoE/E-02/08) of the Hong Kong Special Administrative Region, China.

REFERENCES

[1] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally effient communications in cellular systems," *IEEE Wireless Commun. Mag.,* vol. 13, pp. 56-61, Aug. 2006.

[2] A. Tolli, M. Codreanu, and M. Juntti, "Cooperative MIMO-OFDM cellular system with soft handover between distributed base station antennas," *IEEE Trans. Wireless Commun.,* vol. 7, pp. 1428-1440, Apr. 2008.

[3] H.-F. Chong, M. Motani, and F. Nan, "Transmitter optimization for distributed Gaussian MIMO channels," in *Proc. Inf. Theory and Applications Workshop (ITA), 2010,* San Diego, CA, Feb.2010, pp. 1-6.

[4] R. Zhang, "Cooperative multi-cell block diagonalization with per-base-station power constraints," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp.1435-1445. Dec. 2010.

[5] TR 36.814 V1.0.0, "Evolved universal terrestrial radio access (e-utra); further advancements for e-utra physical layer aspects," Feb. 2009.

[6] F. Xue, and S. Sandhu, "Cooperation in a half-duplex Gaussian diamond relay channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3806-3814, Oct. 2007.

[7] J. N. Laneman, and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415-2425, Oct. 2003.

[8] Y. Jing, and B. Hassibi, "Distributed space-time coding in wireless relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3524-3536, Dec. 2006.

[9] A. M. Tulino, and S. Verdu, "Random matrix theory and wireless communications," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 1, pp.1-182, 2004.

[10] Y. J. Zhang, and A. M.-Cho So, "Optimal spectrum sharing in MIMO cognitive radio networks via semidefinite programming," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp.362-373. Feb. 2011.

[11] F. Mezzadri, "How to generate random matrices from the classical compact groups," *Notices of the AMS*, vol. 54, no. 5, pp. 502-604, May 2007.

[12] J. Zhang, X. Yuan, P. Li, "Hermitian precoding for distributed MIMO systems with individual channel state information," to be submitted.

[13] R. A. Horn, and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1990.

[14] D. N. C. Tse, and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge: Cambridge University Press, 2005.

[15] Y. C. Eldar, and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1931-1946, Jul. 2004.

[16] S. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.