Hermitian Precoding for Distributed MIMO Systems with Individual Channel State Information

Jianwen Zhang, Xiaojun Yuan, *Member, IEEE,* and Li Ping, *Fellow, IEEE,*

***Abstract*—We consider a distributed multiple-input multiple-output (MIMO) system in which multiple transmitters cooperatively serve a common receiver. It is usually very costly to acquire full channel state information at the transmitter (CSIT) in such a scenario, especially for large-scale antenna systems. In this paper, we assume individual CSIT (I-CSIT), i.e., each transmitter has perfect CSI of its own link but only slow fading factors of the others. A linear Hermitian precoding technique is proposed to enhance the system performance. The optimality of the proposed precoding technique is analyzed. Numerical results demonstrate that the performance loss incurred by the I-CSIT assumption is negligible as compared to the full-CSIT case.**

***Index Terms*—Distributed MIMO Systems, Cooperative Cellular Systems, Hermitian Precoding.**

I. Introduction

CONSIDER a distributive MIMO system in which multiple transmitters located at different places cooperatively send a common message to a single receiver. Each terminal is equipped with multiple antennas. This scenario arises, e.g., in the following two applications.

- In a cooperative cellular system, several adjacent base stations simultaneously serve a mobile terminal 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]–[11].

The above scenario has been investigated under various assumptions on the availability of channel state information at the transmitters (CSIT). The works in [9]–[11] assumed no CSIT and proposed distributed space-time coding for efficient transmission. The works in [1]–[5] assumed full CSIT (i.e., every transmitter perfectly knows all the channel state information (CSI)). With CSIT, it is possible to perform beamforming jointly among the transmitters, which can significantly enhance the system performance.

In practice, however, it is very costly to acquire full CSI at all transmitters. This problem is especially serious in a large-scale multiple-input multiple-output (MIMO) system (in which each terminal is equipped with a large number of antennas) [12]. Transmitting CSI from the receiver to the transmitters imposes heavy stress on the feedback link. It is also a daunting task if each transmitter has to share its CSIT with others in a cooperative system.

This difficulty can be partially relieved in the so-called time-division duplex (TDD) system [13] in which the CSI of each link is obtained by assuming channel reciprocity. In this way, no feedback link is required, but then each transmitter has the CSI of its own link only.

This motivates us to investigate the design of transmission strategies with partial CSIT. In particular, we focus on an individual CSIT (I-CSIT) setup, in which each transmitter has the CSI of its own link but only the slow fading factors of the others. We propose a linear Hermitian precoding for transforming the equivalent channel, including a physical channel and a precoder, into a Hermitian matrix form. We show analytically that the proposed scheme provides an efficient solution in the distributive environment. The performance of the proposed scheme with I-CSIT is very close to the system capacity with full CSIT, as demonstrated by analytical and numerical results. In particular, we show that the proposed scheme is asymptotically capacity approaching in a MIMO system with a large number of antennas at the transmitter side.

The notation $(\boldsymbol{A})_{\text{diag}}$ denotes the diagonal matrix formed by the diagonal of $\boldsymbol{A}$; the matrix $\text{Re}\{\boldsymbol{A}\}$ is formed by the real part of matrix $\boldsymbol{A}$; $\boldsymbol{A} \succeq \boldsymbol{B}$ means that $\boldsymbol{A}-\boldsymbol{B}$ is positive semi-definite; $|\boldsymbol{A}|^2$ is a shorthand of $\boldsymbol{A}\boldsymbol{A}^{\text{H}}$; and $\|\boldsymbol{a}\|_2$ represents the Euclidean norm of vector $\boldsymbol{a}$. We say that a function $f(x)$ is $O(g(x))$ if there exists a constant c so that $|f(x)| \leq c\,|g(x)|$ for sufficiently large x.

II. System Model and Preliminaries

A. System Model

Consider a distributed MIMO system, in which K transmitters cooperatively transmit common messages to a single receiver. Each transmitter is equipped with N antennas and the receiver is equipped with M antennas. The received signal 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

Manuscript received 1 February 2012; revised 15 June 2012. This work was supported by a grant from the University Grants Committee (Project No. AoE/E-02/08) of the Hong Kong Special Administrative Region, China. This paper was presented in part at the IEEE International Symposium on Information Theory, Cambridge, MA, USA, July, 2012.

J. Zhang and P. Li are with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong SAR, China. (e-mail: jzhang35@student.cityu.edu.hk; eeliping@cityu.edu.hk).

X. Yuan is with the Institute of Network Coding, Department of Information Engineering, the Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, China. (e-mail: xjyuan@inc.cuhk.edu.hk)

Digital Object Identifier 10.1109/JSAC.2013.130212.

Gaussian noise vector. We assume that each transmitter k has an individual power constraint of

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

where P_k is the maximum transmission power of transmitter k. We always assume that the receiver perfectly knows $\{\boldsymbol{H}_k | k = 1, \ldots, K\}$.

B. Haar Matrix

The following properties of *Haar* matrices are useful for our further discussions.

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 $\boldsymbol{U}$) [14], [15].

Property 1: [p.25, [14]] A *Haar* matrix is unitarily invariant, i.e., the statistical behavior of a *Haar* matrix $\boldsymbol{U}$ remains unchanged when multiplied by any unitary matrix $\boldsymbol{T}$ independent of $\boldsymbol{U}$. In other words, $\boldsymbol{TU}$ has the same distribution as $\boldsymbol{U}$.

Property 2: [p.25, [14]] For any L-by-L *Haar* matrix $\boldsymbol{U}$,

$$\mathrm{E}\left[|U_{i,j}|^2\right] = 1/L, \quad \text{and} \quad \mathrm{E}\left[U_{i,j} U_{k,l}^*\right] = 0,$$

$$\forall 1 \leq i, j, k, l \leq L, \quad i \neq k \quad \text{or} \quad j \neq l. \tag{3}$$

where $(\cdot)^*$ represents the conjugate operation, and the expectation is taken over $\boldsymbol{U}$. Furthermore, given any L-by-L matrix $\boldsymbol{A}$,

$$\mathrm{E}\left[\boldsymbol{U}\boldsymbol{A}\boldsymbol{U}^{\mathrm{H}}\right] = \left(\mathrm{tr}\left\{\boldsymbol{A}\right\}/L\right)\boldsymbol{I}. \tag{4}$$

Property 3: [p.25, [14]] Let $\boldsymbol{U}$ be a *Haar* matrix and $\boldsymbol{\Lambda}$ be a random diagonal matrix. Assume that $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ are independent of each other. Denote $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}$. Then, $\boldsymbol{B}$ is unitarily invariant, i.e., for any unitary matrix $\boldsymbol{T}$ independent of $\boldsymbol{B}$, $\boldsymbol{T}\boldsymbol{B}\boldsymbol{T}^{\mathrm{H}}$ has the same distribution as $\boldsymbol{B}$. Also, $\boldsymbol{B}^*$ has the same distribution as $\boldsymbol{B}$.

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

All channels between the transmitters and the receiver are assumed to be Rayleigh fading, i.e., the entries of the M-by-N channel matrix $\boldsymbol{H}_k$ are i.i.d. drawn from $\mathcal{CN}$(0, 1). Then $\boldsymbol{H}_k\boldsymbol{H}_k^{\mathrm{H}}$ is a central *Wishart* matrix and is unitarily invariant. From Lemma 2.6 in [14], $\boldsymbol{H}_k\boldsymbol{H}_k^{\mathrm{H}}$ can be decomposed as

$$\boldsymbol{H}_k\boldsymbol{H}_k^{\mathrm{H}} = \boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{D}_k^{\mathrm{H}}\boldsymbol{U}_k^{\mathrm{H}} \tag{5a}$$

with $\boldsymbol{U}_k$ a *Haar* matrix independent of the M-by-N diagonal matrix[1] $\boldsymbol{D}_k$. Let the singular value decomposition (SVD) of $\boldsymbol{H}_k$ be

$$\boldsymbol{H}_k = \boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{V}_k^{\mathrm{H}} \tag{5b}$$

Then we conclude that

Property 4: $\boldsymbol{U}_k$ is a *Haar* matrix independent of $\boldsymbol{D}_k$ [14].

[1]For a diagonal matrix with size M-by-N, the only possibly non-zero entries are located at the (i, i)th position with $i = 1, \ldots, \min\{M, N\}$.

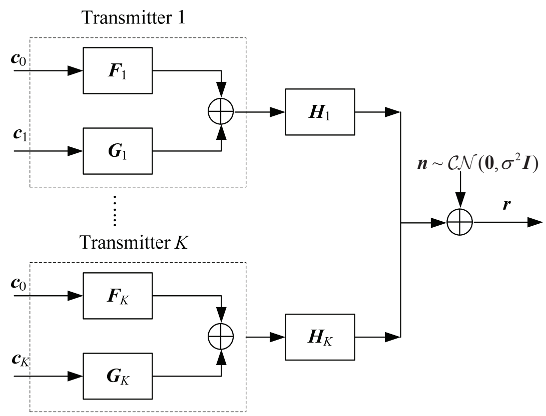


Fig. 1. System model for a distributed MIMO channel.

III. Transmission Strategy with I-CSIT

Linear precoding has been studied to shape the channel input covariance matrix so as to achieve MIMO capacity [16]. The conventional SVD water filling (SVD-WF) approach [16] can be seen as a special case of linear precoding. However, SVD-WF requires global CSIT and thus is not suitable for distributed systems. In this section, we present a distributed linear precoding technique that can overcome this difficulty.

A. Modeling of the Transmit Signal

We assume that the transmitters share the same data to be transmitted. However, the transmitters have the freedom to use the same codebook or different codebooks in channel coding. This implies that the transmitted signals from different transmitters 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 each transmitter k (for $k = 1, \ldots, K$), where $\{\boldsymbol{c}_k\}$ are M-by-1 random vectors with the entries independently drawn from $\mathcal{CN}$(0, 1). By definition, $\mathrm{E}\left[\boldsymbol{c}_k\boldsymbol{c}_k^{\mathrm{H}}\right] = \boldsymbol{I}$ and $\mathrm{E}\left[\boldsymbol{c}_k\boldsymbol{c}_j^{\mathrm{H}}\right] = \boldsymbol{0}, \forall k, j = 0, 1, \ldots, K, k \neq j$. Then, the transmitted signal of transmitter k can be expressed as

$$\boldsymbol{x}_k = \boldsymbol{F}_k\boldsymbol{c}_0 + \boldsymbol{G}_k\boldsymbol{c}_k, \quad k = 1, \ldots, K, \tag{6}$$

where $\boldsymbol{F}_k$ and $\boldsymbol{G}_k$ are N-by-M precoding matrices of transmitter k designed to exploit the available CSI. With (6), the received signal in (1) is rewritten as

$$\boldsymbol{r} = \sum\nolimits_{k=1}^{K} \boldsymbol{H}_k\left(\boldsymbol{F}_k\boldsymbol{c}_0 + \boldsymbol{G}_k\boldsymbol{c}_k\right) + \boldsymbol{n}. \tag{7}$$

The corresponding transmission scheme is illustrated in Fig. 1. The remainder of this paper is concerned with the design of $\{\boldsymbol{F}_k\}$ and $\{\boldsymbol{G}_k\}$ to enhance the system performance.

B. Distributive Precoder Design

We say that a transmitter k has individual CSIT (I-CSIT) if it knows $\boldsymbol{H}_k$ but no knowledge about $\{\boldsymbol{H}_{k'}, k' \neq k\}$. A precoding design is said to be *distributive* if it involves I-CSIT only, which implies that both $\boldsymbol{F}_k$ and $\boldsymbol{G}_k$ in (7) are determined solely by $\boldsymbol{H}_k$, i.e.,

$$\boldsymbol{F}_k = f_k\left(\boldsymbol{H}_k\right), \text{ and } \boldsymbol{G}_k = g_k\left(\boldsymbol{H}_k\right), \; k = 1, \ldots, K, \tag{8}$$

where $f_k(\cdot)$ and $g_k(\cdot)$, $k = 1, \ldots, K$, are precoding functions to be optimized.

We wish to optimize the distributive precoding functions $\{f_k(\cdot)\}$ and $\{g_k(\cdot)\}$, $k = 1, \ldots, K$, so as to maximize the average achievable rate of the system in (7) under the individual transmitter power constraints in (2). This problem can be formulated as

$$\max_{\{f_k(\cdot)\},\{g_k(\cdot)\}} \mathrm{E}\left[\log\det\left(\boldsymbol{I} + \frac{1}{\sigma^2}\left(\Big|\sum_{k=1}^{K}\boldsymbol{H}_k f_k(\boldsymbol{H}_k)\Big|^2 + \sum_{k=1}^{K}\left|\boldsymbol{H}_k g_k(\boldsymbol{H}_k)\right|^2\right)\right)\right] \tag{9a}$$

$$\text{s.t.} \quad \mathrm{tr}\left\{|f_k(\boldsymbol{H}_k)|^2 + |g_k(\boldsymbol{H}_k)|^2\right\} \le P_k, \quad \forall \boldsymbol{H}_k,\ k = 1, \ldots, K \tag{9b}$$

where the expectation in (9a) is taken over the joint distribution of $\{\boldsymbol{H}_k\}$, and the power constraint holds for every possible realization of $\{\boldsymbol{H}_k\}$.

C. Locally Optimal Solution

The problem in (9) requires jointly optimizing $\{f_k(\cdot)\}$ and $\{g_k(\cdot)\}$, which is difficult to tackle with. From now on, we will focus on a locally optimal solution in the sense that any particular transmitter k deviating from its own precoding strategy leads to performance degradation. More rigorously, we say that a set of precoding functions $\{f_k(\cdot), g_k(\cdot)\}_{k=1}^{K}$ is *locally optimal* if, $\forall k$,

$$\{f_k(\boldsymbol{H}_k), g_k(\boldsymbol{H}_k)\} = \underset{\{\boldsymbol{F}_k,\boldsymbol{G}_k\}:\mathrm{tr}\{\boldsymbol{F}_k\boldsymbol{F}_k^{\mathrm{H}}+\boldsymbol{G}_k\boldsymbol{G}_k^{\mathrm{H}}\}\le P_k}{\arg\max} \Psi(\boldsymbol{F}_k, \boldsymbol{G}_k) \tag{10a}$$

with

$$\Psi(\boldsymbol{F}_k, \boldsymbol{G}_k) \equiv \mathrm{E}\left[\log\det\left(\boldsymbol{I} + \frac{1}{\sigma^2}\left(\Big|\boldsymbol{H}_k\boldsymbol{F}_k + \sum_{k'\neq k}^{K}\boldsymbol{H}_{k'} f_{k'}(\boldsymbol{H}_{k'})\Big|^2 + |\boldsymbol{H}_k\boldsymbol{G}_k|^2 + \sum_{k'\neq k}^{K}|\boldsymbol{H}_{k'} g_{k'}(\boldsymbol{H}_{k'})|^2\right)\right)\Bigg|\boldsymbol{H}_k\right] \tag{10b}$$

where $\mathrm{E}[\cdot|\boldsymbol{H}_k]$ means conditional expectation given $\boldsymbol{H}_k$. Thus, the expectation in (10) is taken over $\boldsymbol{H}_{k'}, k' \neq k$. In what follows, we study precoding techniques to meet (10).

IV. Hermitian Precoding

In this section, we propose a distributive Hermitian precoding technique for the system in (7) and prove its local optimality. We further study the power control problem for the proposed Hermitian precoder and analyze its asymptotic performance in large-scale MIMO systems.

A. Basic Precoder Structure

Let $\boldsymbol{U}_k$ and $\boldsymbol{V}_k$ be given by the SVD of $\boldsymbol{H}_k$ in (5b). As $\boldsymbol{U}_k$ and $\boldsymbol{V}_k$ are invertible, the precoders in (8) can always be written into the following form, $\forall k = 1, \ldots, K$

$$\boldsymbol{F}_k = f_k(\boldsymbol{H}_k) = \boldsymbol{V}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}} \tag{11a}$$

$$\boldsymbol{G}_k = g_k(\boldsymbol{H}_k) = \boldsymbol{V}_k\boldsymbol{\Sigma}_k\boldsymbol{U}_k^{\mathrm{H}}. \tag{11b}$$

We first do not impose any restriction on $\boldsymbol{W}_k$ and $\boldsymbol{\Sigma}_k$ except that their sizes are both N-by-M and they are only depend on $\boldsymbol{H}_k$ (due to the I-CSIT assumption). We have the following theorem.

Theorem 1: The optimal $\{\boldsymbol{W}_k\}$ and $\{\boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_k^{\mathrm{H}}\}$ to the problem in (10) (which is a relaxed version of the original problem in (9)) are real diagonal matrices.

The proof of Theorem 1 is given in Appendix. Theorem 1 gives the optimal precoder structures for the problem defined in (10). From Theorem 1, the optimal $\boldsymbol{\Sigma}_k$ is given by $\boldsymbol{\Sigma}_k = \tilde{\boldsymbol{\Sigma}}_k\boldsymbol{Q}_k$, where $\tilde{\boldsymbol{\Sigma}}_k$ is a real diagonal matrix and $\boldsymbol{Q}_k$ is an unitary matrix. Since $\boldsymbol{Q}_k$ has no impact on the achievable rate in (9a), in what follows, we will always assume that $\{\boldsymbol{W}_k\}$ and $\{\boldsymbol{\Sigma}_k\}$ are real diagonal matrices and discuss their optimization techniques.

B. Power Allocation: Determining $\{\boldsymbol{\Sigma}_k\}$

We first consider $\{\boldsymbol{\Sigma}_k\}$. For simplicity of discussion, we only consider the case of $N = M$. The extension of our discussions to the case of $N \neq M$ is straightforward.

Due to the symmetry of the transmitters, it suffices to study the optimization problem in (10) with $k = 1$. Specifically, we assume that $f_{k'}(\boldsymbol{H}_{k'}) = \boldsymbol{V}_{k'}\boldsymbol{W}_{k'}\boldsymbol{U}_{k'}^{\mathrm{H}}$ and $g_{k'}(\boldsymbol{H}_{k'}) = \boldsymbol{V}_{k'}\boldsymbol{\Sigma}_{k'}\boldsymbol{U}_{k'}^{\mathrm{H}}$, $\forall k' \neq 1$. Then the objective function in (10) is given by:

$$\begin{aligned}
&\mathrm{E}\left[\log\det\left(\boldsymbol{I} + \frac{1}{\sigma^2}\left(\Big|\boldsymbol{U}_1\boldsymbol{D}_1\boldsymbol{W}_1\boldsymbol{U}_1^{\mathrm{H}} + \sum_{k=2}^{K}\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}}\Big|^2 + \left|\boldsymbol{U}_1\boldsymbol{D}_1\boldsymbol{\Sigma}_1\boldsymbol{U}_1^{\mathrm{H}}\right|^2 + \sum_{k=2}^{K}\left|\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{\Sigma}_k\boldsymbol{U}_k^{\mathrm{H}}\right|^2\right)\right)\Bigg|\boldsymbol{H}_1\right] \\
=&\mathrm{E}\left[\log\det\left(\boldsymbol{I} + \frac{1}{\sigma^2}\left(\Big|\boldsymbol{D}_1\boldsymbol{W}_1 + \sum_{k=2}^{K}\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}}\boldsymbol{U}_1\Big|^2 + |\boldsymbol{D}_1\boldsymbol{\Sigma}_1|^2 + \sum_{k=2}^{K}\left|\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{\Sigma}_k\boldsymbol{U}_k^{\mathrm{H}}\boldsymbol{U}_1\right|^2\right)\right)\Bigg|\boldsymbol{D}_1\right] \\
=&\mathrm{E}\left[\log\det\left(\boldsymbol{I} + \frac{1}{\sigma^2}\left(\Big|\boldsymbol{D}_1\boldsymbol{W}_1 + \sum_{k=2}^{K}\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}}\Big|^2 + |\boldsymbol{D}_1\boldsymbol{\Sigma}_1|^2 + \sum_{k=2}^{K}\left|\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{\Sigma}_k\boldsymbol{U}_k^{\mathrm{H}}\right|^2\right)\right)\Bigg|\boldsymbol{D}_1\right]
\end{aligned} \tag{12}$$

where (12) follows the fact that $\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}}$ and $\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{\Sigma}_k\boldsymbol{U}_k^{\mathrm{H}}$ are Hermitian matrices and that $\{\boldsymbol{U}_k\}$ are *Haar* matrices and independent of $\{\boldsymbol{W}_k\}$ and $\{\boldsymbol{\Sigma}_k\}$ seen from Property 4. From Property 3 in Section II-B, $\{\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}}\boldsymbol{U}_1\}$ and $\{\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{\Sigma}_k\boldsymbol{U}_k^{\mathrm{H}}\boldsymbol{U}_1\}$ have the same distribution as $\{\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}}\}$ and $\{\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{\Sigma}_k\boldsymbol{U}_k^{\mathrm{H}}\}$, respectively.

Then the related optimization problem is given in (13) at the top of next page. This problem is difficult to solve due to its non-convex nature. Thus, instead of directly solving (13), we resort to a widely used upper-bound technique (cf., [17]). From the Jensen's inequality (as $\log\det(\cdot)$ is concave), the objective function in (13) is upper-bounded by the function $\phi(\boldsymbol{W}_1, \boldsymbol{\Sigma}_1)$ shown in (14) at the top of next page. We formulate the

$$\max_{\{\boldsymbol{W}_1,\boldsymbol{\Sigma}_1\}:\mathrm{tr}\{\boldsymbol{W}_1^2+\boldsymbol{\Sigma}_1^2\}\le P_1} \mathrm{E}\left[\log\det\left(\boldsymbol{I}+\frac{1}{\sigma^2}\Big(\Big|\boldsymbol{D}_1\boldsymbol{W}_1+\sum_{k=2}^{K}\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}}\Big|^2+|\boldsymbol{D}_1\boldsymbol{\Sigma}_1|^2+\sum_{k=2}^{K}\big|\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{\Sigma}_k\boldsymbol{U}_k^{\mathrm{H}}\big|^2\Big)\right)\middle|\boldsymbol{D}_1\right] \tag{13}$$

$$\phi(\boldsymbol{W}_1,\boldsymbol{\Sigma}_1)\equiv\log\det\left(\boldsymbol{I}+\frac{1}{\sigma^2}\mathrm{E}\left[\Big|\boldsymbol{D}_1\boldsymbol{W}_1+\sum_{k=2}^{K}\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}}\Big|^2+|\boldsymbol{D}_1\boldsymbol{\Sigma}_1|^2+\sum_{k=2}^{K}\big|\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{\Sigma}_k\boldsymbol{U}_k^{\mathrm{H}}\big|^2\middle|\boldsymbol{D}_1\right]\right) \tag{14}$$

---

following upper-bound optimization problem:

$$\max_{\{\boldsymbol{W}_1,\boldsymbol{\Sigma}_1\}:\mathrm{tr}\{\boldsymbol{W}_1^2+\boldsymbol{\Sigma}_1^2\}\le P_1} \phi(\boldsymbol{W}_1,\boldsymbol{\Sigma}_1). \tag{15}$$

We have the following result on the solution to (15).

Theorem 2: The optimal $\boldsymbol{\Sigma}_1$ to the problem in (15) is given by $\boldsymbol{\Sigma}_1=\boldsymbol{0}$.

Proof: Using Property 2 in Section II-B, we obtain

$$\begin{aligned}\phi(\boldsymbol{W}_1,\boldsymbol{\Sigma}_1)&=\log\det\Big(\boldsymbol{I}+\frac{1}{\sigma^2}\big(|\boldsymbol{D}_1\boldsymbol{W}_1|^2+|\boldsymbol{D}_1\boldsymbol{\Sigma}_1|^2\\&\quad+\mu\boldsymbol{D}_1\boldsymbol{W}_1+\mu\boldsymbol{W}_1^{\mathrm{H}}\boldsymbol{D}_1^{\mathrm{H}}+\nu\boldsymbol{I}\big)\Big)\\&=\log\det\Big(\boldsymbol{I}+\frac{1}{\sigma^2}\big(\boldsymbol{D}_1\boldsymbol{\Lambda}\boldsymbol{D}_1^{\mathrm{H}}+\mu\boldsymbol{D}_1\boldsymbol{W}_1\\&\quad+\mu\boldsymbol{W}_1^{\mathrm{H}}\boldsymbol{D}_1^{\mathrm{H}}+\nu\boldsymbol{I}\big)\Big)\end{aligned} \tag{16}$$

where

$$\boldsymbol{\Lambda}=\boldsymbol{W}_1\boldsymbol{W}_1^{\mathrm{H}}+\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_1^{\mathrm{H}} \tag{17}$$

$$\mu=\mathrm{E}\left[\frac{1}{M}\sum_{k=2}^{K}\mathrm{tr}\{\boldsymbol{D}_k\boldsymbol{W}_k\}\right] \tag{18}$$

and

$$\nu=\mathrm{E}\left[\frac{1}{M}\mathrm{tr}\left\{\Big|\sum_{k=2}^{K}\boldsymbol{D}_k\boldsymbol{W}_k\Big|^2+\sum_{k=2}^{K}|\boldsymbol{D}_k\boldsymbol{\Sigma}_k|^2\right\}\right]. \tag{19}$$

With $\boldsymbol{\Lambda}$ defined in (17), the problem in (15) can be rewritten as:

$$\max_{\substack{\{\boldsymbol{\Lambda},\boldsymbol{W}_1\}:\\ \mathrm{tr}\{\boldsymbol{\Lambda}\}\le P_1,\boldsymbol{W}_1^2\preceq\boldsymbol{\Lambda}}} \log\det\Big(\boldsymbol{I}+\frac{1}{\sigma^2}\big(\boldsymbol{D}_1\boldsymbol{\Lambda}\boldsymbol{D}_1^{\mathrm{H}}+\mu\boldsymbol{D}_1\boldsymbol{W}_1+\mu\boldsymbol{W}_1^{\mathrm{H}}\boldsymbol{D}_1^{\mathrm{H}}+\nu\boldsymbol{I}\big)\Big) \tag{20a}$$

or equivalently,

$$\max_{\substack{\sum_{i=1}^{M}\Lambda_i\le P_1,\\ w_i^2\le\Lambda_i,i=1,\ldots,M}} \sum_{i=1}^{M}\log\left(1+\frac{1}{\sigma^2}\left(d_i^2\Lambda_i+2\mu d_iw_i+\nu\right)\right) \tag{20b}$$

where w_i, d_i, and Λ_i are the ith diagonal element of $\boldsymbol{W}_1$, $\boldsymbol{D}_1$, and $\boldsymbol{\Lambda}$, respectively. It is seen that the objective in (20b) is an increasing function of w_i. The optimal w_i is $w_i=\sqrt{\Lambda_i}$. From (17), $\Sigma_{1i,i}^2=\Lambda_i-w_i^2=0$, where $\Sigma_{1i,i}$ stands for the (i, i)th element of $\boldsymbol{\Sigma}_1$. Therefore, for the problem in (15), the optimal $\boldsymbol{\Sigma}_1$ is $\boldsymbol{\Sigma}_1=\boldsymbol{0}$. ■

Theorem 2 gives the optimal solution of $\boldsymbol{\Sigma}_1=\boldsymbol{0}$ for the problem in (15) (that is an approximation of the original problem in (13)). Our numerical results in Section V show that this scheme performs very close to the full CSIT upper bound. This demonstrates the effectiveness of the suboptimal solution.

C. Power Allocation: Determining $\{\boldsymbol{W}_k\}$

We next consider $\{\boldsymbol{W}_k\}$. We still focus on transmitter 1 (i.e., k = 1). With Theorem 2, the problem in (20b) is reduced to an optimization problem of $\boldsymbol{W}_1$ as

$$\max_{\sum_{i=1}^{M}w_i^2\le P_1}\sum_{i=1}^{M}\log\left(1+\frac{1}{\sigma^2}\left(d_i^2w_i^2+2\mu d_iw_i+\nu\right)\right) \tag{21}$$

The maximum in (21) is achieved when all $\{w_i\}$ have the same sign of μ, which implies that the optimal $\boldsymbol{W}_1$ must be either positive semi-definite or negative semi-definite. Noting the symmetry, it suffices to study the case that $\boldsymbol{W}_1$ is positive semi-definite.

We next solve (21) using the Karush-Kuhn-Tucker (KKT) conditions [18]. The associated Lagrangian is

$$\begin{aligned}&L(w_1,\ldots,w_M,\lambda)\\&=\sum_{i=1}^{M}\log\left(1+\frac{1}{\sigma^2}\left(d_i^2w_i^2+2\mu d_iw_i+\nu\right)\right)-\lambda\Big(\sum_{i=1}^{M}w_i^2-P_1\Big).\end{aligned} \tag{22}$$

The corresponding KKT conditions are given by

$$\lambda d_i^2w_i^3+2\lambda\mu d_iw_i^2+(\lambda\sigma^2+\lambda\nu-d_i^2)w_i-\mu d_i=0 \tag{23a}$$

$$\lambda\Big(\sum_{i=1}^{M}w_i^2-P_1\Big)=0\text{ and }\lambda\ge0 \tag{23b}$$

$$\sum_{i=1}^{M}w_i^2\le P_1\text{ and }w_i\ge0. \tag{23c}$$

The above KKT conditions are easy to solve since, for any given λ, (23a) is a univariate cubic equation of w_i (with at most three different solutions). What remains is to determine λ, which can be found by a bisection method. The reason is that λ is a monotonically decreasing function with respect to each w_i, since

$$\lambda=\frac{d_i^2w_i+\mu d_i}{d_i^2w_i^3+2\mu d_iw_i^2+\sigma^2w_i+\nu w_i},\forall i=1,\ldots,M,$$

which follows from (23a). Therefore, the KKT conditions in (23) can be numerically solved, which yields the optimal $\boldsymbol{W}_1$ to (21).

From the above discussion, we see that the optimal $\boldsymbol{W}_1$ to (21) can be determined given μ and ν. From (18)-(19), the calculation of μ and ν involves the statistical information of the other transmitters, which may be difficult to acquire in practice. In our numerical study, we approximately calculate μ and ν as follows.

- Initialization: Set $\mu^{(0)}=0$, and $\nu^{(0)}=0$.

- Step 1: Solve the KKT conditions (23) to obtain $\{w_i\}$ based on $\mu = \mu^{(\ell-1)}$ and $\nu = \nu^{(\ell-1)}$ for the ℓth channel realization.
- Step 2: Update $\mu^{(\ell)}$, and $\nu^{(\ell)}$ by:

$$\mu^{(\ell)} = \frac{1}{\ell}\left((\ell-1)\mu^{(\ell-1)} + \frac{1}{M}\left(\sum_{k=2}^{K}\sqrt{\frac{P_k}{P_1}}\right)\sum_{i=1}^{M} d_i w_i\right), \tag{24a}$$

and

$$\nu^{(\ell)} = \frac{1}{\ell}\left((\ell-1)\nu^{(\ell-1)} + \frac{1}{M}\left(\sum_{k=2}^{K}\sqrt{\frac{P_k}{P_1}}\right)^2\sum_{i=1}^{M} d_i^2 w_i^2\right). \tag{24b}$$

- Step 3: Update $\ell = \ell + 1$ and generate a new channel realization, then go to Step 1.

In Step 2, $\{\boldsymbol{D}_k\boldsymbol{W}_k\}, k = 2, \ldots, M$ in (18)-(19) are approximated as $\sqrt{P_k/P_1}\boldsymbol{D}_1\boldsymbol{W}_1$. Intuitively, this is reasonable since every transmitter statistically experience the same channel conditions.

The above algorithm only involves I-CSI and is thus distributive.

D. Intuitions and Discussions

From Theorems 1 and 2, given $\boldsymbol{H}_k = \boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{V}_k^{\mathrm{H}}$, our proposed precoding strategy is, $\forall k$

$$\boldsymbol{F}_k = f_k(\boldsymbol{H}_k) = \boldsymbol{V}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}} \quad \text{and} \quad \boldsymbol{G}_k = g_k(\boldsymbol{H}_k) = \boldsymbol{0}. \tag{25}$$

Then the transmitted and received signals in (6) and (7) can be rewritten respectively as

$$\boldsymbol{x}_k = \boldsymbol{F}_k\boldsymbol{c}_0 = \boldsymbol{V}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}}\boldsymbol{c}_0, \tag{26a}$$

$$\boldsymbol{r} = \left(\sum_{k=1}^{K}\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}}\right)\boldsymbol{c}_0 + \boldsymbol{n} = \boldsymbol{A}\boldsymbol{c}_0 + \boldsymbol{n}, \tag{26b}$$

where $\boldsymbol{A} = \sum_{k=1}^{K}\boldsymbol{A}_k$ with $\boldsymbol{A}_k \equiv \boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}}$. Since each $\boldsymbol{A}_k$ is a positive semi-definite Hermitian matrix, we refer to (25) as a *Hermitian precoding* scheme.

The complexity of Hermitian precoding described above consists of two parts: (i) the SVD operation for determining $\{\boldsymbol{V}_k\}$, $\{\boldsymbol{D}_k\}$, and $\{\boldsymbol{U}_k\}$; (ii) power allocation for determining $\{\boldsymbol{W}_k\}$. The complexity of SVD is roughly $O(MN^2)$ for each transmitter [19], and the complexity for computing $\{\boldsymbol{W}_k\}$ given in Section IV-C is $O(M)$ for each round of searching over λ.

With Hermitian precoding, the received signal power is given by

$$\begin{aligned} P_R &= \mathrm{E}\left[\boldsymbol{c}_0^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{c}_0\right] = \mathrm{tr}\left\{\left(\sum_{k=1}^{K}\boldsymbol{A}_k\right)\left(\sum_{k=1}^{K}\boldsymbol{A}_k\right)^{\mathrm{H}}\right\} \\ &= \sum_{k=1}^{K}\mathrm{tr}\{\boldsymbol{A}_k\boldsymbol{A}_k^{\mathrm{H}}\} + \sum_{i=1}^{K}\sum_{j\neq i}^{K}\mathrm{tr}\{\boldsymbol{A}_i\boldsymbol{A}_j^{\mathrm{H}}\} \end{aligned} \tag{27}$$

It can be shown that $\mathrm{tr}\{\boldsymbol{A}_i\boldsymbol{A}_j^{\mathrm{H}}\} \geq 0, \forall i, j$ when all $\{\boldsymbol{A}_i\}$ are positive semi-definite Hermitian matrices. Thus, all the terms in (27) add together coherently. This coherent-transmission effect provides an intuitive explanation of the related gain of Hermitian precoding.

The proposed Hermitian precoding scheme can be extended to the situation with slow fading effect. Let $\boldsymbol{H}_k = \rho_k\tilde{\boldsymbol{H}}_k$ and $\boldsymbol{H}_k\boldsymbol{H}_k^{\mathrm{H}} = \boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{D}_k^{\mathrm{H}}\boldsymbol{U}_k^{\mathrm{H}}$ as in (5a), where the entries of $\tilde{\boldsymbol{H}}_k$ are i.i.d. drawn from $\mathcal{CN}(0,\,1)$, ρ_k is the slow-fading factor including path-loss and shadow fading. Then Theorems 1 and 2 still hold for $\{\boldsymbol{H}_k\}$ since these two theorems only rely on the *Haar* property of $\{\boldsymbol{U}_k\}$. In computing μ and ν in (18)-(19), we may assume $\{\rho_k\}$ are known to all transmitters. As $\{\rho_k\}$ are real numbers that change much slower than the Rayleigh fading factors $\{\tilde{\boldsymbol{H}}_k\}$, feeding back $\{\rho_k\}$ incurs much less overhead than feeding back $\{\boldsymbol{H}_k\}$. The related performance can be found in Fig. 5 in Section V.

E. Asymptotic Performance in Large-Scale MIMO Systems

In this subsection, we consider systems in which N is sufficient large but M remains small, thus $N > M$. This scenario arises in practice when multiple base stations jointly serve a common mobile terminal. The latter typically has a limited physical size and thus a small M.

For simplicity of discussions, we assume in this subsection that all transmitters have the same power constraint P (i.e. $P_k = P, \forall k = 1, \ldots, K$). Let R_H be the achievable rate of the Hermitian precoding scheme. Denote by $\boldsymbol{I}_{m\times n}$ an m-by-n matrix with the only nonzero elements being 1s located at (i, i)th position for $i = 1, \ldots, \min\{m, n\}$.

Consider the limiting case when $N \to \infty$ and M remains fixed. Then $N^{-1}\boldsymbol{H}_k\boldsymbol{H}_k^{\mathrm{H}} \to \boldsymbol{I}$ with probability one (cf. Lemma 4 in [20]). Consequently, $\boldsymbol{D}_k \to \sqrt{N}\boldsymbol{I}_{M\times N}, \boldsymbol{W}_k \to \sqrt{P/M}\boldsymbol{I}_{N\times M}, \forall k$, and the Hermitian precoder $\boldsymbol{F}_k$ in (25) reduces to

$$\boldsymbol{F}_k = \boldsymbol{V}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}} = \sqrt{P/M}\boldsymbol{V}_k\boldsymbol{I}_{N\times M}\boldsymbol{U}_k^{\mathrm{H}}. \tag{28}$$

Substituting $\boldsymbol{D}_k = \sqrt{N}\boldsymbol{I}_{M\times N}$ and $\boldsymbol{F}_k$ in (28) into (9a), we obtain the achievable rate of the Hermitian precoder:

$$R_H = M\log\left(1 + \frac{K^2NP}{M\sigma^2}\right) \approx M\log\left(\frac{K^2NP}{M\sigma^2}\right). \tag{29}$$

On the other hand, as $N^{-1}\boldsymbol{H}_k\boldsymbol{H}_k^{\mathrm{H}} \to \boldsymbol{I}$, the capacity with full CSIT is also given by (29). Thus a Hermitian precoder is asymptotically capacity approaching when $N \to \infty$ and M remains fixed.

We emphasize that the above result is the consequence of the use of $\boldsymbol{U}_k^{\mathrm{H}}$ in (28), which is a key feature of a Hermitian precoder. Otherwise, suppose that we remove $\boldsymbol{U}_k^{\mathrm{H}}$ in $\boldsymbol{F}_k$ and construct an alternative precoder: $\tilde{\boldsymbol{F}}_k = \sqrt{P/M}\boldsymbol{V}_k\boldsymbol{I}_{N\times M}$. Substituting $\tilde{\boldsymbol{F}}_k$ and $\boldsymbol{D}_k$ into (9a), we obtain the achievable rate of $\tilde{\boldsymbol{F}}_k$:

$$\begin{aligned} R_I &= \mathrm{E}\left[\log\det\left(\boldsymbol{I} + \frac{1}{\sigma^2}\left(\sum_{k=1}^{K}\boldsymbol{U}_k\sqrt{\frac{NP}{M}}\right)\left(\sum_{k=1}^{K}\boldsymbol{U}_k\sqrt{\frac{NP}{M}}\right)^{\mathrm{H}}\right)\right] \\ &= \mathrm{E}\left[\log\det\left(\boldsymbol{I} + \frac{PN}{M\sigma^2}\left(\sum_{k=1}^{K}\boldsymbol{U}_k\boldsymbol{U}_k^{\mathrm{H}} + \sum_{i=1}^{K}\sum_{j=1,j\neq i}^{K}\boldsymbol{U}_i\boldsymbol{U}_j^{\mathrm{H}}\right)\right)\right] \\ &\overset{(a)}{\leq} \log\det\left(\boldsymbol{I} + \frac{KPN}{M\sigma^2}\boldsymbol{I} + \sum_{i=1}^{K}\sum_{j=1,j\neq i}^{K}\mathrm{E}\left[\boldsymbol{U}_i\boldsymbol{U}_j^{\mathrm{H}}\right]\right) \end{aligned}$$

$$\stackrel{(b)}{=} \log\det\left(\boldsymbol{I} + \frac{KPN}{M\sigma^2}\boldsymbol{I}\right) = M\log\left(1 + \frac{KPN}{M\sigma^2}\right)$$
$$\approx M\log\left(\frac{KPN}{M\sigma^2}\right) \quad (30)$$

where step (a) follows from Jensen's inequality (as $\log\det(\cdot)$ is concave), and step (b) follows from the fact that $\mathrm{E}\left[\boldsymbol{U}_i\boldsymbol{U}_j^{\mathrm{H}}\right] = 0, \forall i, j, i \neq j$. From (29) and (30), we have

$$R_H - R_I \geq M\log\left(\frac{K^2NP}{M\sigma^2}\right) - M\log\left(\frac{KPN}{M\sigma^2}\right)$$
$$= M\log K, \quad \text{when} \quad N \to \infty. \quad (31)$$

The above indicates that in a large scale MIMO system with $N \to \infty$ but M fixed, $\boldsymbol{F}_k$ in (28) provides at least $M\log K$ rate gain than $\tilde{\boldsymbol{F}}_k$.

V. Numerical Results

In this section, numerical results are used to demonstrate the performance of the proposed Hermitian precoding technique in large scale distributed MIMO channels. For comparison, we list below a variety of alternative choices of $\boldsymbol{F}_k$ and $\boldsymbol{G}_k$. Full CSIT capacity with total power constraint is also included as an upper bound. For simplicity of discussion, we assume $N = M$.

(i) Full CSIT capacity with total power constraint (Full CSIT): The input covariance matrix is determined by standard SVD water-filling method [16] over $\boldsymbol{H} = [\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_K]$.

(ii) Hermitian Precoding with Optimized Power Allocation (HP-OPA):

$$\boldsymbol{F}_k = \boldsymbol{V}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}} \quad \text{and} \quad \boldsymbol{G}_k = \boldsymbol{0}.$$

Here $\boldsymbol{W}_k$ is obtained using the technique proposed in Section IV.

(iii) Hermitian Precoding with Equal Power Allocation (HP-EPA):

$$\boldsymbol{F}_k = \sqrt{P_k/M}\boldsymbol{V}_k\boldsymbol{U}_k^{\mathrm{H}} \quad \text{and} \quad \boldsymbol{G}_k = \boldsymbol{0}.$$

This is obtained by substituting $\boldsymbol{W}_k = \sqrt{P_k/M}\boldsymbol{I}$ in (25).

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

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

Here $\boldsymbol{W}_k = \beta_k\boldsymbol{D}_k^{-1}$ in (25) with β_k a scalar chosen to meet the power constraint.

(v) No-CSIT with Correlated Signaling (No-CSIT-CS):

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

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

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

(vii) Individual Water-Filling (IWF):

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

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 standard water-filling [16] over $\boldsymbol{D}_k$ at each individual transmitter.

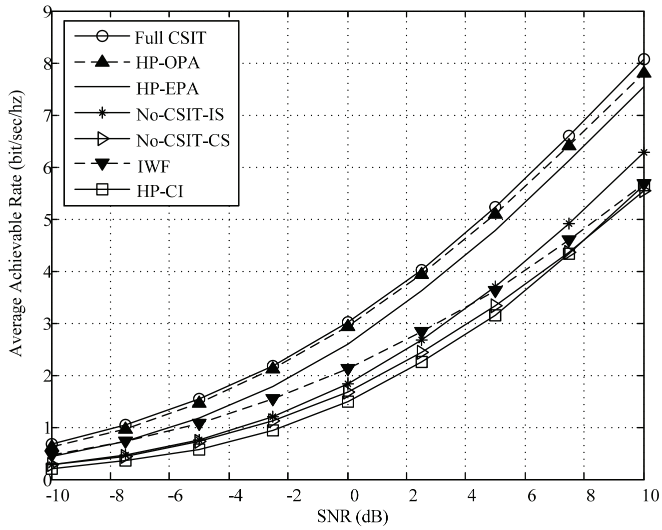


Fig. 2. Performance comparison among different precoding schemes with $M = N = K = 2$ in Rayleigh-fading distributed MIMO Gaussian channels. $\mathrm{SNR} = (P_1 + P_2)/\sigma^2$ with $P_1 = P_2$.

Note that the commonality among choices (ii) - (iv) listed above is the precoding structure $\boldsymbol{V}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}}$. Their only difference is the choice of $\boldsymbol{W}_k$. On the other hand, choices (v) - (vii) do not involve the precoding structure defined in (25), and thus lack the coherent transmission effect discussed in Section IV-D.

Firstly, we consider the scenario without slow-fading factors. Figs. 2-4 show the performance of a Hermitian precoder in this scenario. Note that in these figures, we assume equal power constraints for all transmitters (i.e., $P_k = P, \forall k$) due to the symmetry property among them. The performance of various precoding schemes are compared in Fig. 2 (with $K = M = N = 2$) and Fig. 3 (with $K = 3, M = N = 100$). It is seen that HP-OPA performs very close to the upper bound obtained by assuming full CSIT. This implies that the potential performance loss due to the I-CSIT assumption is marginal. We also see from Fig. 3 that HP-OPA significantly outperforms other alternatives especially when both the transmitters and the receiver equip with large amount of antennas. Note that the performance of No-CSIT-CS, No-CSIT-IS and IWF is relatively poor in both Figs. 2 and 3. The reason is that they cannot achieve the coherent transmission effect provided by the precoding structure defined in (25).

Fig. 4 shows the asymptotic performance of the proposed precoder for large N values and relatively small M values. We fix $K = 3, \mathrm{SNR} = KP/\sigma^2 = 0$dB. From Fig. 4, the gap between the HP-OPA performance and the full-CSIT bound diminishes when N increases. The performance gaps between the HP-OPA and the corresponding IWF scheme are marked. $M\log K$ equals to 3.17 for $M = 2$, 6.34 for $M = 4$, and 25.36 for $M = 16$. By comparing the performance gaps with $M\log K$, we see that these gaps are larger than $M\log K$, which verifies the result in Section IV-E.

In Fig. 5, we consider the scenario with slow-fading factors $\{\rho_k\}$ as discussed in Section IV-D. We set $\rho_k = 1$ for $k = 1, \ldots, K/2$ and $\rho_k = 2$ for $k = K/2 + 1, \ldots, K$. We assume that $\{\rho_k\}$ are known to all transmitters. The power constraints $\{P_k\}$ of different transmitters can be optimized

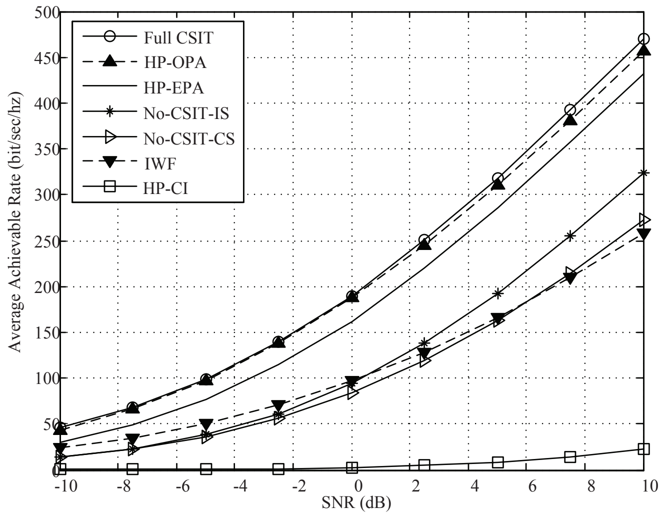


Fig. 3. Performance comparison among different precoding schemes with $K = 3$, $M = N = 100$ in Rayleigh-fading distributed MIMO Gaussian channels. SNR $= KP/\sigma^2$, with $P_k = P, k = 1, 2, 3$.

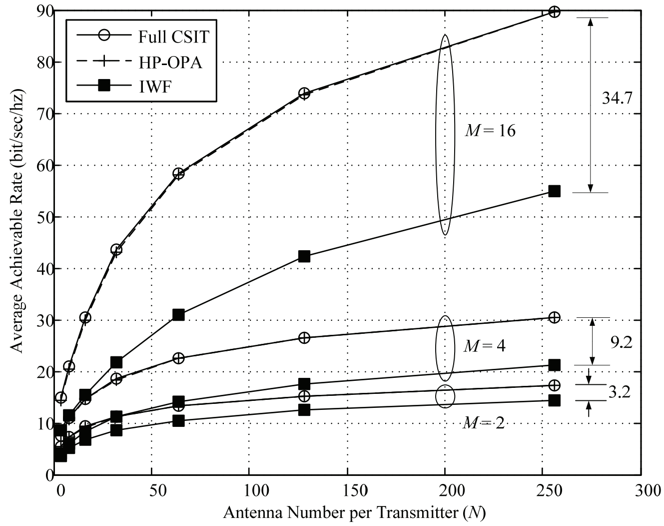


Fig. 4. Comparisons of average achievable rate between the Hermitian precoding scheme with OPA and the channel capacity with full-CSIT. The number of transmitters $K = 3$, and the channel SNR $= KP/\sigma^2$ is fixed to 0dB. The values of M are marked on the curves. $M \log K$ equals to 3.17 for $M = 2$, 6.34 for $M = 4$, and 25.36 for $M = 16$.

according to $\{\rho_k\}$. The performance curve with full-CSIT is included for comparison. It is seen that Hermitian precoding with optimized power constraints slightly outperforms the scheme with equal power constraints. It is also see that the HP-OPA scheme achieves an extra 3dB power gain for every doubling the number of transmitters. This is due to the coherent effect that more transmitters provide more additional power gain.

Finally, we present some simulation results with practical channel coding. Return to the received signal in (26b): $\boldsymbol{r} = \boldsymbol{A}\boldsymbol{c}_0 + \boldsymbol{n}$. The iterative linear minimum-mean-square-error (LMMSE) detection principles developed in [21] is directly applied. Fig. 6 compares the simulated frame error rates (FERs) of various schemes for a distributed MIMO system with $K = 2, M = N = 16$, a (3, 6)-regular LDPC code with codeword length 8192 for channel coding. We see from Fig. 6 that HP-OPA outperforms other alternatives by about 4-6dB. This shows that the potential advantage of HP-OPA analyzed

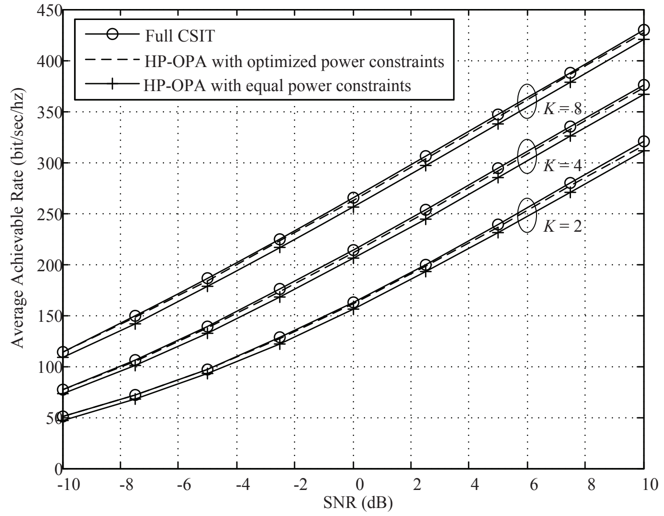


Fig. 5. Comparisons of average achievable rate between the Hermitian precoding scheme with OPA and the channel capacity with full-CSIT in the scenario with slow-fading factors. We set $\rho_k = 1$ for $k = 1, \ldots, K/2$ and $\rho_k = 2$ for $k = K/2 + 1, \ldots, K$. SNR $= \left(\sum_{k=1}^{K} P_k\right)/\sigma^2$. $M = 50, N = 100$. The values of K are marked on the curves.

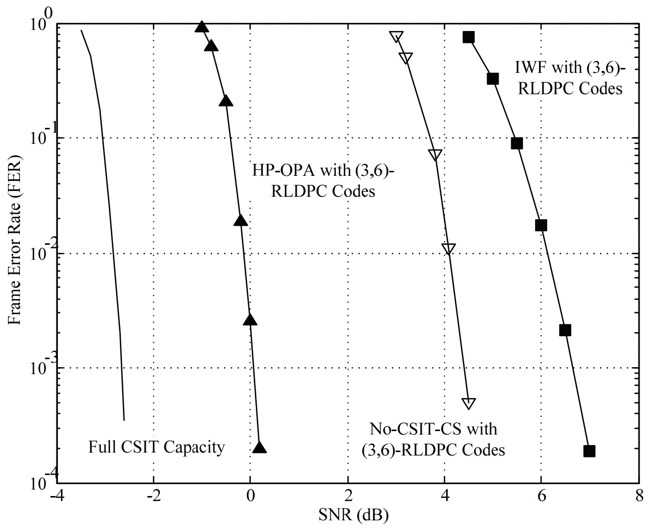


Fig. 6. FER performance on the Rayleigh fading channel with $K = 2$, $N = 16$, and $M = 16$. Regular (3, 6) LDPC codes with codeword length 8192 are used for practical channel coding. All transmitters have the same power constraint P (i.e. $P_1 = P_2 = P$), and SNR $= KP/\sigma^2$.

from an information-theoretic perspective (as demonstrated in Figs. 2-5) is indeed achievable in practice.

VI. Conclusion

We have proposed a Hermitian precoding technique for distributed MIMO channels with I-CSIT and proved its local optimality. Numerical results show that the proposed scheme with I-CSIT can perform close to the channel capacity with full CSIT in various settings. The proposed technique can be used to significantly reduce the overhead related to acquiring CSIT in distributed MIMO channels, especially for large-scale antenna systems.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable suggestions and comments.

$$\max_{\{\boldsymbol{F}_1,\boldsymbol{G}_1\}:\text{tr}\{\boldsymbol{F}_1\boldsymbol{F}_1^{\text{H}}+\boldsymbol{G}_1\boldsymbol{G}_1^{\text{H}}\}\leq P_1} \text{E}\left[\log\det\left(\boldsymbol{I}+\frac{1}{\sigma^2}\left(|\boldsymbol{H}_1\boldsymbol{F}_1+\boldsymbol{H}_2\boldsymbol{F}_2|^2+|\boldsymbol{H}_1\boldsymbol{G}_1|^2+|\boldsymbol{H}_2\boldsymbol{G}_2|^2\right)\right)\middle|\boldsymbol{H}_1\right] \tag{32}$$

$$\begin{aligned}
&\text{E}\left[\log\det\left(\boldsymbol{I}+\frac{1}{\sigma^2}\left(\left|\boldsymbol{U}_1\boldsymbol{D}_1\boldsymbol{V}_1^{\text{H}}\boldsymbol{F}_1+\boldsymbol{U}_1\boldsymbol{H}_2\boldsymbol{F}_2\boldsymbol{U}_1^{\text{H}}\right|^2+\left|\boldsymbol{U}_1\boldsymbol{D}_1\boldsymbol{V}_1^{\text{H}}\boldsymbol{G}_1\right|^2+\boldsymbol{U}_1\left|\boldsymbol{H}_2\boldsymbol{G}_2\right|^2\boldsymbol{U}_1^{\text{H}}\right)\right)\middle|\boldsymbol{H}_1=\boldsymbol{U}_1\boldsymbol{D}_1\boldsymbol{V}_1^{\text{H}}\right]\\
=&\text{E}\left[\log\det\left(\boldsymbol{I}+\frac{1}{\sigma^2}\left(\left|\boldsymbol{D}_1\boldsymbol{V}_1^{\text{H}}\boldsymbol{F}_1\boldsymbol{U}_1+\boldsymbol{H}_2\boldsymbol{F}_2\right|^2+\left|\boldsymbol{D}_1\boldsymbol{V}_1^{\text{H}}\boldsymbol{G}_1\boldsymbol{U}_1\right|^2+\left|\boldsymbol{H}_2\boldsymbol{G}_2\right|^2\right)\right)\middle|\boldsymbol{H}_1\right]\\
=&\text{E}\left[\log\det\left(\boldsymbol{I}+\frac{1}{\sigma^2}\begin{pmatrix}\boldsymbol{D}_1 & \boldsymbol{H}_2\boldsymbol{F}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{V}_1^{\text{H}}\left(\boldsymbol{F}_1\boldsymbol{F}_1^{\text{H}}+\boldsymbol{G}_1\boldsymbol{G}_1^{\text{H}}\right)\boldsymbol{V}_1 & \boldsymbol{V}_1^{\text{H}}\boldsymbol{F}_1\boldsymbol{U}_1\\ \left(\boldsymbol{V}_1^{\text{H}}\boldsymbol{F}_1\boldsymbol{U}_1\right)^{\text{H}} & \boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2\end{pmatrix}\begin{pmatrix}\boldsymbol{D}_1^{\text{H}}\\ \boldsymbol{F}_2^{\text{H}}\boldsymbol{H}_2^{\text{H}}\end{pmatrix}\right)\middle|\boldsymbol{H}_1\right]
\end{aligned} \tag{33}$$

APPENDIX
PROOF OF THEOREM 1

We first consider the case of $K = 2$. Since all transmitters experience the same channel distribution, these two transmitters are symmetric. Then, the problem in (10) can be rewritten in (32) shown at the top of this page, where $\boldsymbol{F}_2 = f_2(\boldsymbol{H}_2) = \boldsymbol{V}_2\boldsymbol{W}_2\boldsymbol{U}_2^{\text{H}}$, and $\boldsymbol{G}_2 = g_2(\boldsymbol{H}_2) = \boldsymbol{V}_2\boldsymbol{\Sigma}_2\boldsymbol{U}_2^{\text{H}}$. 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 $\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_2^{\text{H}}$ here are real diagonal matrices and independent of $\boldsymbol{V}_2$ and $\boldsymbol{U}_2$. Furthermore, note that $\boldsymbol{H}_2\boldsymbol{F}_2 = \boldsymbol{U}_2\boldsymbol{D}_2\boldsymbol{W}_2\boldsymbol{U}_2^{\text{H}}$ and $|\boldsymbol{H}_2\boldsymbol{G}_2|^2 = \boldsymbol{U}_2\boldsymbol{D}_2\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_2^{\text{H}}\boldsymbol{D}_2^{\text{H}}\boldsymbol{U}_2^{\text{H}}$ are Hermitian, and that $\boldsymbol{U}_2$ is a *Haar* matrix and independent of $\boldsymbol{W}_2$ and $\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_2^{\text{H}}$ (cf. Property 4 in Section II-C). From Property 3 stated in Section II-B, $\boldsymbol{U}_1^{\text{H}}\boldsymbol{H}_2\boldsymbol{F}_2\boldsymbol{U}_1$ and $\boldsymbol{U}_1^{\text{H}}|\boldsymbol{H}_2\boldsymbol{G}_2|^2\boldsymbol{U}_1$ 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. $\boldsymbol{H}_1 = \boldsymbol{U}_1\boldsymbol{D}_1\boldsymbol{V}_1^{\text{H}}$) of transmitter 1, the objective function in (32) can be rewritten as (33) given at the top of this page.

From (11), we have $\boldsymbol{W}_1 = \boldsymbol{V}_1^{\text{H}}\boldsymbol{F}_1\boldsymbol{U}_1$ and $\boldsymbol{\Sigma}_1 = \boldsymbol{V}_1^{\text{H}}\boldsymbol{G}_1\boldsymbol{U}_1$. Then define

$$\boldsymbol{\Lambda}_1 = \boldsymbol{V}_1^{\text{H}}\left(\boldsymbol{F}_1\boldsymbol{F}_1^{\text{H}}+\boldsymbol{G}_1\boldsymbol{G}_1^{\text{H}}\right)\boldsymbol{V}_1 \tag{34}$$

From (33) and (34), the problem in (32) is rewritten as an optimization problem for $\boldsymbol{\Lambda}_1$ and $\boldsymbol{W}_1$.

$$\max_{\{\boldsymbol{\Lambda}_1,\boldsymbol{W}_1\}:\text{tr}\{\boldsymbol{\Lambda}_1\}\leq P_1,\boldsymbol{W}_1\boldsymbol{W}_1^{\text{H}}\preceq\boldsymbol{\Lambda}_1} \varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1) \tag{35}$$

where $\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)$ is defined as follow.

$$\begin{aligned}\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1) \equiv \text{E}\Bigg[&\log\det\left(\boldsymbol{I}+\frac{1}{\sigma^2}\begin{pmatrix}\boldsymbol{D}_1 & \boldsymbol{H}_2\boldsymbol{F}_2\end{pmatrix}\right.\\ &\left.\times\begin{pmatrix}\boldsymbol{\Lambda}_1 & \boldsymbol{W}_1\\ \boldsymbol{W}_1^{\text{H}} & \boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2\end{pmatrix}\begin{pmatrix}\boldsymbol{D}_1^{\text{H}}\\ \boldsymbol{F}_2^{\text{H}}\boldsymbol{H}_2^{\text{H}}\end{pmatrix}\right)\Bigg|\boldsymbol{H}_1\Bigg]\end{aligned} \tag{36}$$

The following lemmas are proven in Subsection A, B, and C, respectively.

Lemma 1:

$$\begin{aligned}&\max_{\boldsymbol{\Lambda}_1:\text{tr}\{\boldsymbol{\Lambda}_1\}\leq P_1;\boldsymbol{W}_1:\boldsymbol{W}_1\boldsymbol{W}_1^{\text{H}}\preceq\boldsymbol{\Lambda}_1}\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)\\ \leq&\max_{\substack{\boldsymbol{\Lambda}_1:\text{tr}\{\boldsymbol{\Lambda}_1\}\leq P_1;\\ \boldsymbol{W}_1:(\boldsymbol{W}_1)_{\text{diag}}(\boldsymbol{W}_1)_{\text{diag}}^{\text{H}}\preceq(\boldsymbol{\Lambda}_1)_{\text{diag}}}}\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)\end{aligned} \tag{37}$$

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

$$\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1) \leq \varphi\left((\boldsymbol{\Lambda}_1)_{\text{diag}},(\boldsymbol{W}_1)_{\text{diag}}\right). \tag{38}$$

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

$$\max_{\substack{\{\boldsymbol{\Lambda}_1,\boldsymbol{W}_1\}:\text{tr}\{\boldsymbol{\Lambda}_1\}\leq P_1,\boldsymbol{W}_1\boldsymbol{W}_1^{\text{H}}\preceq\boldsymbol{\Lambda}_1\\ \boldsymbol{\Lambda}_1\text{ and }\boldsymbol{W}_1\text{ are diagonal}}}\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1) \tag{39}$$

From Lemma 1 - Lemma 3, we then have

$$\max_{\{\boldsymbol{\Lambda}_1,\boldsymbol{W}_1\}:\text{tr}\{\boldsymbol{\Lambda}_1\}\leq P_1,\boldsymbol{W}_1\boldsymbol{W}_1^{\text{H}}\preceq\boldsymbol{\Lambda}_1}\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1) \tag{40a}$$

$$\stackrel{(a)}{\leq}\max_{\substack{\boldsymbol{\Lambda}_1:\text{tr}\{\boldsymbol{\Lambda}_1\}\leq P_1\\ \boldsymbol{W}_1:(\boldsymbol{W}_1)_{\text{diag}}(\boldsymbol{W}_1)_{\text{diag}}^{\text{H}}\preceq(\boldsymbol{\Lambda}_1)_{\text{diag}}}}\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)$$

$$\stackrel{(b)}{\leq}\max_{\substack{\boldsymbol{\Lambda}_1:\text{tr}\{(\boldsymbol{\Lambda}_1)_{\text{diag}}\}\leq P_1\\ \boldsymbol{W}_1:(\boldsymbol{W}_1)_{\text{diag}}(\boldsymbol{W}_1)_{\text{diag}}^{\text{H}}\preceq(\boldsymbol{\Lambda}_1)_{\text{diag}}}}\varphi\Big((\boldsymbol{\Lambda}_1)_{\text{diag}},(\boldsymbol{W}_1)_{\text{diag}}\Big)$$

$$\stackrel{(c)}{=}\max_{\substack{\{\boldsymbol{\Lambda}_1,\boldsymbol{W}_1\}:\text{tr}\{\boldsymbol{\Lambda}_1\}\leq P_1,\boldsymbol{W}_1\boldsymbol{W}_1^{\text{H}}\preceq\boldsymbol{\Lambda}_1\\ \boldsymbol{\Lambda}_1\text{ and }\boldsymbol{W}_1\text{ are diagonal}}}\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1) \tag{40b}$$

where step(a)-step(c) follow from Lemma 1-Lemma 3, respectively. Comparing (40a) with (40b), we find that the equalities hold when both $\boldsymbol{\Lambda}_1$ and $\boldsymbol{W}_1$ are real diagonal. Hence, we conclude that the optimal $\boldsymbol{\Lambda}_1$ and $\boldsymbol{W}_1$ to the problem in (35) should be real diagonal matrices. Finally, from (34), we have

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

which is real diagonal provided that $\boldsymbol{\Lambda}_1$ and $\boldsymbol{W}_1$ are. Hence Theorem 1 holds for $K = 2$.

For a general case $K > 2$, the optimization problem in (10) can be rewritten as

$$\begin{aligned}\max_{\substack{\{\boldsymbol{F}_1,\boldsymbol{G}_1\}:\text{tr}\{\boldsymbol{F}_1\boldsymbol{F}_1^{\text{H}}\}\\ +\text{tr}\{\boldsymbol{G}_1\boldsymbol{G}_1^{\text{H}}\}\leq P_1}}\text{E}\Bigg[&\log\det\Bigg(\boldsymbol{I}+\frac{1}{\sigma^2}\Bigg(\Bigg|\boldsymbol{H}_1\boldsymbol{F}_1+\sum_{k=2}^{K}\boldsymbol{H}_k\boldsymbol{F}_k\Bigg|^2\\ &+|\boldsymbol{H}_1\boldsymbol{G}_1|^2+\sum_{k=2}^{K}|\boldsymbol{H}_k\boldsymbol{G}_k|^2\Bigg)\Bigg)\Bigg|\boldsymbol{H}_1\Bigg]\end{aligned} \tag{41}$$

with $\boldsymbol{F}_k = f_k(\boldsymbol{H}_k) = \boldsymbol{V}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\text{H}}$,

and $\boldsymbol{G}_k = g_k(\boldsymbol{H}_k) = \boldsymbol{V}_k\boldsymbol{\Sigma}_k\boldsymbol{U}_k^{\text{H}}, k = 2,\ldots,K.$

Define $\boldsymbol{A}=\sum_{k=2}^{K}\boldsymbol{H}_k\boldsymbol{F}_k=\sum_{k=2}^{K}\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{W}_k\boldsymbol{U}_k^{\mathrm{H}}$,

and $\boldsymbol{B}=\sum_{k=2}^{K}|\boldsymbol{H}_k\boldsymbol{G}_k|^2=\sum_{k=2}^{K}\boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_k^{\mathrm{H}}\boldsymbol{D}_k^{\mathrm{H}}\boldsymbol{U}_k^{\mathrm{H}}$

It can be shown that Property 3 in Section II-B also holds for $\boldsymbol{A}$ and $\boldsymbol{B}$, and so (41) can be rewritten as:

$$\max_{\substack{\{\boldsymbol{F}_1,\boldsymbol{G}_1\}:\\ \mathrm{tr}\{\boldsymbol{F}_1\boldsymbol{F}_1^{\mathrm{H}}+\boldsymbol{G}_1\boldsymbol{G}_1^{\mathrm{H}}\}\le P_1}} \mathrm{E}\Big[\log\det\Big(\boldsymbol{I}+\frac{1}{\sigma^2}\big(|\boldsymbol{H}_1\boldsymbol{F}_1+\boldsymbol{A}|^2 + |\boldsymbol{H}_1\boldsymbol{G}_1|^2+\boldsymbol{B}\big)\Big)\Big|\boldsymbol{H}_1\Big] \tag{42}$$

Following the same reasoning as that from (32) to (40), we can arrive at Theorem 1 for $K>2$.

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., $B_{i,i}-\left(|A_{i,i}|^2+\sum_{j\neq i}|A_{i,j}|^2\right)\ge 0$, 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

$$\left\{\boldsymbol{A}|\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\preceq\boldsymbol{B}\right\}\subseteq\left\{\boldsymbol{A}|(\boldsymbol{A})_{\mathrm{diag}}(\boldsymbol{A})_{\mathrm{diag}}^{\mathrm{H}}\preceq(\boldsymbol{B})_{\mathrm{diag}}\right\}. \tag{43}$$

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 $(\boldsymbol{W}_1)_{\mathrm{diag}}(\boldsymbol{W}_1)_{\mathrm{diag}}^{\mathrm{H}}\preceq(\boldsymbol{\Lambda}_1)_{\mathrm{diag}}$. Therefore, (37) 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. [22]- [23]),

$$\frac{1}{2^M}\sum_{l=1}^{2^M}\bar{\boldsymbol{U}}_l\boldsymbol{A}\bar{\boldsymbol{U}}_l=(\boldsymbol{A})_{\mathrm{diag}} \tag{44}$$

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

For the function defined in (36), we note from Property 4 that both $\boldsymbol{H}_2\boldsymbol{F}_2$ and $\boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2$ are Hermitian matrices with the eigen-values and the eigen-direction being independent of each other, and that $\boldsymbol{U}_2$ is a *Haar* matrix. From Property 3, $\bar{\boldsymbol{U}}_l\boldsymbol{H}_2\boldsymbol{F}_2\bar{\boldsymbol{U}}_l$ and $\bar{\boldsymbol{U}}_l(\boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2)\bar{\boldsymbol{U}}_l$ have the same distribution as $\boldsymbol{H}_2\boldsymbol{F}_2$ and $\boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2$, respectively. Thus, (36) can be rewritten as:

$$\begin{aligned}
&\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)\\
&=\mathrm{E}\Big[\log\det\Big(\boldsymbol{I}+\frac{1}{\sigma^2}\left(\boldsymbol{D}_1\ \ \bar{\boldsymbol{U}}_l\boldsymbol{H}_2\boldsymbol{F}_2\bar{\boldsymbol{U}}_l\right)\\
&\times\begin{pmatrix}\boldsymbol{\Lambda}_1 & \boldsymbol{W}_1\\ \boldsymbol{W}_1^{\mathrm{H}} & \bar{\boldsymbol{U}}_l(\boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2)\bar{\boldsymbol{U}}_l\end{pmatrix}\begin{pmatrix}\boldsymbol{D}_1^{\mathrm{H}}\\ \bar{\boldsymbol{U}}_l\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}\bar{\boldsymbol{U}}_l\end{pmatrix}\Big)\Big|\boldsymbol{H}_1\Big]\\
&=\mathrm{E}\Big[\log\det\Big(\boldsymbol{I}+\frac{1}{\sigma^2}\left(\boldsymbol{D}_1\ \ \boldsymbol{H}_2\boldsymbol{F}_2\right)\\
&\times\begin{pmatrix}\bar{\boldsymbol{U}}_l\boldsymbol{\Lambda}_1\bar{\boldsymbol{U}}_l & \bar{\boldsymbol{U}}_l\boldsymbol{W}_1\bar{\boldsymbol{U}}_l\\ \bar{\boldsymbol{U}}_l\boldsymbol{W}_1^{\mathrm{H}}\bar{\boldsymbol{U}}_l & \boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2\end{pmatrix}\begin{pmatrix}\boldsymbol{D}_1^{\mathrm{H}}\\ \boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}\end{pmatrix}\Big)\Big|\boldsymbol{H}_1\Big]\\
&=\varphi(\bar{\boldsymbol{U}}_l\boldsymbol{\Lambda}_1\bar{\boldsymbol{U}}_l,\bar{\boldsymbol{U}}_l\boldsymbol{W}_1\bar{\boldsymbol{U}}_l).
\end{aligned} \tag{45}$$

Define

$$\psi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)=\boldsymbol{I}+\frac{1}{\sigma^2}\left(\boldsymbol{D}_1\ \ \boldsymbol{H}_2\boldsymbol{F}_2\right)\times\begin{pmatrix}\bar{\boldsymbol{U}}_l\boldsymbol{\Lambda}_1\bar{\boldsymbol{U}}_l & \bar{\boldsymbol{U}}_l\boldsymbol{W}_1\bar{\boldsymbol{U}}_l\\ \bar{\boldsymbol{U}}_l\boldsymbol{W}_1^{\mathrm{H}}\bar{\boldsymbol{U}}_l & \boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2\end{pmatrix}\begin{pmatrix}\boldsymbol{D}_1^{\mathrm{H}}\\ \boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}\end{pmatrix}. \tag{46}$$

Note that $\log\det(\cdot)$ is concave and $\psi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)$ is an affine function of $\boldsymbol{\Lambda}_1$ and $\boldsymbol{W}_1$. As composition with affine mapping preserves concavity [18], $\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)=\mathrm{E}\left[\log\det\left(\psi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)\right)\right]$ is concave in $(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)$. Thus, we have

$$\begin{aligned}
\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)&=\frac{1}{2^M}\sum_{l=1}^{2^M}\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)\\
&\overset{(a)}{=}\frac{1}{2^M}\sum_{l=1}^{2^M}\varphi\left(\bar{\boldsymbol{U}}_l\boldsymbol{\Lambda}_1\bar{\boldsymbol{U}}_l,\bar{\boldsymbol{U}}_l\boldsymbol{W}_1\bar{\boldsymbol{U}}_l\right)\\
&\overset{(b)}{\le}\varphi\Big(\frac{1}{2^M}\sum_{l=1}^{2^M}\bar{\boldsymbol{U}}_l\boldsymbol{\Lambda}_1\bar{\boldsymbol{U}}_l,\frac{1}{2^M}\sum_{l=1}^{2^M}\bar{\boldsymbol{U}}_l\boldsymbol{W}_1\bar{\boldsymbol{U}}_l\Big)\\
&\overset{(c)}{=}\varphi\left((\boldsymbol{\Lambda}_1)_{\mathrm{diag}},(\boldsymbol{W}_1)_{\mathrm{diag}}\right)
\end{aligned} \tag{47}$$

where step (a) follows from (45), step (b) follows from the Jensen's inequality, and step (c) utilizes (44). Thus, (38) holds for the case of $M=N$.

We now discuss the case of $M>N$. We need to replace (44) with

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

where $\boldsymbol{A}$ is an arbitrary M-by-N matrix, $\bar{\boldsymbol{U}}_l^{\mathrm{left}}=\bar{\boldsymbol{U}}_l$, and $\bar{\boldsymbol{U}}_l^{\mathrm{right}}$ is the N-by-N principle submatrix of $\bar{\boldsymbol{U}}_l^{\mathrm{left}}$ with index set $\{1,\ldots,N\}$. 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 (34), $\boldsymbol{\Lambda}_1$ is Hermitian. This implies that, any diagonal $\boldsymbol{\Lambda}_1$ is real-valued. Thus, we only need to consider $\boldsymbol{W}_1$. We first show that, for any diagonal matrix $\boldsymbol{\Lambda}_1$ and $\boldsymbol{W}_1$,

$$\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)=\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1^*), \tag{48}$$

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

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

$$\begin{aligned}
&\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)\\
&=\mathrm{E}\Big[\log\det\Big(\boldsymbol{I}+\frac{1}{\sigma^2}\left(\boldsymbol{D}_1\ \ \boldsymbol{H}_2\boldsymbol{F}_2\right)\\
&\times\begin{pmatrix}\boldsymbol{\Lambda}_1 & \boldsymbol{W}_1\\ \boldsymbol{W}_1^{\mathrm{H}} & \boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2\end{pmatrix}\begin{pmatrix}\boldsymbol{D}_1^{\mathrm{H}}\\ \boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}\end{pmatrix}\Big)^*\Big|\boldsymbol{H}_1\Big]\\
&=\mathrm{E}\Big[\log\det\Big(\boldsymbol{I}+\frac{1}{\sigma^2}\left(\boldsymbol{D}_1\ \ (\boldsymbol{H}_2\boldsymbol{F}_2)^*\right)\\
&\times\begin{pmatrix}\boldsymbol{\Lambda}_1 & \boldsymbol{W}_1^*\\ \boldsymbol{W}_1^{\mathrm{T}} & \left(\boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2\right)^*\end{pmatrix}\begin{pmatrix}\boldsymbol{D}_1^{\mathrm{H}}\\ (\boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}})^*\end{pmatrix}\Big)\Big|\boldsymbol{H}_1\Big].
\end{aligned} \tag{49}$$

From Property 3 in Section II-B, $(\boldsymbol{H}_2\boldsymbol{F}_2)^*$ and $\left(\boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2\right)^*$ have the same distribution as $\boldsymbol{H}_2\boldsymbol{F}_2$ and $\boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2$, respectively. Thus, we obtain

$$\begin{aligned}\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1) = \mathrm{E}\Bigg[\log\det\Bigg(\boldsymbol{I}+\frac{1}{\sigma^2}\left(\boldsymbol{D}_1\ \ \boldsymbol{H}_2\boldsymbol{F}_2\right)\\ \times\begin{pmatrix}\boldsymbol{\Lambda}_1 & \boldsymbol{W}_1^*\\ \boldsymbol{W}_1^{\mathrm{T}} & \boldsymbol{I}+\left|\boldsymbol{F}_2^{-1}\boldsymbol{G}_2\right|^2\end{pmatrix}\begin{pmatrix}\boldsymbol{D}_1^{\mathrm{H}}\\ \boldsymbol{F}_2^{\mathrm{H}}\boldsymbol{H}_2^{\mathrm{H}}\end{pmatrix}\Bigg)\Bigg|\boldsymbol{H}_1\Bigg]\\ = \varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1^*) \qquad (50)\end{aligned}$$

Then

$$\begin{aligned}\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1) &\overset{(a)}{=} \frac{1}{2}\left(\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)+\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1^*)\right)\\ &\overset{(b)}{\le} \varphi\left(\boldsymbol{\Lambda}_1,\frac{1}{2}\left(\boldsymbol{W}_1+\boldsymbol{W}_1^*\right)\right)\\ &= \varphi\left(\boldsymbol{\Lambda}_1,\mathrm{Re}\{\boldsymbol{W}_1\}\right) \qquad (51)\end{aligned}$$

where step (a) follows from (48), and (b) from the Jensen's inequality (as $\varphi(\boldsymbol{\Lambda}_1,\boldsymbol{W}_1)$ is concave). The equality in (51) is achieved when $\boldsymbol{W}_1$ is a real-valued matrix.

Consider the optimization problem in (39). For any diagonal matrix $\boldsymbol{W}_1$ satisfying $\boldsymbol{W}_1\boldsymbol{W}_1^{\mathrm{H}} \preceq \boldsymbol{\Lambda}_1$, we obtain $\mathrm{Re}\{\boldsymbol{W}_1\}\mathrm{Re}\{\boldsymbol{W}_1\}^{\mathrm{H}} \preceq \boldsymbol{\Lambda}_1$. This implies that $\mathrm{Re}\{\boldsymbol{W}_1\}$ falls into the feasible region of (39) if $\boldsymbol{W}_1$ does. Together with (51), we conclude that the optimal $\boldsymbol{W}_1$ to (39) is real-valued.

References

[1] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun. Mag.*, vol. 13, no. 4, pp. 56–61, 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, no. 4, pp. 1428–1440, 2008.

[3] H.-F. Chong, M. Motani, and F. Nan, "Transmitter optimization for distributed gaussian mimo channels," in *Proc. Information Theory and Applications Workshop (ITA)*, 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, 2010.

[5] TR36.814V1.0.0, "Evolved universal terrestrial radio access (e-utra); further advancements for e-utra physical layer aspects," Feb. 2009.

[6] B. Schein and R. Gallager, "The gaussian parallel relay network," in *Proc. IEEE Int Information Theory Symp*, 2000.

[7] 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, 2007.

[8] A. del Coso and C. Ibars, "Bounds on ergodic capacity of multirelay cooperative links with channel state information," in *Proc. IEEE Wireless Communications and Networking Conf. WCNC 2006*, vol. 2, 2006, pp. 902–907.

[9] 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, 2003.

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

[11] S. Yiu, R. Schober, and L. Lampe, "Distributed space-time block coding," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1195–1206, 2006.

[12] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[13] F. Kaltenberger, H. Jiang, M. Guillaud, and R. Knopp, "Relative channel reciprocity calibration in mimo/tdd systems," in *Proc. Future Network and Mobile Summit*, 2010, pp. 1–10.

[14] A. M. Tulino and S. Verdu, "Random matrix theory and wireless communications," *Found. Trends Commun. Inf. Theory*, vol. 1, pp. 1–182, 2004.

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

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.

[17] M. Vu and A. Paulraj, "On the capacity of mimo wireless channels with dynamic csit," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1269–1283, 2007.

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

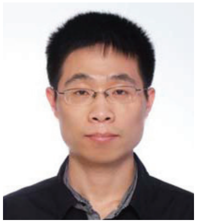
[19] L. N. Trefethen and D. Bau, *Numerical linear algebra*. Society for Industrial and Applied Mathematics (SIAM), 1997.

[20] S. O. Gharan, A. Bayesteh, and A. K. Khandani, "Asymptotic analysis of amplify and forward relaying in a parallel mimo relay network," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2070–2082, 2011.

[21] X. Yuan, C. Xu, L. Ping, and X. Lin, "Precoder design for multiuser mimo isi channels based on iterative lmmse detection," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 6, pp. 1118–1128, 2009.

[22] 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, 2004.

[23] J. Wang and D. P. Palomar, "Robust mmse precoding in mimo channels with pre-fixed receivers," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5802–5818, 2010.



Jianwen Zhang received the B.S. and M.S. degrees in Communication Engineering in 2006 and 2009, respectively, both from Xiamen University, Xiamen, China. He is currently working towards the Ph.D. degree at the City University of Hong Kong. His research interests are communication systems (especially on MIMO relay networks) and coding design.

Xiaojun Yuan (S'04-M'09) received the Ph.D. degree in electrical engineering from the Department of Electronic Engineering, City University of Hong Kong, 2008. During 2009-2010, he was working as a post-doc fellow at the Department of Electric Engineering, University of Hawaii at Manoa. He is now a research assistant professor at the Institute of Network Coding, the Chinese University of Hong Kong.

His research interests are in the general area of wireless communications, signal processing, and information theory, including channel coding and coded modulation, iterative detection and decoding, channel modeling and identification, cooperative communications, physical-layer network coding, etc. He has now published about 40 research papers in the related topics, and is now a member of IEEE.

Li Ping (S'87-M'91-SM'06-F'10) received his Ph.D degree at Glasgow University in 1990. He lectured at Department of Electronic Engineering, Melbourne University, from 1990 to 1992, and worked as a member of research staff at Telecom Australia Research Laboratories from 1993 to 1995. He has been with the Department of Electronic Engineering, City University of Hong Kong, since January 1996, where he is now a chair professor. He received a British Telecom-Royal Society Fellowship in 1986, the IEE J J Thomson premium in 1993, the Croucher Foundation Award in 2005 and a British Royal Academy of Engineering Distinguished Visiting Fellowship in 2010. He is serving as a member of Board of Governors of IEEE Information Theory Society from 2010 to 2012 and is a fellow of IEEE.