DoF-Robust Strategies for the K-user Distributed Broadcast Channel with Weak CSI

Paul de Kerret and David Gesbert

Communication Systems Department, EURECOM, Sophia-Antipolis, France

Abstract

In this paper[1] we consider the Network MIMO channel under the so-called Distributed Channel State Information at the Transmitters (D-CSIT) configuration. In this setting, the precoder is designed in a distributed manner at each Transmitter (TX) on the basis of the locally available multi-user channel estimate. Although the use of simple Zero-Forcing (ZF) was recently shown to reach the optimal DoF for a Broadcast Channel (BC) under noisy, yet centralized, CSIT, it can turn very inefficient in the distributed setting as the achieved number of Degrees-of-Freedom (DoF) is then limited by the worst CSI accuracy across all TXs. To circumvent this effect, we develop a new robust transmission scheme improving the DoF. A surprising result is uncovered by which, in the regime of so-called weak CSIT, the proposed scheme is shown to achieve the centralized outerbound obtained under a genie-aided centralized setting in which the CSI versions available at all TXs are shared among them. Building upon the insight obtained in the weak CSIT regime, we develop a general D-CSI robust scheme which improves over the DoF obtained by conventional ZF approach in an arbitrary CSI quality regime and is shown to achieve the centralized outerbound in some other practically relevant CSIT configurations.

I. Introduction

Multiple-antennas at the TX can be exploited to serve multiple users at the same time, thus offering a strong DoF improvement over time-division schemes [1]. This DoF improvement is however critically dependent on the accuracy of the CSIT. Indeed, the absence of CSIT is known

[1]D. Gesbert and P. de Kerret are supported by the European Research Council under the European Union's Horizon 2020 research and innovation program (Agreement no. 670896).

This paper has been presented in part at the 2016 Information Theory and Applications Workshop and at the 2016 IEEE International Symposium on Information Theory.

to lead to the complete loss of the DoF improvement in the case of a BC with symmetric users [2]. In the noisy (centralized) CSIT regime, a long standing conjecture by Lapidoth, Shamai, and Wigger [3] has been recently settled in [4] by showing that a scaling of the CSIT error in $P^{-\alpha}$ for $\alpha \in [0, 1]$ leads to a DoF of $1 + (K-1)\alpha$ in the K-user BC.

A different line of work in the area of BC with limited feedback has been focused on the exploitation of delayed CSI on the TX side. This research area was triggered by the seminal work from Maddah-Ali and Tse [5] where it was shown that completely outdated CSIT could still be exploited via a multi-phase protocol involving the retransmission of the interference generated. While the original model considered completely outdated CSIT, a large number of works have developed generalized schemes for the case of partially outdated [6], [7], alternating [8], or evolving CSIT [9], to name just a few.

In all the above literature, however, *centralized* CSIT is typically assumed, i.e., precoding is done on the basis of a *single* imperfect/outdated channel estimate being common at *every* transmit antenna. Although meaningful in the case of a BC with a single transmit device, this assumption can be challenged when the joint precoding is carried out across distant TXs linked by heterogeneous and imperfect backhaul links, as in the Network MIMO context. Such a setting could in particular be obtained if the user's data symbols are cached at the different TXs before the transmission [10]. In this case, it is expected that the CSI exchange will introduce further delay and quantization noise such that it becomes necessary to study the impact of *TX dependent* CSI noise.

In order to account for TX dependent limited feedback, a *distributed CSIT* model (here referred to as D-CSIT) was introduced in [11]. In this model, TX j receives its own multi-user imperfect estimate $\hat{\mathbf{H}}^{(j)}$ on the basis of which it designs its transmit coefficients, without additional communications with the other TXs. The finite-SNR performance of regularized ZF under D-CSIT has been computed in the large system limit in [12] while heuristic robust precoding schemes have been provided in [13], [14] for practical cellular networks. In [15], [16], Interference Alignment is studied in a D-CSIT configuration and methods to reduce the required CSIT gaining at each TX are provided. The DoF achieved with delayed and local CSIT, which is hence a particular D-CSIT configuration is also studied in several works [17], [18], but the results provided are restricted to the particular local CSIT configuration considered.

In terms of DoF, it was shown in a previous work [11] that using a conventional ZF precoder (regularized or not) leads to a severe DoF degradation caused by the lack of a consistent CSI shared by the cooperating TXs. A scheme was proposed to lift the DoF in the two-user case but the scheme was relying on the particular structure of the 2-user case and could not be extended to more users. In this paper, we further build up on this concept to establish a general strategy for robust precoding in the distributed setting.

More precisely, the main findings read as follows.

- We show that optimal precoding strategies differ depending on the level of CSI accuracy available at the TXs. To this end we differentiate a weak CSIT regime where the accuracy at the least informed TX lies below a specific threshold, from an arbitrary CSIT regime where such condition does not hold.
- In the *weak CSIT regime* –which will be defined rigorously below– we obtain the surprising result that it is possible to reach the centralized outerbound, where the centralized outerbound is obtained by centralizing all CSI feedback observations across all TXs.
- In the arbitrary CSIT regime, the above D-CSIT robust scheme is extended, helping lift the DoF substantially above what is achieved by conventional ZF precoding and achieving the centralized outerbound in several key CSIT configurations.

Notations: We denote the multivariate circularly symmetric complex Gaussian distribution with zero mean and identity covariance matrix by $\mathcal{N}_{\mathbb{C}}(0,\mathbf{I})$. We use $\doteq$ to denote exponential equality, i.e., we write $f(P)\doteq P^x$ to denote $\lim_{P\to\infty}\frac{\log f(P)}{\log P}=x$. The exponential inequalities $\dot{\leq}$ and $\dot{\geq}$ are defined in the same way. We also use the shorthand notation $[K]$ to denote the set $\{1,\dots,K\}$.

II. System Model

A. Transmission Model

We study a communication system where K TXs jointly serve K Receivers (RXs) over a Network (Broadcast) MIMO channel. We consider that each TX is equipped with a single-antenna. Each RX is also equipped with a single antenna and we further assume that the RXs have perfect CSI so as to focus on the impact of the imperfect CSI on the TX side.

The signal received at RX i is written as

$$y_i=\boldsymbol{h}_i^{\mathrm{H}}\mathbf{x}+z_i \tag{1}$$

where $\boldsymbol{h}_i^{\mathrm{H}} \in \mathbb{C}^{1\times K}$ is the channel to user i, $\mathbf{x} \in \mathbb{C}^{K\times 1}$ is the transmitted multi-user signal, and $z_i \in \mathbb{C}$ is the additive noise at RX i, being independent of the channel and the transmitted signal, and distributed as $\mathcal{N}_{\mathbb{C}}(0,1)$. We further define the channel matrix $\mathbf{H} \triangleq [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]^{\mathrm{H}} \in \mathbb{C}^{K\times K}$ and the channel coefficient from TX j to RX i as $H_{i,j}$. The channel is assumed to be drawn from a continuous ergodic distribution such that all the channel matrices and all their sub-matrices are full rank with probability one.

The transmitted multi-user signal $\mathbf{x}$ is obtained from the symbol vector $\boldsymbol{s} \in \mathbb{C}^{b\times 1}$ having its elements Independently and Identically Distributed (i.i.d.) according to $\mathcal{N}_{\mathbb{C}}(0,1)$, where b is the number of independent data symbols emitted, through joint precoding.

B. *Distributed CSIT Model*

The D-CSIT setting differs from the conventional centralized one in that each TX receives a possibly different (global) CSI based on which it designs its own transmission parameters without any additional communication to the other TXs. Specifically, TX j receives the imperfect multi-user channel estimate $\hat{\mathbf{H}}^{(j)} = [\hat{\boldsymbol{h}}_1^{(j)}, \ldots, \hat{\boldsymbol{h}}_K^{(j)}]^{\mathrm{H}} \in \mathbb{C}^{K\times K}$ where $(\hat{\boldsymbol{h}}_i^{(j)})^{\mathrm{H}}$ refers to the estimate of the channel from all TXs to user i, at TX j. TX j then designs its transmit coefficients solely as a function of $\hat{\mathbf{H}}^{(j)}$ and the statistics of the channel.

Remark 1. It is critical to this work to understand well how the *distributed* CSIT setting differs from (embeds) the many different *heterogeneous* CSIT configurations studied in the literature. Indeed, an heterogeneous CSIT configuration typically refers to a *centralized* CSIT configuration (i.e., with a common channel estimate at all TXs), where each element of the channel is known with a different quality owing to specific feedback mechanisms. In contrast, the distributed setting considered here has as many different channel estimates as there are TXs (where each TX does not have access to the CSIT knowledge at the other TXs), and where different channel coefficients may also be represented with unequal quality (as in heterogeneous case). $\square$

We model the CSI uncertainty at the TXs as

$$\hat{\mathbf{H}}^{(j)} = \mathbf{H} + \sqrt{P^{-\alpha^{(j)}}}\mathbf{\Delta}^{(j)} \tag{2}$$

where $\mathbf{\Delta}^{(j)}$ is a random variable with zero mean and bounded covariance matrix. The scalar $\alpha^{(j)}$ is called the *CSIT scaling coefficient* at TX j.

Remark 2. The CSIT scaling coefficient $\alpha^{(j)}$ takes its value in $[0,1]$ where $\alpha^{(j)}=0$ is generally seen to correspond to a CSIT being useless in terms of DoF. In contrast, $\alpha^{(j)}=1$ is usually equivalent in terms of DoF to a perfect CSIT [4], [19]. □

The multi-user distributed CSIT configuration is represented through the multi-user CSIT scaling vector $\boldsymbol{\alpha}\in\mathbb{R}^K$ defined as

$$\boldsymbol{\alpha}\triangleq\begin{bmatrix}\alpha^{(1)}\\ \vdots\\ \alpha^{(K)}\end{bmatrix}. \tag{3}$$

For ease of notation, we also define the maximum value of these CSIT scaling coefficients

$$\alpha^{\max}\triangleq\max_{j\in[K]}\alpha^{(j)}. \tag{4}$$

In addition, we consider that the channel realizations and the channel estimates are drawn in an i.i.d manner. For a given transmission power P, we further assume that the conditional probability density functions also verify that

$$\max_{\mathbf{H}\in\mathbb{C}^{K\times K}}\left(p_{\mathbf{H}|\hat{\mathbf{H}}^{(1)},\ldots,\hat{\mathbf{H}}^{(K)}}(\mathbf{H})\right)\doteq\sqrt{P^{\alpha^{\max}}}. \tag{5}$$

Remark 3. This condition extends to the distributed CSIT configuration the condition provided in [4], which writes in our setting as

$$\max_{\mathbf{H}\in\mathbb{C}^{K\times K}}\left(p_{\mathbf{H}|\hat{\mathbf{H}}^{(j)}}(\mathbf{H})\right)\doteq\sqrt{P^{\alpha^{(j)}}},\qquad\forall j\in[K]. \tag{6}$$

□

Condition (5) is a mild technical assumption, which holds for the distributions usually considered.

Example 1. *We show in the Appendix that condition* (5) *is satisfied when the noise realizations* $\boldsymbol{\Delta}^{(j)}\in\mathbb{C}^{K\times K}$ *are i.i.d. according to* $\mathcal{N}_{\mathbb{C}}\left(\mathbf{0}_K,\mathbf{I}_K\right)$ *and all the CSI noise error terms* $\boldsymbol{\Delta}^{(j)}$ *are independent of each other.* ■

This D-CSIT setting is illustrated in Fig. 1.

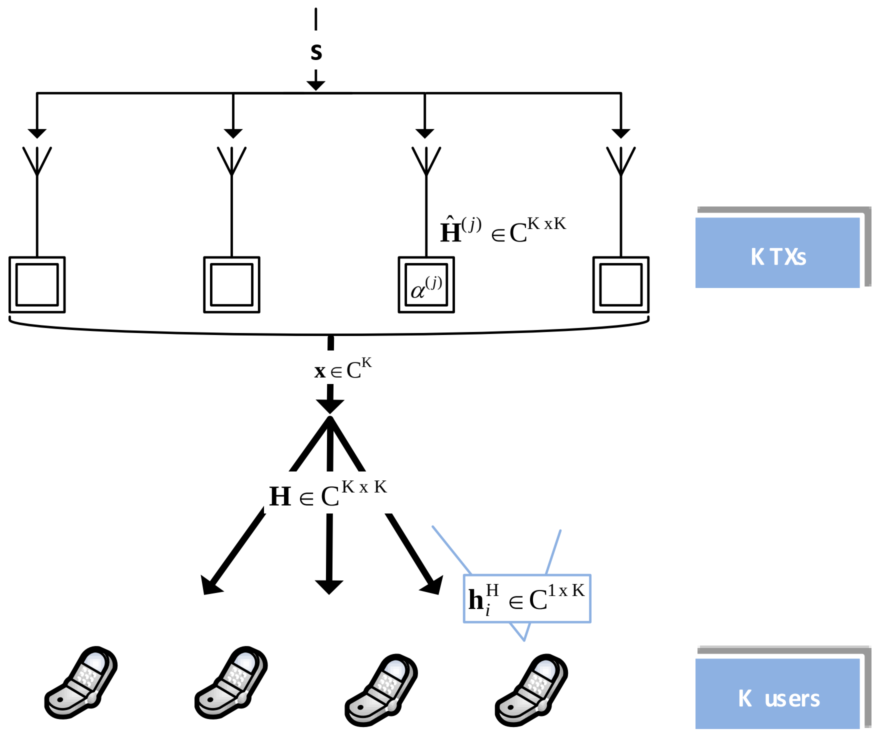

Fig. 1: Network MIMO with Distributed CSIT

C. Degrees-of-Freedom Analysis

Let us denote by $\mathcal{C}(P)$ the sum capacity of the D-CSIT network MIMO channel above. The optimal sum DoF in this distributed CSIT scenario is denoted by $\mathrm{DoF}^{\mathrm{DCSI}}(\boldsymbol{\alpha})$ and defined by

$$\mathrm{DoF}^{\mathrm{DCSI}}(\boldsymbol{\alpha}) \triangleq \lim_{P\to\infty} \frac{\mathcal{C}(P)}{\log_2(P)}. \tag{7}$$

III. A Toy Example

We start by presenting a simple transmission scheme in a Toy-example as it contains some important features of the setting and allows to convey the main intuition in a clear manner. This motivating scheme will then be improved to obtain the results in Section IV.

Let us then consider a 3-user setting in which $\alpha^{(1)} = 0.1$, $\alpha^{(2)} = 0$, and $\alpha^{(3)} = 0$. We will show how it is possible to achieve the DoF of $1 + 2\alpha^{(1)} = 1.2$, which is the value of the DoF that would be achieved in a centralized setting with TX 2 and TX 3 having received the same estimate as TX 1 [19]. In fact, it will be rigorously shown in Section IV that the DoF obtained when forming such a centralized setting is always an outerbound.

A. Encoding

The transmission scheme consists in a single channel use during which 3 private data symbols of rate $\alpha^{(1)}\log_2(P)$ bits are sent to *each* user (thus leading to 9 data symbols being sent in one channel use), while an additional common data symbol of rate $(1-\alpha^{(1)})\log_2(P)$ bits is broadcast *from TX* 1 to all users using superposition coding [20]. Note that the information contained in this common data symbol is not only composed of "fresh" information bits destined to one user, but is also composed of side information necessary for the decoding of the private data symbols, as will be detailed below.

The transmitted signal $\mathbf{x}\in\mathbb{C}^3$ is then equal to

$$\mathbf{x}=\boldsymbol{s}_1+\boldsymbol{s}_2+\boldsymbol{s}_3+\begin{bmatrix}1\\0\\0\end{bmatrix}s_0 \tag{8}$$

where

- $\boldsymbol{s}_i\in\mathbb{C}^3$ is a vector containing private data symbols destined to user i, with power $P^{\alpha^{(1)}}/3$ and rate $\alpha^{(1)}\log_2(P)$ bits.
- s_0 is the common data symbol transmitted by *TX* 1 *only* and destined to all users, with power $P-P^{\alpha^{(1)}}$ and rate $(1-\alpha^{(1)})\log_2(P)$ bits.

The signal received at user i is then equal to

$$y_i=\underbrace{H_{i,1}s_0}_{\doteq P}+\underbrace{\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{s}_1}_{\doteq P^{\alpha^{(1)}}}+\underbrace{\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{s}_2}_{\doteq P^{\alpha^{(1)}}}+\underbrace{\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{s}_3}_{\doteq P^{\alpha^{(1)}}} \tag{9}$$

where we have written under the bracket the power scaling, and where the noise term has been neglected for clarity. The received signals at the users during this transmission are illustrated in Fig. 2.

B. Interference Estimation and Quantization at TX 1

The key element of the scheme is that the common data symbol s_0 is used to convey side information allowing each user to decode their desired data symbols. More specifically, TX 1

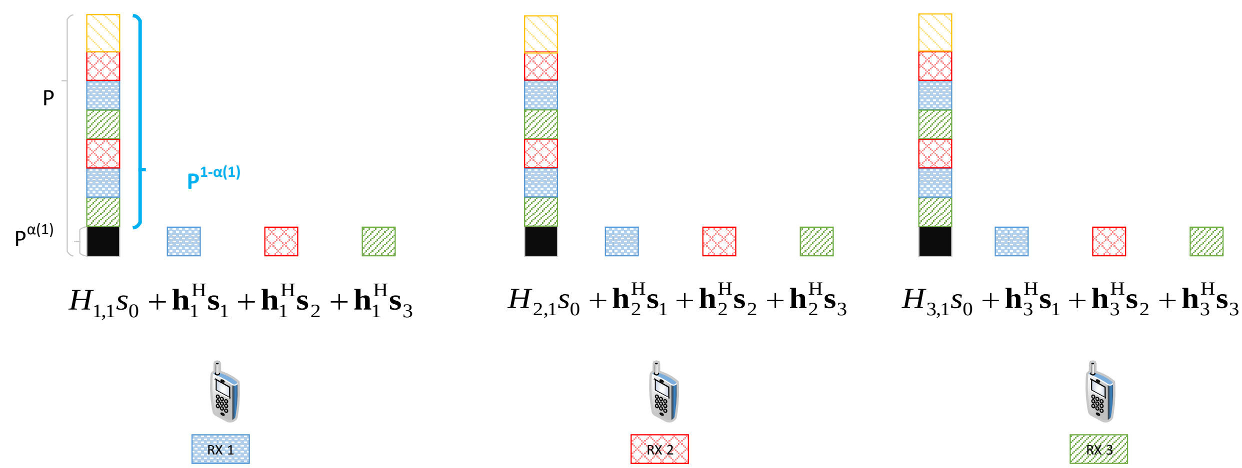

Fig. 2: Illustration of the received signals at the users.

uses its local CSIT $\hat{\mathbf{H}}^{(1)}$ to estimate the interference terms $(\hat{\boldsymbol{h}}_i^{(1)})^{\mathrm{H}}\boldsymbol{s}_k \forall i, k, k \neq i$, quantize them, and then transmit them using the common data symbol s_0. Each interference term is quantized using $\alpha^{(1)}\log_2(P)$ bits such that the quantization noise remains at the noise floor. Hence, the transmission of all the quantized estimated interference terms requires to transmit $6\alpha^{(1)}\log_2(P)$ bits.

These $6\alpha^{(1)}\log_2(P)$ bits can be transmitted via the data symbol s_0 if $6\alpha^{(1)}\log_2(P) \leq (1 - \alpha^{(1)})\log_2(P)$, which is the case for the example considered here. It the inequality is strict, the data symbol s_0 transmits some additional $(1 - 7\alpha^{(1)})\log_2(P)$ fresh information bits to any particular user.

C. *Decoding and DoF Analysis*

It remains to verify that this scheme leads to the claimed DoF. Let us consider without loss of generality the decoding at user 1 as the decoding at the other users will follow with a circular permutation of the user's indices. Note that signals at the noise floor will be systematically omitted.

Using successive decoding [20], the data symbol s_0 is decoded first, followed by the data symbol $\boldsymbol{s}_1$. The data symbol s_0 of rate of $(1-\alpha^{(1)})\log_2(P)$ bits can be decoded with a vanishing probability of error as its SINR can be seen in (9) to scale in $P^{1-\alpha^{(1)}}$.

Upon decoding s_0, the estimated interferences $(\hat{\boldsymbol{h}}_1^{(1)})^{\mathrm{H}}\boldsymbol{s}_2$ are obtained (up to the quantization noise at the noise floor, which is thus omitted). It remains to evaluate the impact over the DoF of the imperfect estimation at TX 1:

$$(\hat{\boldsymbol{h}}_1^{(1)})^{\mathrm{H}}\boldsymbol{s}_2 = \boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{s}_2 + \underbrace{\sqrt{P^{-\alpha^{(1)}}}(\boldsymbol{\delta}_1^{(1)})^{\mathrm{H}}\boldsymbol{s}_k}_{\doteq P^0}. \tag{10}$$

It follows from (10) that the interference terms can be suppressed up to the noise floor at RX 1 using the quantized estimated interference terms received.

After having subtracted the quantized interference terms, the remaining signal at user 1 is then

$$y_1 = \boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{s}_1. \tag{11}$$

Using this signal in combination with the estimated interference terms $(\hat{\boldsymbol{h}}_2^{(1)})^{\mathrm{H}}\boldsymbol{s}_1$ and $(\hat{\boldsymbol{h}}_3^{(1)})^{\mathrm{H}}\boldsymbol{s}_1$ obtained through s_0, user 1 forms a virtual received vector $\boldsymbol{y}_1^{\mathrm{v}} \in \mathbb{C}^3$ defined as

$$\boldsymbol{y}_1^{\mathrm{v}} \triangleq \begin{bmatrix} \boldsymbol{h}_1^{\mathrm{H}} \\ (\hat{\boldsymbol{h}}_2^{(1)})^{\mathrm{H}} \\ (\hat{\boldsymbol{h}}_3^{(1)})^{\mathrm{H}} \end{bmatrix} \boldsymbol{s}_1. \tag{12}$$

Each component of $\boldsymbol{y}_1^{\mathrm{v}}$ has a SINR scaling in $P^{\alpha^{(1)}}$ such that user 1 can decode with a vanishing error probability its destined 3 data symbols having each the rate $\alpha^{(1)}\log_2(P)$ bits.

Considering the 3 users, it is possible to transmit in one channel use $9\alpha^{(1)}\log_2(P)$ bits and $(1-7\alpha^{(1)})\log_2(P)$ bits (through the data symbol s_0), which yields a sum DoF of $1+2\alpha^{(1)}$.

It can be easily seen that this scheme is able to achieve the optimal DoF $1+2\alpha^{(1)}$ as long as $\alpha^{(1)} \leq 1/7$. In Section IV, this scheme will be improved by introducing some ZF precoding to reduce the amount of interference to retransmit, thus allowing for larger values of $\alpha^{(1)}$.

Remark 4. Interestingly, the above scheme builds on the principle of interference estimation, quantization and retransmission, which has already been exploited in the different context of precoding with delayed CSIT (see e.g. [6], [7], [21]). In contrast to these previous works, the distributed nature of the CSIT is exploited here such that the interference terms are estimated and transmitted from the TX having the most accurate CSIT, *during the same channel use in which the interference terms are generated*. □

IV. Main Results

As one of the key observations made in this paper, we found that the DoF behavior in a Network MIMO channel with distributed CSIT quite depends on the CSI quality regime. To this end, notions of "weak CSIT" regime and "arbitrary CSIT" regime are introduced to characterize the interval in which the CSI scaling coefficients $\alpha^{(j)}$ are allowed to take their values. We first provide results for a given weak CSIT regime and then move on to the arbitrary CSIT regime.

We define a weak-CSIT regime as follows:

Definition 1. *In the K-user Network MIMO with distributed CSIT and $K \geq 2$, we define a weak CSIT regime as comprising all the CSIT configurations satisfying that*

$$\alpha^{\max} \leq \frac{1}{1+K(K-2)}. \tag{13}$$

In the two-user case, this condition reduces to $\alpha^{(j)} \leq 1, \forall j \in \{1,2\}$, while in the three-user case, CSIT is said to be weak if $\alpha^{(j)} \leq 1/4, \forall j \in \{1,2,3\}$, and so forth. Before detailing the results of achievability in Subsections IV-B and IV-C, we start by providing an outerbound which will be useful to interpret the following results.

Remark 5. As hinted in Section III, the *weak CSIT regime* it not uniquely defined. In fact, the weak-CSIT regime considered above is larger than the one obtained using the simple scheme described in Section III. □

A. Centralized Outerbound

In the following, we prove rigorously the intuitive result that CSIT discrepancies between TXs does not improve the DoF.

Theorem 1. *In the K-user Network MIMO channel with distributed CSIT, the optimal DoF is upperbounded by the DoF achieved in a centralized CSIT configuration in which all the TXs can share perfectly their CSI estimates. Specifically, it holds that*

$$\mathrm{DoF}^{\mathrm{DCSI}}(\boldsymbol{\alpha}) \leq \mathrm{DoF}^{\mathrm{CCSI}}\left(\alpha^{\max}\right) \tag{14}$$

where $\mathrm{DoF}^{\mathrm{CCSI}}(\alpha)$ denotes the DoF achieved in a centralized CSIT configuration with the CSIT scaling coefficient α [4].

Proof. This intuitive result is proved by considering the outerbound formed by a genie-aided setting where all the channel estimates are perfectly shared between the TXs. This genie-aided configuration is a centralized setting as all TXs share the same CSI. Furthermore, because of the assumption over the probability distribution of the channel in (5), it holds that

$$p_{\mathbf{H}|\hat{\mathbf{H}}^{(1)},\ldots,\hat{\mathbf{H}}^{(K)}}(\mathbf{H}) = O\left(\sqrt{P^{\alpha^{\max}}}\right). \tag{15}$$

After having defined $\mathcal{T} \triangleq \{\hat{\mathbf{H}}^{(1)},\ldots,\hat{\mathbf{H}}^{(K)}\}$ to represent the total available CSIT, it becomes clear that it is possible to apply the outerbound derived in [4] for the centralized case. The DoF of this genie-aided centralized setting is then equal to

$$\mathrm{DoF}^{\mathrm{CCSI}}(\alpha^{\max}) = 1 + (K-1)\alpha^{\max}. \tag{16}$$

□

B. Weak CSIT Regime

Let us now consider the weak-CSIT regime defined in Definition 1.

Theorem 2. *In the K-user Network MIMO with distributed CSIT, the optimal sum DoF in the weak CSIT regime defined in Definition 1 satisfies*

$$\mathrm{DoF}^{\mathrm{DCSI}}(\boldsymbol{\alpha}) = 1 + (K-1)\alpha^{\max}. \tag{17}$$

Proof. The outerbound is obtained from Theorem 1 and a scheme achieving this outerbound is described in Section VI. □

In the weak CSIT regime defined, it is then sufficient in terms of DoF to provide a CSI estimate at a single TX. This surprising result is in strong contrast with the performance obtained using conventional ZF where the DoF is limited by the worst accuracy across all TXs (more precisely the DoF is equal to $1+(K-1)\min_j \alpha^{(j)}$ when considering successive decoding). Note that this is despite the fact that the ZF approach was recently shown to be DoF optimal for the BC under centralized CSIT setting [4].

Remark 6. In the two-user case, the weak CSIT condition considered reduces to $\alpha^{(j)} \leq 1, \forall j \in \{1,2\}$, such that the notion of weak CSIT coincides with the arbitrary regime. This is in agreement with the result in [11] that it is possible to achieve $1+(K-1)\max\left(\alpha^{(1)},\alpha^{(2)}\right)$ in the 2-user case, for any value of $\alpha^{(1)}$ and $\alpha^{(2)}$. □

C. Achievable DoF for Arbitrary CSIT Regime with $K = 3$

Going beyond the weak-CSIT regime, we present below a new transmission scheme exploiting the insights obtained in the weak CSIT regime to achieve a DoF well beyond the DoF achieved using conventional transmission schemes. Deriving a transmission scheme for any CSIT configuration and any number of users is out of the scope of this work and is the topic of ongoing research within our group.

Theorem 3. *In the* 3*-user Network MIMO with distributed CSIT and* $\alpha^{(1)} \geq \alpha^{(2)} \geq \alpha^{(3)}$*, it holds that*

$$\mathrm{DoF}^{\mathrm{DCSI}}(\boldsymbol{\alpha}) \geq \begin{cases} 1 + 2\alpha^{(1)} & \textit{if } \alpha^{(1)} \leq \frac{1}{4} \\ 3\frac{2\alpha^{(1)} - \alpha^{(2)} + 2\alpha^{(1)}\alpha^{(2)}}{4\alpha^{(1)} - \alpha^{(2)}} & \textit{if } \alpha^{(1)} \geq \frac{1}{4}. \end{cases} \tag{18}$$

Proof. See the scheme description in Section VII. □

Comparing this achievability result with the Centralized Outerbound in Theorem 1 gives the following corollary.

Corollary 1. *The lower bound provided in Theorem 3 for* $K = 3$ *users is tight in the weak CSIT regime defined in Definition 1 and in the arbitrary CSIT regime if* $\alpha^{(1)} = \alpha^{(2)}$.

The DoF achieved with the proposed scheme is illustrated in Fig. 3. For $\alpha^{(1)} \leq \frac{1}{4}$, the transmission occurs in the weak CSIT regime, and the proposed scheme coincides with the transmission scheme for the weak CSIT, as described in Section VI. In that regime, the achieved DoF only depends on the value of the best CSI scaling coefficient $\alpha^{(1)}$, while for larger values of $\alpha^{(1)}$, the DoF also depends on the value of the second best CSI scaling coefficient $\alpha^{(2)}$.

The proposed D-CSI robust transmission schemes rely on several ingredients which are (i) Active-Passive (AP-) ZF precoding, (ii) interference quantization, and (iii) superposition coding. AP-ZF was first introduced with a single so-called passive TX in [11] and we present below a non-trivial generalization to an arbitrary number of passive TXs and an arbitrary number of active TXs.

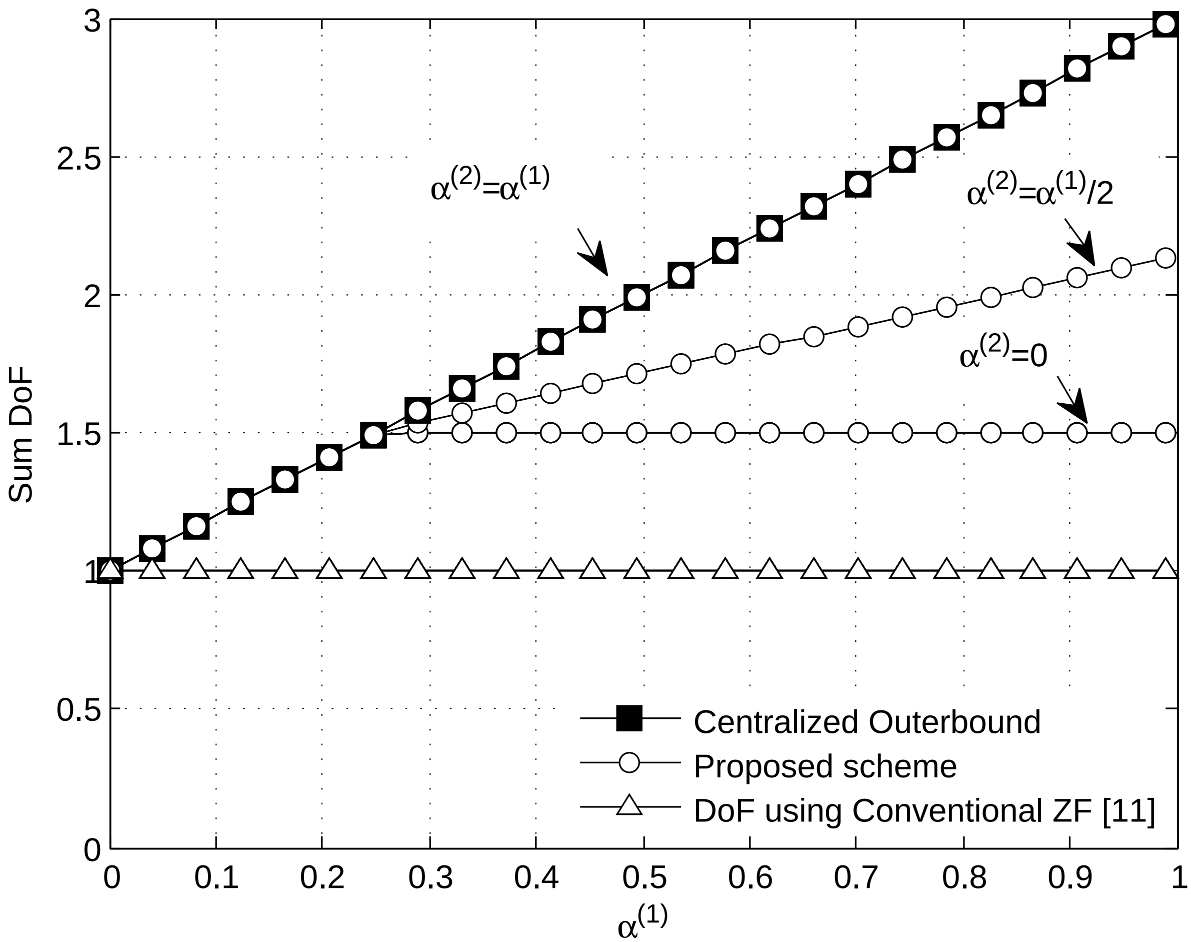

Fig. 3: Sum DoF as a function of $\alpha^{(1)}$. The DoF achieved is presented for exemplary values of $\alpha^{(2)}$, while the value of $\alpha^{(3)}$ is set to 0 to emphasize the sensitivity with respect to one estimate.

V. Preliminaries: Active-Passive Zero Forcing

Let us consider a setting in which K single-antenna TXs aim to transmit $K-n$ data symbols to one user (e.g., a user having $K-n$ antennas) while zero-forcing interference to n other single-antenna users, where $0<n<K$. Within this section, we denote the channel from the K TXs to the interfered users by $\mathbf{H}\in\mathbb{C}^{n\times K}$. We divide the TXs between so-called *active TXs* and *passive TXs* and we consider without loss of generality that the first n TXs are the active TXs while the remaining $K-n$ TXs are the passive ones.

We define the *active channel* as the channel coefficients from the active TXs, denoted by

$\mathbf{H}_{\mathrm{A}} \in \mathbb{C}^{n\times n}$, and the *passive channel* as the channel coefficients from the passive TXs, denoted by $\mathbf{H}_{\mathrm{P}} \in \mathbb{C}^{n\times(K-n)}$, such that

$$\mathbf{H} = \begin{bmatrix}\mathbf{H}_{\mathrm{A}} & \mathbf{H}_{\mathrm{P}}\end{bmatrix}. \tag{19}$$

This transmission setting considered to introduce AP-ZF is illustrated in Fig. 4.

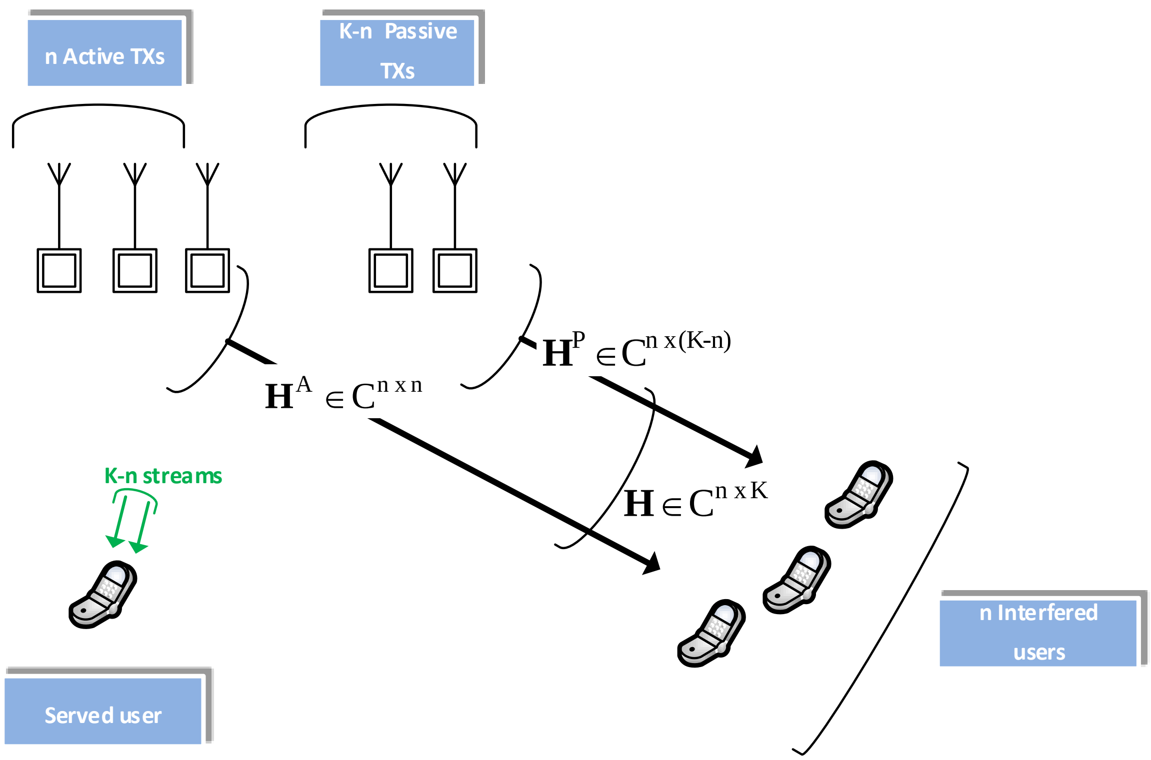

Fig. 4: Illustration of the transmission scenario considered for AP-ZF.

Turning to the CSIT configuration, we assume that an estimate $\hat{\mathbf{H}}^{(j)} \in \mathbb{C}^{n\times K}$ is available at TX j, for $j \in [K]$. We define the estimated active channel $\hat{\mathbf{H}}_{\mathrm{A}}^{(j)} \in \mathbb{C}^{n\times n}$ and the estimated passive channel $\hat{\mathbf{H}}_{\mathrm{P}}^{(j)} \in \mathbb{C}^{n\times(K-n)}$ similarly to their perfectly known counterparts.

Let us now consider the signal processing at TX j. On the basis of the available CSIT $\hat{\mathbf{H}}^{(j)}$, TX j computes $\mathbf{T}^{\mathrm{APZF}(j)} \in \mathbb{C}^{K\times K-n}$ where the part of the precoder which should be implemented at the active TXs is denoted by $\lambda^{\mathrm{APZF}}\mathbf{T}^{\mathrm{A}(j)} \in \mathbb{C}^{n\times K-n}$, where λ^{APZF} is used to satisfy an average sum power constraint (see its exact value in (23)), and is called the *active precoder* (computed at TX j) while the part of the precoder which should be implemented at the passive TXs is denoted by $\lambda^{\mathrm{APZF}}\mathbf{T}^{\mathrm{P}} \in \mathbb{C}^{(K-n)\times(K-n)}$ and is called the *passive precoder*. It then holds that

$$\mathbf{T}^{\mathrm{APZF}(j)} = \lambda^{\mathrm{APZF}} \begin{bmatrix} \mathbf{T}^{\mathrm{A}(j)} \\ \mathbf{T}^{\mathrm{P}} \end{bmatrix}. \tag{20}$$

The precoder $\mathbf{T}^{\mathrm{P}}$ is arbitrarily chosen as any full rank matrix known to all TXs while the precoder $\mathbf{T}^{\mathrm{A}(j)}$ is computed as

$$\mathbf{T}^{\mathrm{A}(j)} = -\left((\hat{\mathbf{H}}_{\mathrm{A}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{\mathrm{A}}^{(j)} + \frac{1}{P}\mathbf{I}_n\right)^{-1}(\hat{\mathbf{H}}_{\mathrm{A}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{\mathrm{P}}^{(j)}\mathbf{T}^{\mathrm{P}} \tag{21}$$

where P is the sum power of the data symbol transmitted. Note that the precoder $\mathbf{T}^{\mathrm{P}}$ is a CSI independent precoder which is commonly agreed upon by all TXs beforehand.

The effective AP-ZF precoder is implemented in a distributed manner and is denoted by $\mathbf{T}^{\mathrm{APZF}} \in \mathbb{C}^{K\times K-n}$. It is a composite version of the precoders computed at each TX and is hence given by

$$\mathbf{T}^{\mathrm{APZF}} \triangleq \lambda^{\mathrm{APZF}} \begin{bmatrix} \boldsymbol{e}_1^{\mathrm{H}}\mathbf{T}^{\mathrm{A}(1)} \\ \vdots \\ \boldsymbol{e}_n^{\mathrm{H}}\mathbf{T}^{\mathrm{A}(n)} \\ \mathbf{T}_{\mathrm{P}} \end{bmatrix} \tag{22}$$

where $\boldsymbol{e}_i \in \mathbb{C}^n$ for $i \in [n]$ is the ith row of the identity $\mathbf{I}_n$ and where the normalization coefficient λ^{APZF} is chosen as

$$\lambda^{\mathrm{APZF}} \triangleq \frac{1}{\sqrt{\mathbb{E}\left[\left\|\begin{bmatrix} -\left(\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{A}} + \frac{1}{P}\mathbf{I}_n\right)^{-1}\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{P}}\mathbf{T}^{\mathrm{P}} \\ \mathbf{T}^{\mathrm{P}} \end{bmatrix}\right\|_{\mathrm{F}}^2\right]}}. \tag{23}$$

This normalization constant λ^{APZF} requires only statistical CSI and can hence be applied at every TX. It ensures that an average normalization constraint is satisfied, i.e., that

$$\mathbb{E}\left[\|\mathbf{T}^{\mathrm{APZF}}\|_{\mathrm{F}}^2\right] = 1. \tag{24}$$

Remark 7. The design of the Active precoder in (21) is an extension of the AP-ZF precoder introduced in [11]. Intuitively, the active precoders invert the channel so as to cancel the interference generated by the passive TXs. Note also that although TX j computes the full precoder $\mathbf{T}^{\mathrm{APZF}(j)}$, only some coefficients will be effectively used for the transmission due to the distributed precoding configuration, as shown in (22). □

The key following properties can be easily shown from the precoder design.

Lemma 1. *With perfect channel knowledge at all (active) TXs, the AP-ZF precoder with n active TXs and $K-n$ passive TXs satisfies*

$$\mathbf{H}\mathbf{T}^{\mathrm{APZF}\star} \xrightarrow[P\to\infty]{} \mathbf{0}_{n\times(K-n)} \tag{25}$$

where $\mathbf{T}^{\mathrm{APZF}\star}$ denotes the AP-ZF precoder based on perfect CSIT and is given as

$$\mathbf{T}^{\mathrm{APZF}\star} \triangleq \lambda^{\mathrm{APZF}} \begin{bmatrix} \mathbf{T}^{\mathrm{A}\star} \\ \mathbf{T}^{\mathrm{P}} \end{bmatrix}. \tag{26}$$

Proof. Using the well known Resolvent identity [22, Lemma 6.1], we can write that

$$\left(\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{A}}+\frac{1}{P}\mathbf{I}_n\right)^{-1}-\left(\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{A}}\right)^{-1}=-\left(\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{A}}\right)^{-1}\left(\frac{1}{P}\mathbf{I}_n\right)\left(\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{A}}+\frac{1}{P}\mathbf{I}_n\right)^{-1}. \tag{27}$$

We can then compute the leaked interference as

$$\mathbf{H}\mathbf{T}^{\mathrm{APZF}\star} = \lambda^{\mathrm{APZF}}\mathbf{H}_{\mathrm{A}}\mathbf{T}^{\mathrm{A}\star}+\lambda^{\mathrm{APZF}}\mathbf{H}_{\mathrm{P}}\mathbf{T}^{\mathrm{P}} \tag{28}$$

$$\overset{(a)}{=} \lambda^{\mathrm{APZF}}\mathbf{H}_{\mathrm{A}}\left(\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{A}}\right)^{-1}\left(\frac{1}{P}\mathbf{I}_n\right)\left(\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{A}}+\frac{1}{P}\mathbf{I}_n\right)^{-1}\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{P}}\mathbf{T}^{\mathrm{P}} \tag{29}$$

where equality (a) follows from inserting (27) inside the AP-ZF precoder and simplifying. Letting the available power P tend to infinity, the leaked interference tends to zero. □

Lemma 2. *The AP-ZF precoder with n active TXs and $K-n$ passive TXs is of rank $K-n$.*

Proof. The passive precoder was chosen such that $\mathbf{T}^{\mathrm{P}}$ is full rank, i.e., of rank $K-n$. The precoder $\mathbf{T}^{\mathrm{A}(j)}$ is a linear combination of $\mathbf{T}^{\mathrm{P}}$ for each j, such that the effective AP-ZF precoder $\mathbf{T}^{\mathrm{APZF}}$ resulting from distributed precoding is exactly of rank $K-n$. □

Lemma 3. *If $\hat{\mathbf{H}}^{(j)} \triangleq \mathbf{H}+\sqrt{P^{-\alpha^{(j)}}}\mathbf{\Delta}^{(j)}$ for $\alpha^{(j)}\in[0,1]$ with $\mathbf{\Delta}^{(j)}$ being drawn from a continuous ergodic distribution with zero mean and bounded full rank covariance matrix, it then holds that*

$$\left\|\mathbf{H}\mathbf{T}^{\mathrm{APZF}}\right\|_{\mathrm{F}}^2 \dot{\leq} P^{-\min_{j\in[n]}\alpha^{(j)}}. \tag{30}$$

Proof. Following a similar approach as in [11], we can use once more the resolvent identity [22, Lemma 6.1] to approximate the matrix inverse and show that

$$\left\|\mathbf{T}^{\mathrm{APZF}(j)}-\mathbf{T}^{\mathrm{APZF}\star}\right\|_{\mathrm{F}}^2 \dot{\leq} P^{-\alpha^{(j)}}. \tag{31}$$

It then follows that

$$\left\|\mathbf{H}\mathbf{T}^{\mathrm{APZF}}\right\|_{\mathrm{F}}^{2} \dot{\leq} \left\|\mathbf{H}\left(\mathbf{T}^{\mathrm{APZF}}-\mathbf{T}^{\mathrm{APZF}\star}\right)\right\|_{\mathrm{F}}^{2} \tag{32}$$

$$\dot{\leq} \left\|\mathbf{H}\right\|_{\mathrm{F}}^{2}\left\|\mathbf{T}^{\mathrm{APZF}}-\mathbf{T}^{\mathrm{APZF}\star}\right\|_{\mathrm{F}}^{2} \tag{33}$$

$$\dot{\leq} \left\|\mathbf{H}\right\|_{\mathrm{F}}^{2}\left(\sum_{j=1}^{n}\left\|\mathbf{T}^{\mathrm{APZF}(j)}-\mathbf{T}^{\mathrm{APZF}\star}\right\|_{\mathrm{F}}^{2}\right) \tag{34}$$

$$\dot{\leq} P^{-\min_{j\in[n]}\alpha^{(j)}} \tag{35}$$

where (32) follows from Lemma 1 and (35) follows from Lemma 3. □

The interpretation behind this result is that the interference attenuation of AP-ZF precoding is only limited by the CSIT accuracy at the active TXs, and does not depend on the CSI accuracy at the passive TXs.

Remark 8. Interestingly, it can be seen that the number of passive TXs determines the rank of the precoder, i.e., the number of streams, and the number of active TXs determines the number of ZF constraints that are satisfied. □

VI. Weak CSIT Regime: Achievable Scheme

We now consider the weak CSIT regime and we describe the transmission scheme achieving the DoF expression given in Theorem 2. Without loss of generality, we assume that the TX with the best CSIT accuracy is TX 1, i.e., that $\alpha^{(1)}=\max_{j\in[K]}\alpha^{(j)}$.

A. Encoding

The proposed transmission scheme consists in only one channel use during which $K-1$ data symbols of rate $\alpha^{(1)}\log_2(P)$ bits are sent to *each* user (thus leading to $K(K-1)$ data symbols being sent in one channel use), while an additional data symbol of data rate $(1-\alpha^{(1)})\log_2(P)$ bits is broadcast *from TX* 1. Note that the information which is contained in this broadcast symbol is not only composed of information bits destined to one user, but is also composed of side information necessary for decoding the other –private– data symbols, as will be detailed in Subsection VI-B. The data symbol vector destined to user i is denoted by $\boldsymbol{s}_i\in\mathbb{C}^{K-1}$ while the broadcast data symbol is denoted by $s_0\in\mathbb{C}$.

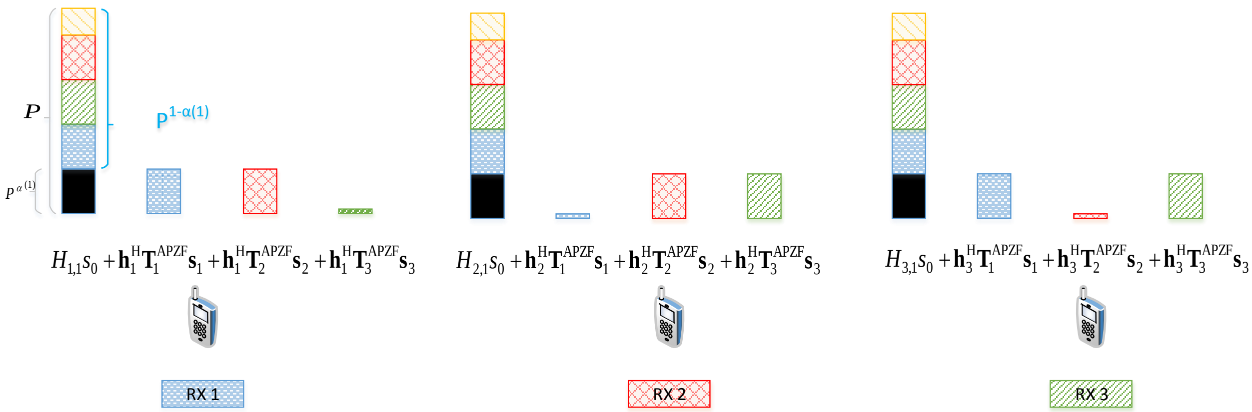

Fig. 5: Illustration of the transmission scheme for the weak CSIT regime in the case of $K = 3$ users.

The transmitted signal $\mathbf{x} \in \mathbb{C}^K$ is then equal to

$$\mathbf{x} = \begin{bmatrix} 1 \\ \mathbf{0}_{K-1\times 1} \end{bmatrix} s_0 + \sum_{i=1}^{K} \mathbf{T}_i^{\mathrm{APZF}} \boldsymbol{s}_i \tag{36}$$

where

- $\boldsymbol{s}_i \in \mathbb{C}^{K-1}$ contains $K-1$ data symbols of rate $\alpha^{(1)} \log_2(P)$ bits and power $P^{\alpha^{(1)}}/(K(K-1))$ while $\mathbf{T}_i^{\mathrm{APZF}} \in \mathbb{C}^{K\times(K-1)}$ is the AP-ZF precoder described in Section V, with TX 1 being the only active TX such that the interference are zero-forced at a single user, which we choose to be user $\overline{i+1}$ where $\overline{i+1} = i \bmod [K] + 1$.
- $\boldsymbol{s}_0 \in \mathbb{C}$ is a data symbol of rate $(1-\alpha^{(1)}) \log_2(P)$ bits and power $P - P^{\alpha^{(1)}}$, and is transmitted *from TX* 1 *only*.

The signal received at user i is then

$$y_i = \underbrace{H_{i,1} s_0}_{\doteq P} + \underbrace{\boldsymbol{h}_i^{\mathrm{H}} \mathbf{T}_i^{\mathrm{APZF}} \boldsymbol{s}_i}_{\doteq P^{\alpha^{(1)}}} + \underbrace{\boldsymbol{h}_i^{\mathrm{H}} \sum_{k=1,k\neq i,k\neq \overline{i-1}}^{K} \mathbf{T}_k^{\mathrm{APZF}} \boldsymbol{s}_k}_{\doteq (K-2)P^{\alpha^{(1)}}} + \underbrace{\boldsymbol{h}_i^{\mathrm{H}} \mathbf{T}_{\overline{i-1}}^{\mathrm{APZF}} \boldsymbol{s}_{\overline{i-1}}}_{\doteq P^0} \tag{37}$$

where the noise term has been neglected for clarity. The last term in (37) scales as P^0 following the attenuation by $P^{-\alpha^{(1)}}$ due to AP-ZF with TX 1 being the only active TX, as shown in Lemma 3. The received signals during this transmission are illustrated in Fig. 5.

B. Interference Estimation and Quantization at TX 1

The data symbol s_0 is used to convey to the users side information allowing to decode their destined data symbols. More specifically, TX 1 uses its local CSIT $\hat{\mathbf{H}}^{(1)}$ to estimate the interference terms $\hat{\boldsymbol{h}}_i^{\mathrm{H}}\mathbf{T}_k^{\mathrm{APZF}}\boldsymbol{s}_k$ for $k \neq i, k \neq \overline{i-1}$ that *is going to be* generated by the private data symbols $\boldsymbol{s}_i, \forall i \in [K]$. Each interference term has a power scaling in $P^{\alpha^{(1)}}$ such that using $\alpha^{(1)} \log_2(P)$ bits, each term can be quantized with a quantization noise at the noise floor [20].

It can be seen by inspection that there are in total $K(K-2)$ such interference terms. In the weak interference regime considered, it holds by definition that $K(K-2)\alpha^{(1)} \leq 1-\alpha^{(1)}$ such that these $K(K-2)\alpha^{(1)}\log_2(P)$ bits can be transmitted via the common data symbol s_0 of data rate $(1-\alpha^{(1)})\log_2(P)$ bits. If the previous inequality is strict, these bits are completed with information bits destined to any particular user.

C. Decoding and DoF Analysis

It remains to verify that this scheme leads to the claimed DoF. Let us consider without loss of generality the decoding at user 1 as the decoding at the other users will follow with a circular permutation of the user's indices.

Using successive decoding, the data symbol s_0 is decoded first, followed by $\boldsymbol{s}_1$. The data symbol s_0 of rate of $(1-\alpha^{(1)})\log_2(P)$ bits can be decoded with a vanishing probability of error as its SINR can be seen in (37) to scale as $P^{1-\alpha^{(1)}}$.

Upon decoding s_0, the estimated interferences $(\hat{\boldsymbol{h}}_1^{(1)})^{\mathrm{H}}\mathbf{T}_k^{\mathrm{APZF}}\boldsymbol{s}_k, k \in \{2,\dots,K-1\}$ are obtained (up to the quantization noise at the noise floor). It holds that

$$(\hat{\boldsymbol{h}}_1^{(1)})^{\mathrm{H}}\mathbf{T}_k^{\mathrm{APZF}}\boldsymbol{s}_k = \boldsymbol{h}_1^{\mathrm{H}}\mathbf{T}_k^{\mathrm{APZF}}\boldsymbol{s}_k + \underbrace{P^{-\alpha^{(1)}}(\boldsymbol{\delta}_1^{(1)})^{\mathrm{H}}\mathbf{T}_k^{\mathrm{APZF}}\boldsymbol{s}_k}_{\doteq P^0} \tag{38}$$

such that subtracting the estimated interference from the received signals can be done perfectly in terms of DoF.

Remark 9. Note that TX 1 knows perfectly the effective precoder used in the transmission as he is the only active TX. □

After having decoded s_0 and subtracted the quantized interference terms, the remaining signal at user 1 is then (up to the noise floor)

$$y_1 = \boldsymbol{h}_1^{\mathrm{H}}\mathbf{T}_1^{\mathrm{APZF}}\boldsymbol{s}_1. \tag{39}$$

The estimated interference terms $(\hat{\boldsymbol{h}}_i^{(1)})^{\mathrm{H}}\mathbf{T}_1^{\mathrm{APZF}(1)}\boldsymbol{s}_1, i = 3, \dots, K$, which have been obtained through s_0, are then used by user 1 to form a virtual received vector $\boldsymbol{y}_1^{\mathrm{v}} \in \mathbb{C}^{K-1}$ equal to

$$\boldsymbol{y}_1^{\mathrm{v}} \triangleq \begin{bmatrix} \boldsymbol{h}_1^{\mathrm{H}} \\ (\hat{\boldsymbol{h}}_3^{(1)})^{\mathrm{H}} \\ \vdots \\ (\hat{\boldsymbol{h}}_K^{(1)})^{\mathrm{H}} \end{bmatrix} \mathbf{T}_1^{\mathrm{APZF}}\boldsymbol{s}_1. \tag{40}$$

Each component of $\boldsymbol{y}_1^{\mathrm{v}}$ has a SINR scaling in $P^{\alpha^{(1)}}$ and the AP-ZF precoder is of rank $K-1$ (See Lemma 2) such that user 1 can decode its desired $K-1$ data symbols, each with the rate of $\alpha^{(1)}\log_2(P)$ bits.

Considering all users, $K(K-1)\alpha^{(1)}\log_2(P)$ bits are transmitted through the private data symbols and $(1-\alpha^{(1)}-K(K-2)\alpha^{(1)})\log_2(P)$ bits to any particular user through the common data symbol s_0. Adding the two expressions yields the claimed DoF.

VII. Arbitrary CSIT Regime for $K = 3$

In the arbitrary CSIT setting, finding an efficient transmission scheme adapted to all the possible CSIT configurations is made challenging by the fact that the CSIT configuration is characterized by K CSIT scaling coefficients. Consequently, there are many different CSIT regimes and the transmission scheme needs to be very adaptive. Therefore, we provide in the following an efficient heuristic scheme for the case of $K=3$ users, while the extension of the same ideas to K users is left for further works.

A. Main Principle

The first phase of the scheme is exactly the same as the scheme presented for the weak CSIT regime in Section VI. If the transmission occurs in the weak CSIT regime, then the scheme coincides with the scheme presented above. However, outside the weak CSIT regime, it holds that $K(K-2)\alpha^{(1)} > 1-\alpha^{(1)}$ such that all it is not possible to transmit to each user sufficient side information to decode its desired private data symbols. Consequently, we resort to a second phase of the transmission, in successive channel uses, during which the interference terms computed at TX 1 that could not yet be transmitted, are taken care of.

During the second phase, the private data symbols are transmitted using AP-ZF with 2 active TXs (TX 1 and TX 2) and the power $P^{\alpha^{(2)}}$. The use of 2 Active TXs with a lower power has for consequence that interferences generated remain at the noise floor. Indeed, AP-ZF with k active TXs is able to ZF interference at k users (See Section V), and the interference attenuation is limited by the worst accuracy across the Active TXs (See Lemma 3). As a consequence, this second phase does not generate any additional interference (in terms of DoF), such that the common data symbol can be used to retransmit solely the interference generated during the first phase.

This second phase is repeated until sufficient side information (i.e., quantized estimated interference terms) have been transmitted to the users to decode the private data symbols emitted during the first phase.

B. *Encoding*

As explained above, only the transmission during the second phase needs to be described. The transmitted signal is then given by

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} s_0 + \sum_{i=1}^{3} \boldsymbol{t}_i'^{\mathrm{APZF}} s_i' \tag{41}$$

where

- $s_i' \in \mathbb{C}$ is a private data symbol destined to user i, with power $P^{\alpha^{(2)}}/3$ and rate $\alpha^{(2)} \log_2(P)$ bits. The AP-ZF precoder $\boldsymbol{t}_i'^{\mathrm{APZF}}$ is designed with 2 active TXs –TX 1 and TX 2– and with one passive TX –TX 3–. Consequently, the interferences can be zero-forced at both interfered users (See Section V).
- $s_0' \in \mathbb{C}$ is a common data symbol, with power $P - P^{\alpha^{(2)}}$ and rate $(1 - \alpha^{(2)}) \log_2(P)$ bits, transmitted *from TX* 1 *only*.

The received signals at the users are then given by

$$\begin{aligned}
y_1' &= \underbrace{H_{1,1} s_0}_{\doteq P} + \underbrace{\boldsymbol{h}_1^{\mathrm{H}} \boldsymbol{t}_1'^{\mathrm{APZF}} s_1'}_{\doteq P^{\alpha^{(2)}}} + \underbrace{\boldsymbol{h}_1^{\mathrm{H}} \boldsymbol{t}_2'^{\mathrm{APZF}} s_2' + \boldsymbol{h}_1^{\mathrm{H}} \boldsymbol{t}_3'^{\mathrm{APZF}} s_3'}_{\doteq P^0} \\
y_2' &= \underbrace{H_{2,1} s_0}_{\doteq P} + \underbrace{\boldsymbol{h}_2^{\mathrm{H}} \boldsymbol{t}_2'^{\mathrm{APZF}} s_2'}_{\doteq P^{\alpha^{(2)}}} + \underbrace{\boldsymbol{h}_2^{\mathrm{H}} \boldsymbol{t}_1'^{\mathrm{APZF}} s_1' + \boldsymbol{h}_2^{\mathrm{H}} \boldsymbol{t}_3'^{\mathrm{APZF}} s_3'}_{\doteq P^0} \\
y_3' &= \underbrace{H_{3,1} s_0}_{\doteq P} + \underbrace{\boldsymbol{h}_3^{\mathrm{H}} \boldsymbol{t}_3'^{\mathrm{APZF}} s_3'}_{\doteq P^{\alpha^{(2)}}} + \underbrace{\boldsymbol{h}_3^{\mathrm{H}} \boldsymbol{t}_1'^{\mathrm{APZF}} s_1' + \boldsymbol{h}_3^{\mathrm{H}} \boldsymbol{t}_2'^{\mathrm{APZF}} s_2'}_{\doteq P^0}
\end{aligned} \tag{42}$$

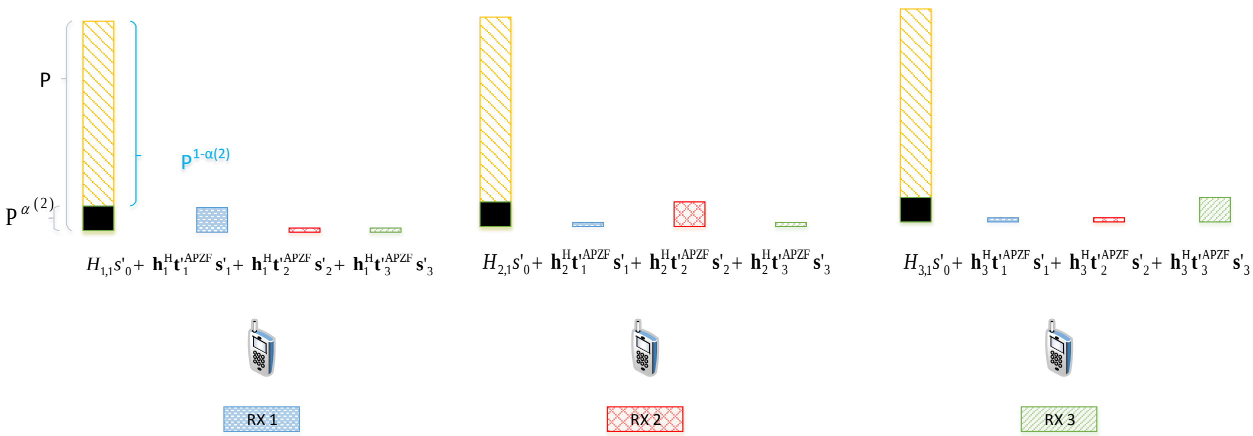

Fig. 6: Illustration of the second phase of the transmission in the arbitrary CSIT regime for $K = 3$.

where the noise realizations have been neglected. The last terms of the received signals scale as P^0 due to the attenuation by $P^{-\alpha^{(2)}}$ from AP-ZF with TX 1 and TX 2 being Active TXs (See Lemma 3).

This transmission scheme is illustrated in Fig. 6.

C. DoF Analysis

The common data symbol is decoded first and its contribution to the received signal is removed. This is possible with a vanishing probability of error as the SNR at each user scales in $P^{1-\alpha^{(2)}}$. Using successive decoding, each user can then decode with a vanishing probability of error its desired data symbol from the received signal. Thus, a sum DoF equal to $3\alpha^{(2)}$ is achieved during each channel use of the second phase.

This second phase lasts until the totality of the quantized estimated interferences have been successfully broadcast, i.e., during $\lceil(4\alpha^{(1)} - 1)/(1 - \alpha^{(2)})\rceil$ channel uses. The impact of the ceiling operator is made arbitrary small by repeating the first phase n_1 times and the second phase n_2 times, with n_1 and n_2 chosen such that $(4\alpha^{(1)} - 1)/(1 - \alpha^{(2)})$ is arbitrarily close to its next integer. Consequently, we omit in the following the ceiling operator for the sake of clarity.

A sum DoF of $6\alpha^{(1)}$ is achieved during the first phase (conditioned on the successful retransmission of the quantized estimated interferences) while a sum DoF of $3\alpha^{(2)}$ is achieved during each channel use of the second phase. As the second phase lasts for $(4\alpha^{(1)} - 1)/(1 - \alpha^{(2)})$

channel uses, the DoF achieved by the full transmission scheme is

$$\begin{aligned}\mathrm{DoF} &= \frac{6\alpha^{(1)} + \frac{4\alpha^{(1)}-1}{1-\alpha^{(2)}} 3\alpha^{(2)}}{1 + \frac{4\alpha^{(1)}-1}{1-\alpha^{(2)}}} \\ &= 3\frac{2\alpha^{(1)} - \alpha^{(2)} + 2\alpha^{(1)}\alpha^{(2)}}{4\alpha^{(1)} - \alpha^{(2)}}\end{aligned} \tag{43}$$

which is the claimed DoF, and concludes the proof.

VIII. Conclusion

We have described a new D-CSIT robust transmission schemes improving over the DoF achieved by conventional precoding approaches when faced with distributed CSIT. As a first step, we have derived an outerbound for the DoF achieved with D-CSIT, coined as the *Centralized Outerbound*, and consisting in a genie-aided setting where all the CSI versions are made available at all TXs. We have then uncovered the surprising result that in a certain “weak CSIT regime”, it is possible to achieve this Centralized Outerbound in a D-CSIT configuration with CSIT handed at a single TX. The robust precoding schemes proposed rely on new methods such as the estimation of the interference and their transmission from a single TX, and the AP-ZF precoding with multiple Passive TXs and multiple Active TXs. These new methods have a strong potential for improvement in other wireless configurations with distributed CSIT. Deriving an optimal transmission scheme for an arbitrary number of users and an arbitrary CSIT configuration is a challenging and interesting research problem.

IX. Acknowledgment

The authors are grateful to Petros Elia, Dirk Slock and Shlomo Shamai for interesting and helpful discussions.

Appendix

For the sake of completeness, we start by recalling the following result on multivariate Gaussian distribution.

Theorem 4. *[23] Let $\mathbf{X}$ and $\mathbf{Y}$ be centered and jointly Gaussian with covariance matrix $\mathbf{K}_{\mathbf{XX}}$ and $\mathbf{K}_{\mathbf{YY}}$. Assume that $\mathbf{K}_{\mathbf{YY}} > 0$. Then the conditional distribution of $\mathbf{X}$ conditional on $\mathbf{Y} = \mathbf{y}$*

is a multivariate Gaussian of mean

$$\mathbf{E}\left[\mathbf{X}\mathbf{Y}^{\mathrm{H}}\right]\mathbf{K}_{\mathbf{Y}\mathbf{Y}}^{-1}\mathbf{y}. \tag{44}$$

and covariance matrix

$$\mathbf{K}_{\mathbf{X}\mathbf{X}} - \mathbb{E}\left[\mathbf{X}\mathbf{Y}^{\mathrm{H}}\right]\mathbf{K}_{\mathbf{Y}\mathbf{Y}}^{-1}\mathbb{E}\left[\mathbf{Y}\mathbf{X}^{\mathrm{H}}\right]. \tag{45}$$

For ease of notation, we will in the following consider the vectorized versions of the channel and channel estimates, which we denote by $\boldsymbol{h}$ and $\hat{\boldsymbol{h}}^{(j)}, \forall j \in [K]$, respectively. Applying Theorem 4, the conditional distribution $p_{\boldsymbol{h}|\hat{\boldsymbol{h}}^{(1)},\dots,\hat{\boldsymbol{h}}^{(K)}}$ is multivariate Gaussian. Our goal is to compute the covariance matrix of this conditional distribution, denoted by $\mathbf{K}$. Before writing it explicitly, we introduce the shorthand notation $\bar{\mathbf{I}} \in \mathbb{C}^{K^3\times K^2}$ as

$$\bar{\mathbf{I}} \triangleq \underbrace{\begin{bmatrix}\mathbf{I}_{K^2} & \dots & \mathbf{I}_{K^2}\end{bmatrix}^{\mathrm{T}}}_{K\text{times}} \tag{46}$$

and $\mathbf{\Sigma} \in \mathbb{C}^{K^3\times K^3}$ as

$$\mathbf{\Sigma} \triangleq \begin{bmatrix} P^{-\alpha^{(1)}}\mathbf{I}_{K^2} & & \\ & \ddots & \\ & & P^{-\alpha^{(K)}}\mathbf{I}_{K^2}\end{bmatrix}. \tag{47}$$

With these notations, the covariance matrix $\mathbf{K}$ is then written using Theorem 4 as

$$\begin{aligned}
\mathbf{K} &= \mathbf{I}_{K^2} - \bar{\mathbf{I}}^{\mathrm{T}}\ \left[\bar{\mathbf{I}}\,\bar{\mathbf{I}}^{\mathrm{T}} + \mathbf{\Sigma}\right]^{-1}\ \bar{\mathbf{I}} && (48)\\
&\overset{(a)}{=} \mathbf{I}_{K^2} - \bar{\mathbf{I}}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\bar{\mathbf{I}} + \bar{\mathbf{I}}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\bar{\mathbf{I}}\left[\mathbf{I}_{K^2} + \bar{\mathbf{I}}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\bar{\mathbf{I}}\right]^{-1}\bar{\mathbf{I}}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\bar{\mathbf{I}} && (49)\\
&\overset{(b)}{=} \mathbf{I}_{K^2} - \left[\mathbf{I}_{K^2} + \bar{\mathbf{I}}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\bar{\mathbf{I}}\right]^{-1}\bar{\mathbf{I}}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\bar{\mathbf{I}} && (50)\\
&= \mathbf{I}_{K^2} - \left[\mathbf{I}_{K^2} + \sum_{j=1}^{K} P^{\alpha^{(j)}}\mathbf{I}_{K^2}\right]^{-1}\left(\sum_{j=1}^{K} P^{\alpha^{(j)}}\mathbf{I}_{K^2}\right) && (51)\\
&= \frac{1}{1+\sum_{j=1}^{K} P^{\alpha^{(j)}}}\mathbf{I}_{K^2} && (52)
\end{aligned}$$

where equality (a) follows from the Matrix Inversion Lemma [24, Chapter 3.1.1] and equality (b) follows from basic algebraic manipulations. Hence, the conditional probability density function is Gaussian with the variance of its elements scaling in $P^{-\alpha^{\max}}$ such that it satisfies that

$$\max(p_{\boldsymbol{h}|\hat{\boldsymbol{h}}^{(1)},\dots,\hat{\boldsymbol{h}}^{(K)}}(h)) \doteq \sqrt{P^{\alpha^{\max}}}. \tag{53}$$

REFERENCES

[1] G. Caire and S. Shamai (Shitz), “On the achievable throughput of a multiantenna Gaussian Broadcast Channel,” *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

[2] S. A. Jafar and A. J. Goldsmith, “Isotropic fading vector Broadcast Channels: The scalar upper bound and loss in degrees of freedom,” *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 848–857, Mar. 2005.

[3] M. W. Amos Lapidoth, Shlomo Shamai, “On the capacity of fading MIMO Broadcast Channels with imperfect transmitter side-information,” 2006. [Online]. Available: http://arxiv.org/pdf/cs/0605079.pdf

[4] A. G. Davoodi and S. A. Jafar, “Aligned image sets under channel uncertainty: Settling conjectures on the collapse of Degrees of Freedom under finite precision CSIT,” *IEEE Trans. Inf. Theo.*, vol. 62, no. 10, pp. 5603–5618, Oct. 2016.

[5] M. Maddah-Ali and D. Tse, “Completely stale transmitter channel state information is still very useful,” *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4418–4431, Jul. 2012.

[6] T. Gou and S. Jafar, “Optimal use of current and outdated channel state information: Degrees of freedom of the MISO BC with mixed CSIT,” *IEEE Communications Letters*, vol. 16, no. 7, pp. 1084–1087, Jul. 2012.

[7] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, “Degrees of freedom of time correlated MISO Broadcast Channel with delayed CSIT,” *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 315–328, Jan. 2013.

[8] R. Tandon, S. A. Jafar, S. Shamai (Shitz), and H. V. Poor, “On the synergistic benefits of alternating CSIT for the MISO BC,” *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4106–4128, 2013.

[9] J. Chen and P. Elia, “Toward the performance versus feedback tradeoff for the two-user MISO Broadcast Channel,” *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8336–8356, Dec. 2013.

[10] M. A. Maddah-Ali and U. Niesen, “Fundamental limits of Caching,” *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[11] P. de Kerret and D. Gesbert, “Degrees of freedom of the network MIMO channel with distributed CSI,” *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6806–6824, Nov. 2012.

[12] Q. Li, P. de Kerret, D. Gesbert, and N. Gresset, “Robust regularized ZF in decentralized Broadcast Channel with correlated CSI noise,” in *Proc. Allerton Conference on Communication, Control, and Computing (Allerton)*, 2015.

[13] P. de Kerret, R. Fritzsche, D. Gesbert, and U. Salim, “Robust precoding for network MIMO with hierarchical CSIT,” in *Proc. IEEE International Symposium on Wireless Communication Systems (ISWCS)*, 2014.

[14] R. Fritzsche and G. Fettweis, “Distributed robust sum rate maximization in cooperative cellular networks,” in *Proc. IEEE Workshop on Cooperative and Cognitive Mobile Networks (CoCoNet)*, 2013.

[15] X. Rao, L. Ruan, and V. K. N. Lau, “CSI feedback reduction for MIMO interference alignment,” *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4428–4437, Sept. 2013.

[16] P. de Kerret and D. Gesbert, “Interference alignment with incomplete CSIT sharing,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2563–2573, May. 2014.

[17] C. Hao and B. Clerckx, “Degrees-of-Freedom of the K-user MISO interference channel with delayed local CSIT,” in *Proc. IEEE International Conference on Communications (ICC)*, 2015.

[18] N. Lee, R. Tandon, and R. W. Heath, “Distributed space-time Interference Alignment with moderately Delayed CSIT,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1048–1059, Feb. 2015.

[19] N. Jindal, “MIMO Broadcast Channels with finite-rate feedback,” *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[20] T. Cover and A. Thomas, *Elements of information theory*. Wiley-Interscience, Jul. 2006.

[21] P. de Kerret, D. Gesbert, J. Zhang, and P. Elia, "Optimally Bridging the Gap from Delayed to Perfect CSIT in the K-user MISO BC," in *Proc. IEEE Information Theory Workshop (ITW)*, 2016.

[22] R. Couillet and M. Debbah, *Random matrix methods for wireless Communications*. Cambridge University Press, 2011.

[23] A. Lapidoth, *A foundation in Digital communication*. Cambridge University Press, 2009.

[24] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, Nov. 2012. [Online]. Available: http://matrixcookbook.com