Robust Regularized ZF in Decentralized Broadcast Channel with Correlated CSI Noise

Qianrui Li*‡, Paul de Kerret†*, David Gesbert*, and Nicolas Gresset‡
* Mobile Communication Department, Eurecom
† Télécom Bretagne, IMT, UMR CNRS 3192 Lab-STICC
‡Mitsubishi Electric R&D Centre Europe

***Abstract*—We consider in this work [1] the Distributed Channel State Information (DCSI) Broadcast Channel (BC) setting, in which the various Transmitters (TXs) compute elements of the precoder based on their individual estimates of the global multi-user channel matrix. Previous works relative to the DCSI setting assume the estimation errors at different TXs to be uncorrelated, while we consider in contrast in this work that the CSI noises can be correlated. This generalization bridges the gap between the fully distributed and the centralized setting, and offers an avenue to analyze partially centralized networks. In addition, we generalize the regularized Zero Forcing (ZF) precoding by letting each TX use a different regularization coefficient. Building upon random matrix theory tools, we obtain a deterministic equivalent for the rate achieved in the large system limit from which we can optimize the regularization coefficients at different TXs. This extended precoding scheme in which each TX applies the optimal regularization coefficient is denoted as "DCSI Regularized ZF" and we show by numerical simulations that it allows to significantly reduce the negative impact of the distributed CSI configuration and is robust to the distribution of CSI quality level across all TXs.**

***Index Terms*—Multiuser channels, Cooperative communication, MIMO, Feedback Communications**

I. Introduction

Network (or Multi-cell) MIMO methods, whereby multiple interfering TXs share user messages and allow for joint precoding, are currently considered for next generation wireless networks [1]. With perfect message and CSI sharing, the different TXs can be seen as a unique virtual multiple-antenna array serving all RXs in a multiple-antenna BC fashion [2]. Joint precoding however requires the feedback of an accurate multi-user CSI to each TX in order to achieve near optimal sum rate performance [3].

Although the case of imperfect, noisy or delayed, CSI has been considered in past work [3], [4], literature typically assumes *centralized* CSIT, i.e., that the precoding is done on the basis of a *single* imperfect channel estimate which is common at every TX. This assumption, albeit rather meaningful in the case of a broadcast with a single transmit device, can be challenged when the joint precoding is carried out across distant TXs linked by heterogeneous and imperfect backhaul links or having to communicate without backhaul (over the air) among each other. In these cases, it is expected that the CSI exchange will introduce further delay and quantization noise. It is thus practically relevant for joint precoding across distant TXs to consider a CSI setting where each TX receives its *own* channel estimate, which we denote as the *distributed CSI* configuration [5].

From an information theoretic perspective, the study of TX cooperation in the DCSI BC setting raises several intriguing and challenging questions. First, the capacity region of the BC under a DCSI setting is unknown. In [6], a rate characterization at high SNR is carried out using a Degree-of-freedom (DoF) analysis for the two TXs scenario. This study highlights the severe penalty caused by the lack of a consistent CSI shared by the cooperating TXs from a DoF point of view, when using a conventional ZF precoder. Although a new DoF-restoring decentralized precoding strategy was presented in [6] for the two TXs case, the problem of designing precoders that optimally tackle the DCSI setting at finite SNR is open for any number of TXs. Hence, an important question is how to reduce the losses due to the DCSI configuration, i.e., how to derive a DCSI robust precoding scheme.

The DCSI model used in literature represents the local degradation of the CSI by channel independent CSI noise at each TX. When the CSI is exchanged between the TXs, it is however expected that there will be some correlation between the CSI noises at different TXs. Furthermore, there is a growing interest in the industry for partially centralized networks as a practical approach to improve the performance of the network [7]. To model such settings, we extend the DCSI model by allowing for correlation between the CSI noise at the different TXs. This models is extremely general and allows to bridge the gap between centralized and decentralized setting and is a very promising tool for studying the partially centralized networks.

This work builds upon a previous work [8] in which the authors have derived a deterministic equivalent for the sum rate achieved in the DCSI setting in the large antenna regime akin to massive MIMO regime [9]. In this paper, our contributions are threefold:

- The DCSI model is now generalized such that the CSI noise at the different TXs can be correlated.
- The regularized ZF precoding is extended such that each TX can transmit using a different regularization coefficient. Regularization coefficient at each TX can be optimized so as to maximize the sum rate.

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

- An efficient, low complexity and robust distributed regularized ZF scheme is implemented by counting on the distribution of CSI quality level across *all* transmitters.

Notations: During the calculation we use the notation $x \asymp y$ to denote that $x - y \xrightarrow[K,M\to\infty]{a.s.} 0$.

II. System Model

A. Transmission Model

We study a communication system where n TXs serve jointly K Receivers (RXs) over a network MIMO channel. Each TX is equipped with M_{TX} antennas such that the total number of transmit antennas is denoted by $M = nM_{\mathrm{TX}}$. Every RX is equipped with a single antenna. We assume that the ratio of transmit antennas with respect to the number of users is fixed and given by

$$\beta = \frac{M}{K} \geq 1. \tag{1}$$

We further assume that the RXs have perfect CSI so as to focus on the imperfect CSI feedback and exchange among the TXs. We consider that the RXs treat interference as noise. The channel from the n TXs to the K RXs is represented by the multi-user channel matrix $\mathbf{H} \in \mathbb{C}^{K\times M}$.

Considering linear precoding, the transmission is then described as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{h}_1^{\mathrm{H}} \\ \vdots \\ \boldsymbol{h}_K^{\mathrm{H}} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_K \end{bmatrix} \tag{2}$$

where $\boldsymbol{h}_i^{\mathrm{H}} = \boldsymbol{e}_i^{\mathrm{H}}\mathbf{H} \in \mathbb{C}^{1\times M}$ is the channel from all transmit antennas to RX i with its elements i.i.d. as $\mathcal{N}_{\mathbb{C}}(0,1)$. $\boldsymbol{\eta} = [\eta_1,\ldots,\eta_K]^{\mathrm{T}} \in \mathbb{C}^{K\times 1}$ is the normalized Gaussian noise with its elements i.i.d. as $\mathcal{N}_{\mathbb{C}}(0,1)$.

The transmitted multi-user signal $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is obtained from the symbol vector $\boldsymbol{s} = [\boldsymbol{s}_1^{\mathrm{T}},\ldots,\boldsymbol{s}_K^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{K\times 1}$ with its elements i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ as

$$\mathbf{x} = \mathbf{T}\boldsymbol{s} = \begin{bmatrix} \boldsymbol{t}_1,\ldots,\boldsymbol{t}_K \end{bmatrix} \begin{bmatrix} \boldsymbol{s}_1 \\ \vdots \\ \boldsymbol{s}_K \end{bmatrix} \tag{3}$$

with $\mathbf{T} \in \mathbb{C}^{M\times K}$ being the *multi-user* precoder and $\boldsymbol{t}_i = \mathbf{T}\boldsymbol{e}_i \in \mathbb{C}^{M\times 1}$ being the beamforming vector used to transmit to RX i. We consider for clarity the sum power constraint $\|\mathbf{T}\|_{\mathrm{F}}^2 = P$.

Our main figure-of-merit is the average rate per user

$$\mathrm{R} = \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}\left[\log_2\left(1+\mathrm{SINR}_k\right)\right] \tag{4}$$

where SINR_k denotes the Signal-to-Interference and Noise Ratio (SINR) at RX k and is defined as

$$\mathrm{SINR}_k = \frac{\left|\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{t}_k\right|^2}{1+\sum_{\ell=1,\ell\neq k}^{K}\left|\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{t}_\ell\right|^2}. \tag{5}$$

B. Distributed CSIT Model

In the distributed CSIT model studied here, each TX receives its own CSI based on which it designs its transmission parameters without any additional communication to the other TXs. The actual feedback and exchange mechanism based on which the TXs receive the multi-user channel estimate is completely arbitrary, yet it can itself be the topic of some interesting trade-off and optimization beyond the scope of this paper [10], [11].

Specifically, TX j receives the multi-user channel estimate $\hat{\mathbf{H}}^{(j)} = \begin{bmatrix} \hat{\boldsymbol{h}}_1^{(j)} & \ldots & \hat{\boldsymbol{h}}_K^{(j)} \end{bmatrix}^{\mathrm{H}} \in \mathbb{C}^{K\times M}$ and designs its transmit coefficients $\mathbf{x}_j \in \mathbb{C}^{M_{\mathrm{TX}}\times 1}$ solely as a function of $\hat{\mathbf{H}}^{(j)}$ and the statistics of the channel. We model the imperfect channel estimate for RX k at TX j as

$$\hat{\boldsymbol{h}}_k^{(j)} \triangleq \sqrt{1-(\sigma_k^{(j)})^2}\boldsymbol{h}_k + \sigma_k^{(j)}\boldsymbol{\delta}_k^{(j)} \tag{6}$$

with $\boldsymbol{\delta}_k^{(j)} \in \mathbb{C}^{M\times 1}$ having its elements i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$, and all the CSI noise error terms $\boldsymbol{\delta}_k^{(j)}$ being independent of the channel and jointly Gaussian such that

$$\mathbb{E}\left[\boldsymbol{\delta}_k^{(j)}(\boldsymbol{\delta}_k^{(j')})^{\mathrm{H}}\right] = (\rho_k^{(j,j')})^2\mathbf{I}_M \tag{7}$$

where the parameters $\rho_k^{(j,j')}$ is the *CSI noise correlation factor* between TX j and TX j'. Note that $\rho_k^{(j,j)} = 1$.

$\sigma_k^{(j)} \in [0,1]$ is a parameter indicating the accuracy of the kth user channel available at TX j.

This distributed CSI model allowing for correlation between the errors at the different TXs is very general and is one of the main contributions of this paper. This model allows to bridge the gap between the two extreme configuration: Distributed with *independent* CSI errors and centralized CSI. Indeed, the CSI configuration where

$$\sigma_k^{(j)} = \sigma_k^{(j')},\ \rho_k^{(j,j')} = 1, \forall j,j' = 1,\ldots,n,\ k = 1,\ldots,K \tag{8}$$

corresponds to the centralized CSI configuration [3], [12] while taking $\rho_k^{(j,j')} = 0$ is the distributed CSI configuration as previously studied in the literature [6]. The major interest of this model is to allow for all the intermediate CSI configuration where the CSI can then be seen as *partially centralized*. This is particularly adapted to model *imperfect* CSI exchange where delay and/or imperfections are introduced.

C. Regularized Zero Forcing with Distributed CSIT

We address the performance of a classical MISO broadcast precoder, namely *regularized ZF* [13], [14], when faced with the above distributed CSIT model in the large system regime. Hence, the precoder designed at TX j is assumed to take the form

$$\mathbf{T}_{\mathrm{rZF}}^{(j)} = \left((\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)} + M\alpha^{(j)}\mathbf{I}_M\right)^{-1}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\frac{\sqrt{P}}{\sqrt{\Psi^{(j)}}} \tag{9}$$

with regularization factor $\alpha^{(j)} > 0$. We also define

$$\mathbf{C}^{(j)} = \frac{(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}}{M} + \alpha^{(j)}\mathbf{I}_M \tag{10}$$

such that the precoder at TX j can be rewritten as

$$\mathbf{T}_{\mathrm{rZF}}^{(j)}=\frac{1}{M}(\mathbf{C}^{(j)})^{-1}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\frac{\sqrt{P}}{\sqrt{\Psi^{(j)}}}. \tag{11}$$

The scalar $\Psi^{(j)}$ corresponds to the power normalization at TX j. Hence, it holds that

$$\Psi^{(j)}=\|\left((\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}+M\alpha^{(j)}\mathbf{I}_M\right)^{-1}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\|_{\mathrm{F}}^2. \tag{12}$$

Upon concatenation of all TX's precoding vectors, the effective global precoder denoted by $\mathbf{T}_{\mathrm{rZF}}^{\mathrm{DCSI}}$, is equal to

$$\mathbf{T}_{\mathrm{rZF}}^{\mathrm{DCSI}}=\begin{bmatrix}\mathbf{E}_1^{\mathrm{H}}\mathbf{T}_{\mathrm{rZF}}^{(1)}\\ \mathbf{E}_2^{\mathrm{H}}\mathbf{T}_{\mathrm{rZF}}^{(2)}\\ \vdots\\ \mathbf{E}_n^{\mathrm{H}}\mathbf{T}_{\mathrm{rZF}}^{(n)}\end{bmatrix} \tag{13}$$

where $\mathbf{E}_j^{\mathrm{H}}\in\mathbb{C}^{M_{\mathrm{TX}}\times M}$ is defined as

$$\mathbf{E}_j^{\mathrm{H}}=\begin{bmatrix}\mathbf{0}_{M_{\mathrm{TX}}\times(j-1)M_{\mathrm{TX}}} & \mathbf{I}_{M_{\mathrm{TX}}} & \mathbf{0}_{M_{\mathrm{TX}}\times(n-j)M_{\mathrm{TX}}}\end{bmatrix}. \tag{14}$$

We furthermore denote the kth column of $\mathbf{T}_{\mathrm{rZF}}^{\mathrm{DCSI}}$ (used to serve RX k) by $\boldsymbol{t}_{\mathrm{rZF},k}^{\mathrm{DCSI}}$.

Remark 1. It is important to note that TX j transmits using the regularization coefficient $\alpha^{(j)}$, which means that all the TXs may not use the same regularization coefficient. □

D. Naive Regularized ZF

When TX j is not aware of the differences between imperfection levels at all TXs, it chooses its regularization parameter on the basis of its own CSI quality, which yields a naive (suboptimal) precoding scheme. In general, this can be done using a linear search and the large system approximation is given in [12].

In the particular case with same quality across all links, i.e., $\sigma_k^{(j)}=\sigma^{(j)}$, a closed form solution exists [12]

$$\alpha^{(j),\mathrm{CCSI}}=\frac{1+(\sigma^{(j)})^2P}{1-(\sigma^{(j)})^2}\frac{1}{\beta P}. \tag{15}$$

Naive Regularized ZF will be our benchmark precoding scheme. However, as it neglects the distributed nature of CSI at different TXs, a better scheme can be designed below.

E. DCSI Robust Regularized ZF

In the DCSI configuration, each TX is aware of the statistics of the estimates at *all* the TXs. Hence, it takes this knowledge into account when choosing its regularization coefficient. As the performance depends on the regularization coefficients of all the TXs, this means that each TX effectively solves

$$(\alpha^{(1),\star},\dots,\alpha^{(n),\star})=\underset{(\alpha^{(1)},\dots,\alpha^{(n)})}{\operatorname{argmax}}\ R. \tag{16}$$

To reduce complexity, a simpler design can be considered where the *same* regularization coefficient is used at each TX but still on the basis of CSI quality levels at *all* TX. Hence, we will study

$$\alpha^{\star}=\underset{(\alpha,\dots,\alpha)}{\operatorname{argmax}}\,R. \tag{17}$$

Although the finite SNR rate analysis under the precoding structure (13) and the distributed CSI model in (6) is challenging in the general case because of the dependency of one user performance on all the channel estimates, some useful insights can be obtained in the large antenna regime, as shown below.

III. Deterministic Equivalent of the SINR

They key of our approach consists in studying a large antenna regime where the number of transmit antennas and the number of receive antennas jointly grow large with a fixed ratio, thus allowing to use efficient tools from the field of Random Matrix Theory (RMT). The Stieltjes transform has proven very useful in the analysis of wireless networks [See [12], [15]–[18] among others] and we will also follow this approach. Hence, our approach is built on the following fundamental result.

Theorem 1. *[17], [19] Consider the resolvent matrix* $\mathbf{Q}=\left(\frac{\mathbf{H}^{\mathrm{H}}\mathbf{H}}{M}+\alpha\mathbf{I}_M\right)^{-1}$ *with the matrix* $\mathbf{H}\in\mathbb{C}^{K\times M}$ *with* $\beta=M/K$ *having its elements i.i.d. as* $\mathcal{N}_{\mathbb{C}}(0,1)$ *and* $\alpha>0$. *Let us note that the equation*

$$x=\left(\alpha+\frac{1}{\beta\,(1+x)}\right)^{-1} \tag{18}$$

admits a unique fixed point which we will denote by δ *in the following and can be obtained in closed form as*

$$\delta=\frac{\beta-1-\alpha\beta+\sqrt{(\alpha\beta-\beta+1)^2+4\alpha\beta^2}}{2\alpha\beta}. \tag{19}$$

Let us furthermore define

$$\mathbf{Q}_o\triangleq\left(\alpha\mathbf{I}_M+\frac{\mathbf{I}_M}{\beta\,(1+\delta)}\right)^{-1} \tag{20}$$

and let the matrix $\mathbf{U}$ *be any matrix with bounded spectral norm. Then,*

$$\frac{1}{M}\operatorname{tr}(\mathbf{U}\mathbf{Q})-\frac{1}{M}\operatorname{tr}(\mathbf{U}\mathbf{Q}_o)\xrightarrow[K,M\to\infty]{a.s.}0. \tag{21}$$

Using this theorem, we can now state our main result. Yet, following the shorthand notation used in [12], we introduce the following:

$$c_{0,k}^{(j)}\triangleq 1-(\sigma_k^{(j)})^2,\quad c_{1,k}^{(j)}\triangleq(\sigma_k^{(j)})^2,\quad c_{2,k}^{(j)}\triangleq\sigma_k^{(j)}\sqrt{1-(\sigma_k^{(j)})^2}. \tag{22}$$

Theorem 2. *Considering the D-CSI network MIMO channel described in Section II, then* $\mathrm{SINR}_k-\mathrm{SINR}_k^o\to 0$ *with* SINR_k^o *defined as*

$$\mathrm{SINR}_k^o\triangleq\frac{P\left(\frac{1}{n}\sum_{j=1}^{n}\sqrt{\frac{1-(\sigma_k^{(j)})^2}{\Gamma_{j,j}^o}}\frac{\delta^{(j)}}{1+\delta^{(j)}}\right)^2}{1+I_k^o} \tag{23}$$

with $I_k^o \in \mathbb{R}$ defined as

$$I_k^o \triangleq P - P\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{\Gamma^o_{j,j'}}{\sqrt{\Gamma^o_{j,j}\Gamma^o_{j',j'}}}\Bigg[\frac{2c_{0,k}^{(j)}}{n^2}\frac{\delta^{(j)}}{1+\delta^{(j)}} - \frac{\left((\rho_k^{(j,j')})^2 c_{2,k}^{(j)}c_{2,k}^{(j')} + c_{0,k}^{(j)}c_{0,k}^{(j')}\right)\delta^{(j)}\delta^{(j')}}{n^2\left(1+\delta^{(j)}\right)\left(1+\delta^{(j')}\right)}\Bigg] \tag{24}$$

where $\delta^{(j)}$ is obtained from Theorem 1 using $\alpha^{(j)}$ while $\Gamma^o_{j,j'} \in \mathbb{R}$ is defined as

$$\Gamma^o_{j,j'} \triangleq \frac{\frac{1}{M}\sum_{\ell=1}^{K}\sqrt{c_{0,\ell}^{(j)}c_{0,\ell}^{(j')}} + \sqrt{c_{1,\ell}^{(j)}c_{1,\ell}^{(j')}}(\rho_\ell^{(j,j')})^2}{\frac{1+\delta^{(j)}}{\delta^{(j)}}\frac{1+\delta^{(j')}}{\delta^{(j')}} - \frac{\sum_{\ell=1}^{K}\left(\sqrt{c_{0,\ell}^{(j)}c_{0,\ell}^{(j')}}+\sqrt{c_{1,\ell}^{(j)}c_{1,\ell}^{(j')}}(\rho_\ell^{(j,j')})^2\right)^2}{M}}. \tag{25}$$

The above deterministic SINR expression is very generic and encompasses important sub-cases listed below.

Corollary 1. *Let us consider $n > 1$ TXs with $\sigma_k^{(j)} = \sigma_k^{(j')} = \sigma_k$ and $\rho_k^{(j,j')} = 1, \forall j, j' = 1, \dots, n,\ k = 1, \dots, K$, which corresponds to the D-CSI setting with IDentical channel estimate at each TX (ID-DCSI).*

Let $\alpha^{(j)} = \alpha^{(j')} = \alpha, \forall j, j' = 1, \dots, n$, which indicates that each TX implements the RZF precoder with the same regularization parameter (such that $\delta^{(j)} = \delta^{(j')} = \delta, \forall j, j' = 1, \dots, n$).

The deterministic SINR becomes:

$$\mathrm{SINR}_k^{ID-DCSI,o} = \frac{(1-\sigma_k^2)\delta^2}{\Gamma^o\left(1-\sigma_k^2+(1+\delta)^2\sigma_k^2+\frac{(1+\delta)^2}{P}\right)}$$

with

$$\Gamma^o = \frac{\delta^2}{\beta\left(1+\delta\right)^2-\delta^2}.$$

The above expression matches with the results for the C-CSI expression in [12] just as expected.

Corollary 2. *Let us consider $n > 1$ Txs and DCSI with EQual $\alpha^{(j)} = \alpha^{(j')} = \alpha, \forall j, j' = 1, \dots, n$ (EQ-DCSI). The deterministic SINR becomes:*

$$\mathrm{SINR}_k^{EQ-DCSI,o} = \frac{\frac{P}{\Gamma^o}\left(\frac{1}{n}\sum_{j=1}^{n}\sqrt{c_{0,k}^{(j)}}\right)^2\frac{\delta^2}{(1+\delta)^2}}{I_k^{EQ-DCSI,o}+1}$$

with

$$I_k^{EQ-DCSI} = P - P\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{\delta\Gamma^o_{j,j'}}{n^2(1+\delta)^2\Gamma^o}\cdot\Big[2c_{0,k}^{(j)} + \delta\left(2c_{0,k}^{(j)} - c_{0,k}^{(j)}c_{0,k}^{(j')} - (\rho_k^{(j,j')})^2 c_{2,k}^{(j)}c_{2,k}^{(j')}\right)\Big]$$

If $\rho_k^{(j,j')} = 0, \forall j \neq j',\ j, j' = 1, \dots, n,\ k = 1, \dots, K$, the above expression matches with the results for the uncorrelated DCSI expression in [8].

IV. Proof of Theorem 2

Due to space limitation, we will omit some steps of the derivations. The omitted steps are presented in a longer version paper [20]. The following proof significantly extends the proof of the Theorem in [8] and is built upon the approaches in [12] and [17]. We also make extensive use of classical RMT lemmas recalled in the Appendix.

As a preliminary step, we introduce the following notations.

$$\mathbf{C}_{[k]}^{(j)} = \frac{\hat{\mathbf{H}}_{[k]}^{(j)}(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}}}{M} + \alpha^{(j)}\mathbf{I}_M, \qquad \forall j$$

with

$$(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}} = \begin{bmatrix}\hat{\boldsymbol{h}}_1^{(j)} & \dots & \hat{\boldsymbol{h}}_{k-1}^{(j)} & \hat{\boldsymbol{h}}_{k+1}^{(j)} & \dots & \hat{\boldsymbol{h}}_K^{(j)}\end{bmatrix}, \qquad \forall j,$$

which is the matrix $(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}$ with the kth column removed.

A. Deterministic Equivalent for $\Psi^{(j)}$

A deterministic equivalent for $\Psi^{(j)}$ can be found in [12], or can be obtained using Lemma 5 with $\sigma^{(j)} = \sigma^{(j')} = 0$:

$$\Psi^{(j)} \asymp \Gamma^o_{j,j}$$

with $\Gamma^o_{j,j}$ defined in (25). As expected, this deterministic equivalent does not depend on $\sigma^{(j)}$.

It should be noted that when the system becomes large, the effective global precoder $\mathbf{T}_{\mathrm{rZF}}^{\mathrm{DCSI}}$ satisfies the total power constraint, since

$$\begin{aligned}
&\| \mathbf{T}_{\mathrm{rZF}}^{\mathrm{DCSI}} \|_{\mathrm{F}}^2 \\
&= \sum_{j=1}^{n}\mathrm{tr}\left(\mathbf{E}_j^{\mathrm{H}}\mathbf{T}_{\mathrm{rZF}}^{(j)}(\mathbf{T}_{\mathrm{rZF}}^{(j)})^{\mathrm{H}}\mathbf{E}_j\right) \\
&= \sum_{j=1}^{n}\frac{P}{\Gamma^o_{j,j}}\mathrm{tr}\left(\frac{1}{M^2}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)})^{-1}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}(\mathbf{C}^{(j)})^{-1}\right) \\
&\overset{(a)}{\asymp} \sum_{j=1}^{n}\frac{P}{\Gamma^o_{j,j}}\frac{\Gamma^o_{j,j}}{n} \\
&= P
\end{aligned}$$

where (a) follows from Lemma 5, the isotropy of the channel, and $\Psi^{(j)} \asymp \Gamma^o_{j,j'}$.

B. Deterministic Equivalent for $\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{t}_{\mathrm{rZF},k}^{\mathrm{DCSI}}$:

For the desired signal part at RX k, we can write

$$\begin{aligned}\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{t}_{\mathrm{rZF},k}^{\mathrm{DCSI}} &= \sum_{j=1}^{n}\frac{1}{M}\frac{\sqrt{P}}{\sqrt{\Psi^{(j)}}}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)})^{-1}\hat{\boldsymbol{h}}_k^{(j)} \\ &\overset{(a)}{\asymp} \sqrt{P}\sum_{j=1}^{n}\sqrt{\frac{1}{\Gamma_{j,j}^{o}}}\frac{\frac{1}{M}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}_{[k]}^{(j)})^{-1}\hat{\boldsymbol{h}}_k^{(j)}}{1+\frac{1}{M}(\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}(\mathbf{C}_{[k]}^{(j)})^{-1}\hat{\boldsymbol{h}}_k^{(j)}} \\ &\overset{(b)}{\asymp} \sqrt{P}\sum_{j=1}^{n}\sqrt{\frac{1-(\sigma_k^{(j)})^2}{\Gamma_{j,j}^{o}}}\frac{\frac{1}{M}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}_{[k]}^{(j)})^{-1}\boldsymbol{h}_k}{1+\frac{1}{M}(\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}(\mathbf{C}_{[k]}^{(j)})^{-1}\hat{\boldsymbol{h}}_k^{(j)}} \\ &\overset{(c)}{\asymp} \sqrt{P}\sum_{j=1}^{n}\sqrt{\frac{1-(\sigma_k^{(j)})^2}{\Gamma_{j,j}^{o}}}\frac{\frac{1}{M}\operatorname{tr}\left(\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}_{[k]}^{(j)})^{-1}\right)}{1+\frac{1}{M}\operatorname{tr}\left((\mathbf{C}_{[k]}^{(j)})^{-1}\right)} \\ &\overset{(d)}{\asymp} \frac{\sqrt{P}}{n}\sum_{j=1}^{n}\sqrt{\frac{1-(\sigma_k^{(j)})^2}{\Gamma_{j,j}^{o}}}\frac{\delta^{(j)}}{1+\delta^{(j)}}.\end{aligned}$$

Equality (a) follows then from Lemma 1 and the use of the deterministic equivalent derived for $\Psi^{(j)}$, (b) from Lemma 3, (c) from Lemma 2, (d) from Lemma 4 and the fundamental Theorem 1. Note that $\delta^{(j)}$ can be calculated as illustrated in Theorem 1 and can in fact even be calculated in closed form. Taking the squared absolute value provides the desired expression.

C. Deterministic Equivalent for the Interference Term

According to the definition of $\mathbf{T}^{\mathrm{DCSI}}$ in (13) and replacing $\Psi^{(j)}$ by its deterministic equivalent yields

$$\begin{aligned}\mathcal{I}_k &= \sum_{\ell=1,\ell\neq k}^{K}|\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{t}_{\mathrm{rZF},\ell}^{\mathrm{DCSI}}|^2 \\ &= \boldsymbol{h}_k^{\mathrm{H}}\mathbf{T}_{\mathrm{rZF}}^{\mathrm{DCSI}}(\mathbf{T}_{\mathrm{rZF}}^{\mathrm{DCSI}})^{\mathrm{H}}\boldsymbol{h}_k - \boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{t}_{\mathrm{rZF},k}^{\mathrm{DCSI}}(\boldsymbol{t}_{\mathrm{rZF},k}^{\mathrm{DCSI}})^{\mathrm{H}}\boldsymbol{h}_k \\ &= \frac{1}{M^2}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{P}{\sqrt{\Psi^{(j)}\Psi^{(j')}}}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)})^{-1}(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}} \\ &\cdot\hat{\mathbf{H}}_{[k]}^{(j')}(\mathbf{C}^{(j')})^{-1}\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{h}_k \\ &\asymp \frac{P}{M^2}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{1}{\sqrt{\Gamma_{j,j}^{o}\Gamma_{j',j'}^{o}}}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}_{[k]}^{(j)})^{-1}(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}} \\ &\qquad\cdot\hat{\mathbf{H}}_{[k]}^{(j')}(\mathbf{C}^{(j')})^{-1}\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{h}_k \\ &+\frac{P}{M^2}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{1}{\sqrt{\Gamma_{j,j}^{o}\Gamma_{j',j'}^{o}}}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}} \\ &\cdot\left((\mathbf{C}^{(j)})^{-1}-(\mathbf{C}_{[k]}^{(j)})^{-1}\right)(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{[k]}^{(j')}(\mathbf{C}^{(j')})^{-1}\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{h}_k.\end{aligned}$$

Applying Lemma 5, we obtain a deterministic equivalent for the first summation term. For the second summation term, we use the following relation

$$\begin{aligned}&(\mathbf{C}^{(j)})^{-1}-(\mathbf{C}_{[k]}^{(j)})^{-1} \\ &=(\mathbf{C}^{(j)})^{-1}\left(\mathbf{C}_{[k]}^{(j)}-\mathbf{C}^{(j)}\right)(\mathbf{C}_{[k]}^{(j)})^{-1} \\ &=-\frac{1}{M}(\mathbf{C}^{(j)})^{-1}\Big(c_{0,k}^{(j)}\boldsymbol{h}_k\boldsymbol{h}_k^{\mathrm{H}}+c_{1,k}^{(j)}\boldsymbol{\delta}_k^{(j)}(\boldsymbol{\delta}_k^{(j)})^{\mathrm{H}} \\ &\qquad+c_{2,k}^{(j)}\boldsymbol{\delta}_k^{(j)}\boldsymbol{h}_k^{\mathrm{H}}+c_{2,k}^{(j)}\boldsymbol{h}_k(\boldsymbol{\delta}_k^{(j)})^{\mathrm{H}}\Big)(\mathbf{C}_{[k]}^{(j)})^{-1} \end{aligned}\tag{26}$$

It is important to note that $c_{0,k}^{(j)}c_{1,k}^{(j)}=(c_{2,k}^{(j)})^2$ and $c_{0,k}^{(j)}+c_{1,k}^{(j)}=1$ as these relations will be used several times through the proof in order to simplify the expressions obtained. Inserting (26) in $\mathcal{I}_k$ yields

$$\begin{aligned}\mathcal{I}_k &\asymp \frac{P}{M^2}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{1}{\sqrt{\Gamma_{j,j}^{o}\Gamma_{j',j'}^{o}}}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}_{[k]}^{(j)})^{-1}(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}} \\ &\qquad\cdot\hat{\mathbf{H}}_{[k]}^{(j')}(\mathbf{C}^{(j')})^{-1}\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{h}_k \\ &-\frac{P}{M^3}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)})^{-1}\left[\boldsymbol{h}_k c_{0,k}^{(j)}\boldsymbol{h}_k^{\mathrm{H}}\right]}{\sqrt{\Gamma_{j,j}^{o}\Gamma_{j',j'}^{o}}} \\ &\qquad\cdot(\mathbf{C}_{[k]}^{(j)})^{-1}(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{[k]}^{(j')}(\mathbf{C}^{(j')})^{-1}\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{h}_k \\ &-\frac{P}{M^3}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)})^{-1}\left[\boldsymbol{\delta}_k^{(j)}c_{1,k}^{(j)}(\boldsymbol{\delta}_k^{(j)})^{\mathrm{H}}\right]}{\sqrt{\Gamma_{j,j}^{o}\Gamma_{j',j'}^{o}}} \\ &\qquad\cdot(\mathbf{C}_{[k]}^{(j)})^{-1}(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{[k]}^{(j')}(\mathbf{C}^{(j')})^{-1}\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{h}_k \\ &-\frac{P}{M^3}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)})^{-1}\left[\boldsymbol{\delta}_k^{(j)}c_{2,k}^{(j)}\boldsymbol{h}_k^{\mathrm{H}}\right]}{\sqrt{\Gamma_{j,j}^{o}\Gamma_{j',j'}^{o}}} \\ &\qquad\cdot(\mathbf{C}_{[k]}^{(j)})^{-1}(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{[k]}^{(j')}(\mathbf{C}^{(j')})^{-1}\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{h}_k \\ &-\frac{P}{M^3}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)})^{-1}\left[\boldsymbol{h}_k c_{2,k}^{(j)}(\boldsymbol{\delta}_k^{(j)})^{\mathrm{H}}\right]}{\sqrt{\Gamma_{j,j}^{o}\Gamma_{j',j'}^{o}}} \\ &\qquad\cdot(\mathbf{C}_{[k]}^{(j)})^{-1}(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{[k]}^{(j')}(\mathbf{C}^{(j')})^{-1}\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{h}_k \\ &\triangleq A+B+C+D+E.\end{aligned}\tag{27}$$

We proceed by calculating each of the 5 terms in (27) successively. Using in particular Lemma 6, we obtain

$$\begin{aligned}A &\asymp P\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{1}{\sqrt{\Gamma_{j,j}^{o}\Gamma_{j',j'}^{o}}} \\ &\cdot\Bigg[\frac{\operatorname{tr}\left(\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}_{[k]}^{(j)})^{-1}(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{[k]}^{(j')}(\mathbf{C}_{[k]}^{(j')})^{-1}\right)}{M^2} \\ &-c_{0,k}^{(j')}\frac{\operatorname{tr}\left(\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}_{[k]}^{(j)})^{-1}(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{[k]}^{(j')}(\mathbf{C}_{[k]}^{(j')})^{-1}\right)}{M^2} \\ &\cdot\frac{\operatorname{tr}\left(\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}(\mathbf{C}_{[k]}^{(j')})^{-1}\right)}{M}\frac{1}{1+\frac{\operatorname{tr}\left((\mathbf{C}_{[k]}^{(j')})^{-1}\right)}{M}}\Bigg].\end{aligned}\tag{28}$$

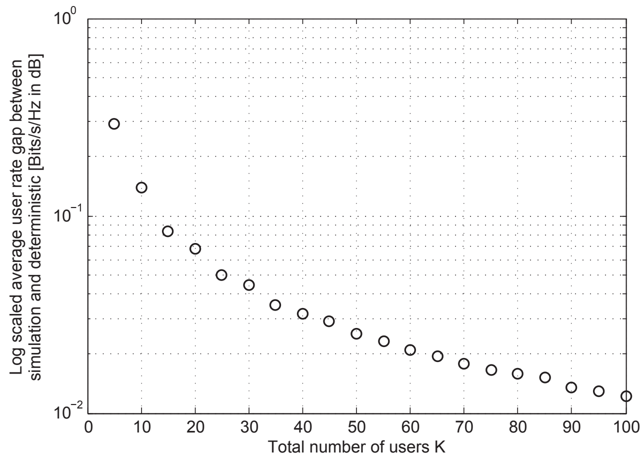

Fig. 1: Difference for the average rate per user between the numerical simulations and the deterministic equivalent as a function of the number of users K.

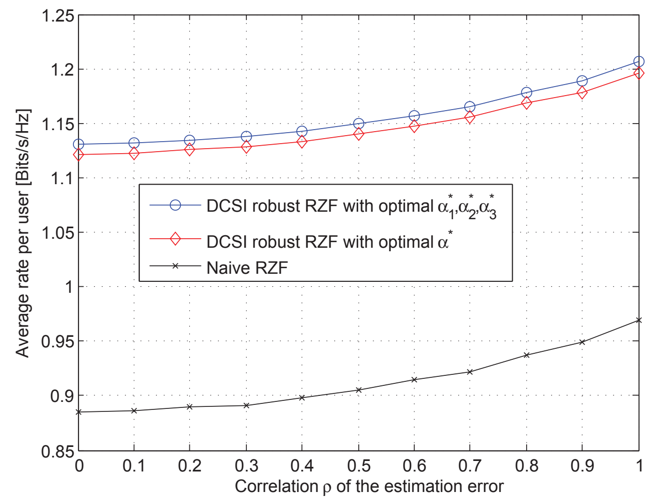

Fig. 2: Average rate per user as a function of estimation error correlation ρ.

Since $\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}} = \mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}\cdot 1_{j=j'}$, where $1_{j=j'}$ is a function that returns 1 when $j = j'$ and 0 otherwise, according to Lemma 5, it can be shown that

$$\frac{\mathrm{tr}\left(\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}_{[k]}^{(j)})^{-1}(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{[k]}^{(j')}(\mathbf{C}_{[k]}^{(j')})^{-1}\right)}{M^2} \asymp \frac{1}{n}\Gamma^o_{j,j'} \quad (29)$$

Inserting (29) in (28) and using the fundamental Theorem 1 yields

$$\begin{aligned} A &\asymp P\sum_{j=1}^{n}\sum_{j'=1}^{n}\left(\frac{1}{n}1_{j=j'} - c_{0,k}^{(j')}\frac{\Gamma^o_{j,j'}}{n\sqrt{\Gamma^o_{j,j}\Gamma^o_{j',j'}}}\frac{\delta^{(j')}}{n}\frac{1}{1+\delta^{(j')}}\right)\\ &= P - P\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{\Gamma^o_{j,j'}}{\sqrt{\Gamma^o_{j,j}\Gamma^o_{j',j'}}}\frac{c_{0,k}^{(j')}\delta^{(j')}}{n^2\left(1+\delta^{(j')}\right)}. \quad (30)\end{aligned}$$

The derivation of the B, C, D, and E terms follows exactly in the same way after applying Lemma 6 and Lemma 5. Adding all the terms together gives the interference term.

V. Simulation Results

In the following, we keep $\beta = 1$ and $P = 20$ dB while further simulation results will be provided in the journal version [20].

A. Numerical Verification of Theorem 2

We now verify using Monte-Carlo simulations the accuracy of the asymptotic expression derived in Theorem 2. We consider a network MIMO system consisting of $n = 5$ TXs and we assume that $\sigma_k^{(j)}, \forall j = 1,\ldots,n,\ k = 1,\ldots,K$ is uniformly distributed between $(0,1)$. $\rho_k^{(j,j')}, j \neq j', \forall j,j' = 1,\ldots,n,\ \forall k = 1,\ldots,K$ is also uniformly distributed between $(0,1)$ and the $n\times n$ error correlation matrix for which the (j,j') entry is $\rho_k^{(j,j')}$ is symmetric positive semi-definite. The regularization parameter $\alpha^{(j)}, \forall j = 1,\ldots,n$ is chosen uniformly distributed between $(0,1)$.

In Fig. 1, we show the average rate per user between the simulation and the deterministic equivalent as a function of the number of users for a square setting where $M = nM_{\mathrm{TX}} = K$. It confirms that Monte-Carlo simulation converges to the deterministic equivalent as the system becomes large.

B. Analysis of the CSI Noise Correlation Factor

We now discuss the cost of the distributiveness of the CSI. We consider a network consisting of $n = 3$ TXs and $K = 30$ RXs and we assume that at TX 1, $\sigma_k^{(1)} = 0.1, \forall k = 1,\ldots,K$, at TX 2, $\sigma_k^{(2)} = 0.4, \forall k = 1,\ldots,K$, and at TX 3, $\sigma_k^{(3)} = 0.7, \forall k = 1,\ldots,K$. This corresponds to an asymmetric CSI setting where TX 1 has relatively good CSI, TX 2 has moderate CSI and TX 3 has relatively bad CSI.

We choose CSI noise correlation factors $\rho_k^{(j,j')} = \rho, j \neq j', \forall j,j' = 1,\ldots,n, \forall k = 1,\ldots,K$ and we let ρ vary from 0 to 1. The parameter ρ controls the *level of centralization* such that the CSI model varies from fully distributed CSIT to centralized CSIT. We compare the performance obtained using DCSI regularized ZF with the optimal regularization coefficients $\alpha_1^*, \alpha_2^*, \alpha_3^*$ obtained numerically in Section II-E and the naive regularized ZF in Section II-D where the CSI inconsistencies between the TXs are not considered. A less complex and more efficient case where all TXs use a common regularization coefficient $\alpha^\star$ which is optimized numerically is also considered in the simulations.

It can be seen in Fig. 2 that the optimization of the regularization coefficient leads to a gain of about 30%, which is particularly interesting given that this coefficient can be optimized only on the basis of the long term statistics, i.e., at a low cost. It can also be seen that using the same coefficient at each TX leads to a even lower complexity single parameter optimization. If chosen the coefficient correctly, the performance is close to the different regularization coefficients case.

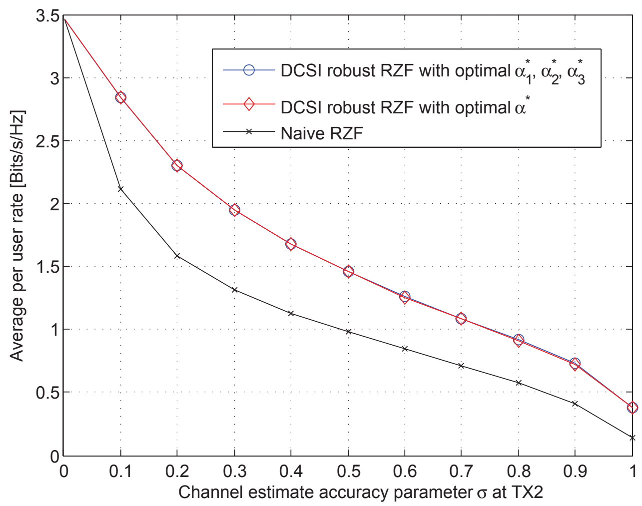

Fig. 3: Average rate per user as a function of CSI estimate accuracy parameter σ at TX2, with TX1 having perfect CSI.

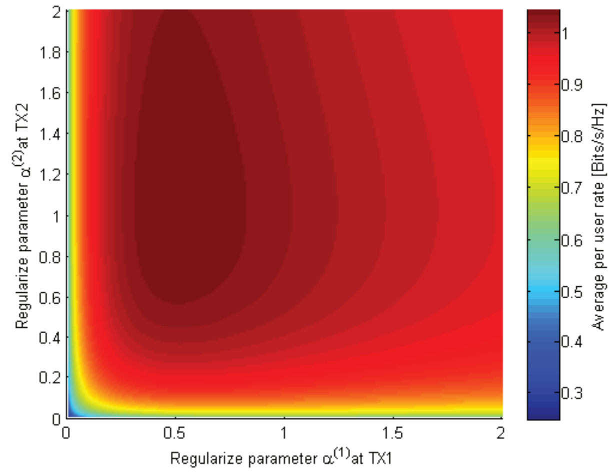

Fig. 4: Average per user rate as a function of regularization parameter at TX 1, TX 2, TX 1 has good CSI estimate and TX 2 has bad CSI estimate.

C. Impact of the Asymmetrical CSIT

In Fig. 3, we discuss the performance obtained . In this simulation, we consider $n = 2$ TXs and $K = 30$ RXs with $\rho_k^{(1,2)} = \rho_k^{(2,1)} = 0.4,\ \forall k = 1,\ldots,K$. At TX 1, $\sigma_k^{(1)} = 0, \forall k = 1,\ldots,K$, which indicates that the CSI is perfect at TX 1. At TX 2, $\sigma_k^{(2)} = \sigma, \forall k = 1,\ldots,K$ with σ varying from 0 to 1, meaning that the CSI at TX 2 varies from perfect CSI to no CSI case.

When both TX have symmetrically good CSI, the optimization of the regularization coefficient does not significantly enhance the system performance. In contrast, when the CSI quality at the two TXs becomes more and more asymmetric, the gap between the proposed D-CSI robust Regularized ZF and the naive Regularized ZF becomes non-negligible.

D. Optimal Regularization Coefficient

We now analysis how the choice of the regularization parameter will interact with the user rate. We consider a network consisting of $n = 2$ TXs and $K = 50$ RXs. $\rho_k^{(1,2)} = \rho_k^{(2,1)} = 0.4,\ \forall k = 1,\ldots,K$.

Fig. 4 exploits the case when TX 1 has rather good CSI estimate for all the user channels where $(\sigma_k^{(1)})^2 = 0.1, \forall\ k = 1,\ldots,K$, while TX 2 has rather bad CSI estimates for all user channels where $(\sigma_k^{(2)})^2 = 0.7, \forall\ k = 1,\ldots,K$. The heat map indicates the average per user rate when different regularization parameters are chosen at TX 1 and TX 2. Dark red represents higher rate and dark blue represents lower rate. We can observe that in general the optimal regularization coefficients at different TXs are non-equal, however, assuming each TX has the same regularization coefficient and optimize this coefficient based on CSI quality level for all TXs will only have a small performance degradation from the optimal different regularization coefficients case.

VI. Conclusion

We have studied regularized ZF joint precoding in a distributed CSI configuration. We extend the conventional distributed CSI scenario by allowing the CSI errors at the different TXs to be correlated. This novelty offers new perspectives for modeling the CSI in partially centralized setting. In addition, we extend the analysis of regularized ZF by allowing each TX to choose its own regularization coefficient. Using RMT tools, an analytical expression is derived to approximate the average rate per user in the large system limit. This analytical expression is then used to optimize the regularization coefficients at the different TXs in order to reduce the impact of the distributed CSI configuration.

Appendix

A. Classical Lemmas from the Literature

Lemma 1 (Resolvent Identities [17], [18]). *Given any matrix $\mathbf{H} \in \mathbb{C}^{K\times M}$, let $\boldsymbol{h}_k^{\mathrm{H}}$ denote its kth row and $\mathbf{H}_{[k]} \in \mathbb{C}^{(K-1)\times M}$ denote the matrix obtained after removing the kth row from $\mathbf{H}$. The resolvent matrices of $\mathbf{H}$ and $\mathbf{H}_{[k]}$ are denoted by $\mathbf{Q} \triangleq \left(\mathbf{H}^{\mathrm{H}}\mathbf{H} + \alpha\mathbf{I}_M\right)^{-1}$ and $\mathbf{Q}_{[k]} \triangleq \left(\mathbf{H}_{[k]}^{\mathrm{H}}\mathbf{H}_{[k]} + \alpha\mathbf{I}_M\right)^{-1}$, with $\alpha > 0$, respectively. It then holds that*

$$\mathbf{Q} = \mathbf{Q}_{[k]} - \frac{1}{M}\frac{\mathbf{Q}_{[k]}\boldsymbol{h}_k\boldsymbol{h}_k^{\mathrm{H}}\mathbf{Q}_{[k]}}{1 + \frac{1}{M}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{Q}_{[k]}\boldsymbol{h}_k} \tag{31}$$

and

$$\boldsymbol{h}_k^{\mathrm{H}}\mathbf{Q} = \frac{\boldsymbol{h}_k^{\mathrm{H}}\mathbf{Q}_{[k]}}{1 + \frac{1}{M}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{Q}_{[k]}\boldsymbol{h}_k}. \tag{32}$$

Lemma 2 ([17], [18]). *Let $(\mathbf{A}_N)_{N\geq 1}, \mathbf{A}_N \in \mathbb{C}^{N\times N}$ be a sequence of matrices such that $\limsup \|\mathbf{A}_N\| < \infty$, and $(\mathbf{x}_N)_{N\geq 1}, \mathbf{x}_N \in \mathbb{C}^{N\times 1}$ be a sequence of random vectors of* *i.i.d. entries of zero mean, unit variance, and finite eighth order moment independent of* $\mathbf{A}_N$. *Then,*

$$\frac{1}{N}\mathbf{x}_N^{\mathrm{H}}\mathbf{A}_N\mathbf{x}_N - \frac{1}{N}\operatorname{tr}(\mathbf{A}_N) \xrightarrow[N\to\infty]{a.s.} 0. \tag{33}$$

Lemma 3 ([17], [18]). *Let* $(\mathbf{A}_N)_{N\geq 1}, \mathbf{A}_N \in \mathbb{C}^{N\times N}$ *be a sequence of matrices such that* $\limsup \|\mathbf{A}_N\| < \infty$, *and* $\mathbf{x}_N, \mathbf{y}_N$ *be random, mutually independent with i.i.d. entries of zero mean, unit variance, finite eighth order moment, and independent of* $\mathbf{A}_N$. *Then,*

$$\frac{1}{N}\mathbf{x}_N^{\mathrm{H}}\mathbf{A}_N\mathbf{y}_N \xrightarrow[N\to\infty]{a.s.} 0. \tag{34}$$

Lemma 4 ([12], [18]). *Let* $\mathbf{Q}$ *and* $\mathbf{Q}_{[k]}$ *be as given in Lemma 1. Then, for any matrix* $\mathbf{A}$, *we have*

$$\operatorname{tr}\left(\mathbf{A}\left(\mathbf{Q}-\mathbf{Q}_{[k]}\right)\right) \leq \|\mathbf{A}\|_2. \tag{35}$$

B. New Lemmas

Lemma 5. *Let* $\boldsymbol{h}_k' = \sqrt{1-(\sigma_k')^2}\boldsymbol{h}_k + \sigma_k'\boldsymbol{\delta}_k'$ *and* $\boldsymbol{h}_k'' = \sqrt{1-(\sigma_k'')^2}\boldsymbol{h}_k + \sigma_k''\boldsymbol{\delta}_k''$. $\sigma_k', \sigma_k'' \in [0,1]$ *with* $\boldsymbol{h}_k$ *independent with* $\boldsymbol{\delta}_k', \boldsymbol{\delta}_k''$. $\boldsymbol{\delta}_k', \boldsymbol{\delta}_k''$ *have their elements i.i.d.* $\mathcal{N}_{\mathbb{C}}(0,1)$, $\mathbb{E}\left[\boldsymbol{\delta}_k'(\boldsymbol{\delta}_k'')^H\right] = \rho_k^2\mathbf{I}_M$. $\boldsymbol{h}_k^H, \boldsymbol{h}_k'^H, \boldsymbol{h}_k''^H, \forall k = 1\ldots K$ *are the row vectors for* $\mathbf{H}, \mathbf{H}', \mathbf{H}''$ *respectively. Let* $\mathbf{Q}' = \left(\frac{\mathbf{H}'^{\mathrm{H}}\mathbf{H}'}{M} + \alpha'\mathbf{I}_M\right)^{-1}$ *and* $\mathbf{Q}'' = \left(\frac{\mathbf{H}''^{\mathrm{H}}\mathbf{H}''}{M} + \alpha''\mathbf{I}_M\right)^{-1}$ *with* $\alpha', \alpha'' > 0$. *Let* $\mathbf{A} \in \mathbb{C}^{M\times M}$ *be of uniformly bounded spectral norm with respect to* M. *Then,*

$$\frac{1}{M^2}\operatorname{tr}\left(\mathbf{A}\mathbf{Q}'\mathbf{H}'^{\mathrm{H}}\mathbf{H}''\mathbf{Q}''\right) - \frac{\operatorname{tr}(\mathbf{A})}{M}Y_0 \xrightarrow{a.s.} 0 \tag{36}$$

where Y_0 *is defined as*

$$Y_o \triangleq \frac{\frac{\delta'}{1+\delta'}\frac{\delta''}{1+\delta''}\frac{1}{M}\sum_{\ell=1}^{K}\sqrt{(1-(\sigma_\ell')^2)(1-(\sigma_\ell'')^2)}+\sigma_\ell'\sigma_\ell''\rho_\ell^2}{1-\frac{\delta'}{1+\delta'}\frac{\delta''}{1+\delta''}\frac{1}{M}\sum_{\ell=1}^{K}\left(\sqrt{(1-(\sigma_\ell')^2)(1-(\sigma_\ell'')^2)}+\sigma_\ell'\sigma_\ell''\rho_\ell^2\right)^2} \tag{37}$$

Lemma 6. *Let* $\mathbf{L}, \mathbf{R}, \bar{\mathbf{A}} \in \mathbb{C}^{M\times M}$ *be of uniformly bounded spectral norm with respect to* M *and let* $\bar{\mathbf{A}}$ *be invertible. Let* $\mathbf{x}, \mathbf{x}', \mathbf{y}$ *have i.i.d. complex entries of zero mean, variance* $1/M$ *and finite* 8*th order moment.* $\mathbf{x}, \mathbf{y}$ *and* $\mathbf{x}', \mathbf{y}$ *are mutually independent as well as independent of* $\mathbf{L}, \mathbf{R}, \bar{\mathbf{A}}$. $\mathbf{x}, \mathbf{x}'$ *satisfies* $\mathbb{E}\left[\mathbf{x}'\mathbf{x}^H\right] = \frac{\rho^2}{M}\mathbf{I}_M$. *Then we have:*

$$\begin{aligned}
\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}^{-1}\mathbf{R}\mathbf{x} &\asymp u_{\mathrm{LR}} - \frac{c_0 u_{\mathrm{L}} u_{\mathrm{R}}}{1+u}\\
\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}^{-1}\mathbf{R}\mathbf{y} &\asymp -\frac{c_2 u_{\mathrm{L}} u_{\mathrm{R}}}{1+u}\\
\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}'^{-1}\mathbf{R}\mathbf{y} &\asymp -\rho^2\frac{c_2 u_{\mathrm{L}} u_{\mathrm{R}}}{1+u}
\end{aligned}$$

with

$$\begin{aligned}
\mathbf{A} &= \bar{\mathbf{A}} + c_0\mathbf{x}\mathbf{x}^{\mathrm{H}} + c_1\mathbf{y}\mathbf{y}^{\mathrm{H}} + c_2\mathbf{x}\mathbf{y}^{\mathrm{H}} + c_2\mathbf{y}\mathbf{x}^{\mathrm{H}}\\
\mathbf{A}' &= \bar{\mathbf{A}} + c_0\mathbf{x}'\mathbf{x}'^{\mathrm{H}} + c_1\mathbf{y}\mathbf{y}^{\mathrm{H}} + c_2\mathbf{x}'\mathbf{y}^{\mathrm{H}} + c_2\mathbf{y}\mathbf{x}'^{\mathrm{H}}
\end{aligned}$$

with $c_0 + c_1 = 1$ *and* $c_0c_1 - c_2^2 = 0$, *and*

$$\begin{aligned}
u &= \frac{\operatorname{tr}(\bar{\mathbf{A}}^{-1})}{M}, & u_{\mathrm{L}} &= \frac{\operatorname{tr}(\mathbf{L}\bar{\mathbf{A}}^{-1})}{M},\\
u_{\mathrm{R}} &= \frac{\operatorname{tr}(\bar{\mathbf{A}}^{-1}\mathbf{R})}{M}, & u_{\mathrm{LR}} &= \frac{\operatorname{tr}(\mathbf{L}\bar{\mathbf{A}}^{-1}\mathbf{R})}{M}.
\end{aligned}$$

References

[1] D. Gesbert, S. Hanly, H. Huang, S. Shamai (Shitz), O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: a new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[2] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Communications*, vol. 13, no. 4, pp. 56–61, Aug. 2006.

[3] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[4] 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.

[5] P. de Kerret, "Transmitter cooperation with distributed feedback in wireless networks," Ph.D. dissertation, TELECOM ParisTech, 2013. [Online]. Available: http://tel.archives-ouvertes.fr/tel-00952820

[6] 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.

[7] *Final definition and evaluation of PHY layer approaches for RANaaS and joint backhaul-access layer*. Available: http://www.ict-ijoin.eu/deliverables/, [Accessed: 21/09/2015], May. 2011.

[8] P. de Kerret, D. Gesbert, and U. Salim, "Large system analysis of joint regularized Zero Forcing precoding with distributed CSIT," 2015, to be published in ISIT 2015.

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

[10] M. Kobayashi, N. Jindal, and G. Caire, "Training and feedback optimization for multiuser MIMO downlink," *IEEE Trans. Commun.*, vol. 59, no. 8, pp. 2228–2240, Aug. 2011.

[11] Y. Cheng, V. K. N. Lau, and Y. Long, "A scalable limited feedback design for network MIMO using per-cell product codebook," *IEEE Trans. Wireless Commun.*, vol. 9, no. 10, pp. 3093–3099, Oct. 2010.

[12] S. Wagner, R. Couillet, M. Debbah, and D. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, July 2012.

[13] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO Channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[14] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Trans. on Commun.*, vol. 53, no. 1, pp. 195–202, 2005.

[15] B. Hochwald and S. Vishwanath, "Space-time multiple access: Linear growth in the sum rate," in *Proc. Allerton Conference on Communication, Control, and Computing (Allerton)*, 2002.

[16] A. Tulino and S. Verdu, *Random matrix theory and wireless communications*. Now Publisher Inc., 2004.

[17] A. Müller, A. Kammoun, E. Björnson, and M. Debbah, "Linear precoding based on polynomial expansion: reducing complexity in massive MIMO," 2013. [Online]. Available: http://arxiv.org/abs/1310.1806

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

[19] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for mutual information analysis of larger dimensional multi-antenna channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3987–4004, Sep. 2008.

[20] Q. Li, P. de Kerret, D. Gesbert, and N. Gresset, "Robust Regularized ZF in Cooperative Broadcast Channel under Distributed CSIT," 2015, submitted to IEEE Trans. Inf. Theo.