

Robust Regularized ZF in Cooperative Broadcast Channel under Distributed CSIT

Qianrui Li, *Member, IEEE*, Paul de Kerret, David Gesbert, *Fellow, IEEE* and Nicolas Gresset, *Senior member, IEEE*

Abstract—In this work, we consider the sum rate performance of joint processing coordinated multi-point transmission network (JP-CoMP, a.k.a Network MIMO) in a so-called distributed channel state information (D-CSI) setting. In the D-CSI setting, the various transmitters (TXs) acquire a local, TX-dependent, estimate of the global multi-user channel state matrix obtained via terminal feedback and limited backhauling. The CSI noise across TXs can be independent or correlated, so as to reflect the degree to which TXs can exchange information over the backhaul, hence allowing to model a range of situations bridging fully distributed and fully centralized CSI settings. In this context we aim to study the price of CSI distributiveness in terms of sum rate at finite SNR when compared with conventional centralized scenarios. We consider the family of JP-CoMP precoders known as regularized zero-forcing (RZF). We conduct our study in the large scale antenna regime, as it is currently envisioned to be used in real 5G deployments. It is then possible to obtain accurate approximations on so-called deterministic equivalents of the signal to interference and noise ratios. Guided by the obtained deterministic equivalents, we propose an approach to derive a RZF scheme that is robust to the distributed aspect of the CSI, whereby the key idea lies in the optimization of a TX-dependent power level and regularization factor. Our analysis confirms the improved robustness of the proposed scheme with respect to CSI inconsistency at different TXs, even with moderate number of antennas and receivers (RXs).

Index Terms—Multiuser channels, cooperative communication, coordinated multi-point transmission network, random matrix theory, limited feedback, limited backhaul, linear precoding

I. INTRODUCTION

Joint processing CoMP, whereby multiple cooperating TXs share the data streams and perform joint precoding [1], is considered for use in current and next generation wireless networks. Theoretically, with perfect data and CSI sharing, TXs at different locations can be seen as a unique virtual multiple-antenna array serving all RXs in a multiple-antenna broadcast channel (BC) fashion and well known precoding algorithms from the literature can be used [2]. However, in real systems both the feedback through the wireless medium

and the information exchange through the backhaul place a burden on overall resources and therefore must be limited.

Joint processing CoMP under limited feedback and imperfect backhaul (or fronthaul for cloud radio access network, a.k.a C-RAN systems) has been investigated in many works. In [3], [4], the capacity limited backhaul is considered and an information theoretic analysis of the system performance for joint processing CoMP is provided. In [5]–[7], the compress-and-forward schemes, cooperative beamforming and resource allocation for a C-RAN with capacity-limited fronthaul links are considered. In [8]–[10], the effect of imperfect CSIT due to limited feedback and/or delay is investigated in a single TX multiple antennas broadcast channel setting. In [11], [12], precoder designs for the joint processing CoMP with limited backhaul are provided. However, most of these contributions typically assume a *centralized* CSIT setting, i.e., the precoding is done on the basis of a *single* imperfect channel estimate which is commonly known at every TX.

This assumption of a centralized computing unit is relevant in the so-called C-RAN architecture, yet it is more and more challenged in other forms of networks where a pre-existing optical fiber backhaul is lacking or is considered too expensive in terms of CAPEX. Other emerging deployment scenarios are those with a fully heterogeneous infrastructure where the network's edge is composed of not just fixed macro base stations but also small cell base stations, mobile (possibly flying [13]) access points or relays. In such settings, exchanging CSI over limited and unreliable backhaul is likely to lead to additional quantization noise and latencies. As a result, the global downlink CSI estimate collected by any TX is unique to that TX, although the CSI noise can exhibit some degree of correlation from TX to TX. In the rest of this paper, we refer to this setting as a Distributed CSI setting, which considers implicitly the possible correlation between the estimates. In this context we are interested in the design of a distributed precoder whereby each TX computes the elements of the precoder used for its transmission based solely on its own channel estimate.

From an information theoretic perspective, the study of joint processing CoMP in D-CSI setting raises several intriguing and challenging questions.

First, while the JP-CoMP with perfect user message sharing is akin to the information theoretic MISO broadcast channel, the capacity region of the broadcast channel under a general D-CSI setting is unknown. In [14], a rate characterization at high SNR is carried out using degree of freedom (DoF) analysis for the two TXs scenario. This study highlights the severe

Manuscript received February 6, 2017; revised October 29, 2018; accepted October 28, 2019. Part of this work has been presented at the 2015 IEEE International Symposium on Information Theory (ISIT 2015) and 53rd Annual Allerton Conference on Communication, Control and Computing (Allerton 2015).

Q. Li and N. Gresset are with Mitsubishi Electric R&D Centre Europe, 35708 Rennes, France. (e-mail: q.li, n.gresset@fr.mercedes.mee.com) Q. Li was also with Communication System Department, EURECOM during this work.

D. Gesbert is with Communication System Department, EURECOM, 06410 Biot, France. (e-mail: gesbert@eurecom.fr) P. de Kerret was also with Communication System Department, EURECOM during this work, he is now with Mantu AI Lab, 06410 Biot, France. (e-mail: pdekerret@mantu.com)

penalty caused by the lack of a consistent CSI shared by the cooperating TXs from a DoF point of view, when using a conventional precoder. It was also shown that classical RZF [15] do not restore the DoF. Although a new DoF-restoring decentralized precoding strategy was presented in [14] for the two TXs case, only partial results are known for the case of an arbitrary number of users [16]. Furthermore, at finite SNR, the problem of designing precoders that optimally tackle the D-CSI setting is fully open. The use of conventional linear precoders that are unaware of the D-CSI structure is expected to yield a significant loss with respect to a centralized (and imperfect) CSI setting. Hence, an important question is how to reduce the losses due to the D-CSI configuration, i.e., how to derive a D-CSI-robust precoding scheme.

In this work, we study the average rate achieved when 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). Although RMT has been applied in many works to the analysis of wireless communications (See [8], [17]–[20] among others), its role in helping to analyze cooperative systems with distributed information has received little attention before.

In this work, our main contribution are threefold:

- A novel general D-CSI channel model that allows to study distributed CoMP networks ranging from fully distributed to fully centralized is introduced.
- A deterministic equivalent of the SINR in D-CSI setting in the limit of a large number of antennas is derived.
- Building upon this deterministic equivalent, the sum rate maximization regularization coefficient for the RZF precoder and the local optimal power allocation for each TX under a total power constraint can be found. This leads to a robust distributed RZF precoder design for the D-CSI setting. The regularization coefficient can either be optimized individually by each TX or be found by a low complexity heuristic algorithm assuming that a single common regularization coefficient is used at all TX. Simulations show that the low complexity approach approximates well the performance of the per-TX individually optimization.

Notations: In the following, boldface lower-case and upper-case characters denote vectors and matrices, respectively. The operators $(\cdot)^T, (\cdot)^H, \text{tr}(\cdot), \mathbb{E}(\cdot)$ denote transpose, conjugate transpose, trace and expectation, respectively. The $N \times N$ identity matrix is denoted \mathbf{I}_N . The notation $[\mathbf{A}]_{i,j}, [\mathbf{b}]_i$ denotes the (i,j) th entry of matrix \mathbf{A} and the i th entry of vector \mathbf{b} , respectively. $\text{diag}(\cdot)$ creates a diagonal matrix with given entries in the diagonal.

The notation $x \asymp y$ denotes that $x - y \xrightarrow[K, M_{\text{TX}} \rightarrow \infty]{a.s.} 0$. The notation $1_{a=b}$ returns 1 when $a = b$ and 0 otherwise. The notation \mathbf{i} denotes the imaginary unit. A random vector $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Theta})$ is circular symmetric complex Gaussian distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Theta}$. The notation \triangleq is used in a definition of a scalar, vector or matrix.

II. SYSTEM MODEL

A. Transmission Model

We consider a communication system where n TXs jointly serve K RXs over a joint processing CoMP transmission network. Each TX is equipped with M_{TX} antennas, while the total number of transmit antennas is denoted by $M = nM_{\text{TX}}$. Every RX is equipped with a single antenna. We assume that n is a fixed value, which corresponds to a finite TX node CoMP transmission. We also assume that the ratio of transmit antennas with respect to the number of users is fixed and given by

$$\beta \triangleq \frac{M}{K} \geq 1. \quad (1)$$

The signal y_k received at RX k reads as

$$y_k = \mathbf{h}_k^H \mathbf{x} + n_k \quad (2)$$

and the overall receiving signal at all RXs is described as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (3)$$

where $\mathbf{y} \triangleq [y_1 \dots y_K]^T \in \mathbb{C}^{K \times 1}$, $\mathbf{H} \triangleq [\mathbf{h}_1 \dots \mathbf{h}_K]^H \in \mathbb{C}^{K \times M}$ is the CoMP channel. $\mathbf{h}_k^H \in \mathbb{C}^{1 \times M}$ is the channel from all transmit antennas to RX k . $\mathbf{x} \in \mathbb{C}^{M \times 1}$ is the transmitted signal and $\mathbf{n} \triangleq [n_1 \dots n_K]^T \in \mathbb{C}^{K \times 1}$ is the noise at the K RXs. The transmission noise has i.i.d entry $n_k \sim \mathcal{N}_{\mathbb{C}}(0, 1), \forall k = 1, \dots, K$.

The multi-user transmit signal $\mathbf{x} \in \mathbb{C}^{M \times 1}$ is obtained from the symbol vector $\mathbf{s} \triangleq [s_1, \dots, s_K]^T \in \mathbb{C}^{K \times 1}$:

$$\mathbf{x} = \mathbf{T}\mathbf{s} = \sum_{k=1}^K \mathbf{t}_k s_k \quad (4)$$

with $\mathbf{T} \triangleq [\mathbf{t}_1, \dots, \mathbf{t}_K] \in \mathbb{C}^{M \times K}$ being the *multi-user* precoder, $\mathbf{t}_k \in \mathbb{C}^{M \times 1}$ being the beamforming vector for RX k . We consider an average sum power constraint

$$\text{tr}(\mathbf{T}\mathbf{T}^H) = P, \quad (5)$$

where P is the average total transmit power for all TXs. Here we adopt a sum power constraint, a per TX power constraint can also be used and the derivation will be similar to the sum power constraint case.

In addition, the channel to RX k is modeled as:

$$\mathbf{h}_k = \sqrt{M} \boldsymbol{\Theta}_k^{\frac{1}{2}} \mathbf{z}_k \quad (6)$$

where $\boldsymbol{\Theta}_k \in \mathbb{C}^{M \times M}$ is the channel correlation matrix of RX k and \mathbf{z}_k has i.i.d complex entries of zero mean, variance $\frac{1}{M}$ and eighth order moment of order $O(\frac{1}{M^4})$. The channel correlation matrices $\boldsymbol{\Theta}_k, \forall k = 1, \dots, K$ are assumed to be slowly varying compared to the channel coherence time and therefore to be perfectly known by *all* TXs.

With the assumption of Gaussian signaling $s_k \sim \mathcal{N}_{\mathbb{C}}(0, 1), \forall k$ and each user decoding with perfect CSIR, the signal-to-interference-plus-noise ratio (SINR) at RX k is given by [21]

$$\text{SINR}_k = \frac{|\mathbf{h}_k^H \mathbf{t}_k|^2}{1 + \sum_{\ell=1, \ell \neq k}^K |\mathbf{h}_k^H \mathbf{t}_\ell|^2}. \quad (7)$$

The ergodic sum rate for the CoMP network is then equal to

$$R_{sum} \triangleq \sum_{k=1}^K \mathbb{E} [\log_2 (1 + \text{SINR}_k)] \quad (8)$$

where the expectation is taken over the random channel realizations.

B. D-CSIT Model

Note that while we assume all TXs are endowed with a perfect copy of the user message packet to be sent on the downlink to the user terminal (e.g. user contents have been pre-routed or pre-cached at the TXs), we instead focus on the limitation of instantaneous CSI acquisition. This assumption is relevant to all the scenarios where the time scale for payload distribution to the transmitters is less than the time scale for CSI distribution. An example is joint precoding from multiple cache-enabled BS. In this case, the payload is cached at the BS at off peak hours, while the CSI still required to be collected on an instantaneous basis. In a setting with any sort of user mobility (hence finite CSI coherence time), CDI feedback and distribution is bound to generate the non-idealities on the CSIT. The transmitter-dependent nature of such CSI uncertainties will arise from the fact each TX may combine higher quality locally collected CSI with missing non-local CSI obtained from backhaul exchange.

For most of previous works on CoMP transmission with imperfect CSI, a centralized precoding system is considered [8], [11], [12]. There is a general assumption that there exists a central unit which can gather instantaneous CSI for all CoMP channel. Upon receiving this imperfect CoMP CSI, the central unit can make a robust precoder design based on the single imperfect CSI and redistribute the precoder choices to all TX nodes in the CoMP system. However, in this work, we consider a distributed precoding system, where the decision is kept at the local level. We assumed that each TX node preserves its own (presumably high quality) CSI and completes the full CoMP CSI with the missing (lower quality) CSI that is collected from backhaul or fronthaul sharing. This is an interesting case for high mobility systems, for example a heterogeneous infrastructure where the network's edge is composed of not just fixed macro base stations but also small cell base stations, mobile (possibly flying [13]) access points or relays. In those scenarios, centralizing the signal processing is hard to achieve. It also leads to a very bad scalability and the equation of the size of the cooperation. In contrast, considering decentralized or partially centralized signal processing allows to flexibly adapt to any scenario.

In the D-CSIT model, each TX receives its own CSI estimate for the CoMP channel. This multi-user estimate received at the TXs is the result of feedback and CSI sharing protocols and is imperfect due to the limited resources available. The actual feedback and exchange mechanism based on which the TXs receive the multi-user channel estimate is left unspecified and arbitrary [22], [23]. Apart from the feedback and exchange of instantaneous CSI, we assume that each TX can obtain perfect the statistical information for the full CoMP channel and all feedback/backhaul/fronthaul links. This is due to the

fact that statistical information has slower variation compared to instantaneous CSI, therefore we assume perfect knowledge of statistical information at each TX.

After this CSI sharing step, TX j acquires $\hat{\mathbf{H}}^{(j)} \triangleq \left[\hat{\mathbf{h}}_1^{(j)} \dots \hat{\mathbf{h}}_K^{(j)} \right]^H \in \mathbb{C}^{K \times M}$ which is the multi-user channel estimate and designs its transmit coefficients *without any further exchange of information or iterations with the other TXs*.

In state of the art, there exist two models for the imperfect CSI: the statistical model and the deterministic worst-case model [24]. In statistical model, the error is viewed as a random variable which is subjected to certain distribution; whereas in the deterministic model, the imperfect CSI lies in the neighborhood of the true channel with a bounded uncertainty region. In this paper we adopt a statistical model following the conventional models in the literature [8], [9], [20], the imperfect channel estimate $\hat{\mathbf{h}}_k^{(j)}$ for RX k at TX j is then modeled as

$$\begin{aligned} \hat{\mathbf{h}}_k^{(j)} &= \sqrt{M} \Theta_k^{\frac{1}{2}} \left(\sqrt{1 - (\sigma_k^{(j)})^2} \mathbf{z}_k + \sigma_k^{(j)} \mathbf{q}_k^{(j)} \right) \\ &= \sqrt{1 - (\sigma_k^{(j)})^2} \mathbf{h}_k + \sigma_k^{(j)} \boldsymbol{\delta}_k^{(j)}. \end{aligned} \quad (9)$$

The estimation error $\boldsymbol{\delta}_k^{(j)} = \sqrt{M} \Theta_k^{\frac{1}{2}} \mathbf{q}_k^{(j)} \in \mathbb{C}^{M \times 1}$, where $\mathbf{q}_k^{(j)}$ has i.i.d complex entries of zero mean, variance $\frac{1}{M}$, eighth order moment of order $O(\frac{1}{M^4})$ and are independent of \mathbf{z}_k and n_k . The parameter $\sigma_k^{(j)} \in [0, 1]$ indicates the accuracy of the CSIT relative to the channel to RX k , as seen at TX j . For example, $\sigma_k^{(j)} = 0$ correspond to perfect CSIT, whereas $\sigma_k^{(j)} = 1$ corresponds to the channel estimate being completely uncorrelated with the true channel.

Further, we assume that the estimation errors at TX j and TX j' satisfy

$$\mathbf{q}_k^{(j)} = \rho_k^{(j,j')} \mathbf{q}_k^{(j')} + \sqrt{1 - (\rho_k^{(j,j')})^2} \mathbf{e}_k^{(j,j')}, \forall j, j', k, \quad (10)$$

where $\rho_k^{(j,j')} \in [0, 1]$ is the correlation between $\mathbf{q}_k^{(j)}$ and $\mathbf{q}_k^{(j')}$. The vector $\mathbf{e}_k^{(j,j')}$ has i.i.d complex entries of zero mean, variance $\frac{1}{M}$, eighth order moment of order $O(\frac{1}{M^4})$ and are independent of $\mathbf{q}_k^{(j')}$. Hence, the CSI estimation errors satisfy

$$\mathbb{E} \left[\boldsymbol{\delta}_k^{(j)} (\boldsymbol{\delta}_k^{(j')})^H \right] = \Theta_k^{\frac{1}{2}} \mathbb{E} \left[\mathbf{q}_k^{(j)} (\mathbf{q}_k^{(j')})^H \right] \Theta_k^{\frac{H}{2}} = \rho_k^{(j,j')} \Theta_k. \quad (11)$$

Note that $\rho_k^{(j,j)} = 1, \forall j, k$.

It should be noticed that in both limited feedback and limited backhaul/fronthaul sharing, the limited information exchange is modeled by a quantization procedure which leads to an additive noise which has the same covariance as the quantization input. This model is based a specific quantization implementation with optimal lattice quantizer. Unlike the instantaneous CSI, the statistical information has slower variation and therefore we can assume that it is perfectly obtained by each TX. Since each TX has perfect knowledge of statistical information for the full CoMP channel, a whitening procedure can be applied on the channel vector before the quantization. The merit of this whitening procedure is that

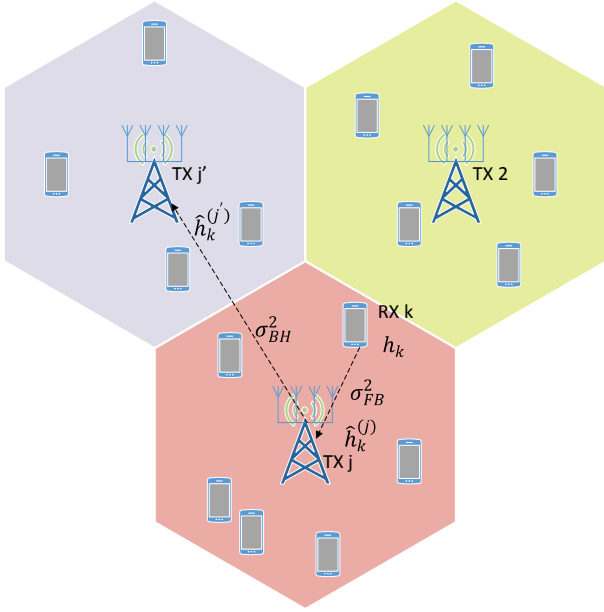


Fig. 1: CoMP transmission network with limited CSI feedback and limited CSI sharing

the whitened channel vector has equal average strength per dimension. The quantization will not be skewed by the per-dimensional power differences. Furthermore, this quantization process also allows to only feedback a reduced number of dimensions in case the channel is not full rank. After the whitening procedure, the quantization input vector is quantized by an optimal lattice quantizer, according to [25], this will introduce an additive wide-sense-stationary and white quantization noise as $\mathbf{q}_k^{(j)}$ and $\mathbf{e}_k^{(j,j')}$ in Eq. (9) and Eq. (10). At the receiving side of the limited CSI feedback or backhaul/fronthaul sharing, since the statistical information is assumed to be known perfectly, each TX node can apply a de-whitening procedure to obtain the local version of CSI estimate. This limited information exchange procedure with (i) channel whitening (ii) optimal lattice quantization (iii) de-whitening corresponds to the statistical model presented in Eq. (9) and Eq. (10).

This D-CSI model which allows for correlation between the estimate errors at different TXs is very general. It is particularly adapted to model *imperfect* CSI backhaul between TXs where delay and/or imperfections are introduced.

Example 1. Consider a particular CoMP network setting illustrated in Fig. 1. In a LTE FDD downlink channel estimation scenario, each base station (TX) sends pilots to all the served users (RXs). The RX k only feedback its downlink CSI to its associated base station, the TX j . The CSIT seen at TX j for RX k can then be modeled as

$$\hat{\mathbf{h}}_k^{(j)} = \sqrt{1 - \sigma_{\text{FB}}^2} \mathbf{h}_k + \sigma_{\text{FB}} \boldsymbol{\delta}_k^{(j)},$$

where $\sigma_{\text{FB}}^2 \in (0, 1)$ parameterizes the feedback quality and $\boldsymbol{\delta}_k^{(j)} \sim \mathcal{N}_{\mathbb{C}}(0, \mathbf{I})$ is the channel independent feedback noise.

Following the LTE-architecture, this channel estimate is then shared to the other TXs through backhaul/fronthaul

links, e.g., X2 interfaces. During this sharing step, channel estimate is further degraded due to the limited information exchange capability between these two TXs. For the TX nodes information sharing depicted in Fig. 1, the estimate received at TX j' is written as

$$\hat{\mathbf{h}}_k^{(j')} = \sqrt{1 - \sigma_{\text{BH}}^2} \hat{\mathbf{h}}_k^{(j)} + \sigma_{\text{BH}} \boldsymbol{\epsilon}_k^{(j,j')},$$

where $\sigma_{\text{BH}} \in (0, 1)$ parameterizes the backhaul quality and $\boldsymbol{\epsilon}_k^{(j,j')}$ is the sharing noise independent from $\mathbf{h}_k, \boldsymbol{\delta}_k^{(j)}$.

After basic algebraic operation, we exhibit that this scenario fits a D-CSIT configuration with the following parameters:

$$\begin{aligned} \sigma_k^{(j)} &= \sigma_{\text{FB}} \\ \sigma_k^{(j')} &= \sqrt{1 - (1 - \sigma_{\text{BH}}^2)(1 - \sigma_{\text{FB}}^2)} \\ \rho_k^{(j,j')} &= \frac{\sigma_{\text{FB}} \sqrt{1 - \sigma_{\text{BH}}^2}}{\sqrt{1 - (1 - \sigma_{\text{BH}}^2)(1 - \sigma_{\text{FB}}^2)}} \end{aligned} \quad (12)$$

□

Remark 1. The D-CSIT model bridges the gap between the two extreme configuration: centralized CSIT and fully distributed CSIT. Indeed, choosing

$$\begin{aligned} \sigma_k^{(j)} &= \sigma_k^{(j')}, \quad \rho_k^{(j,j')} = 1, \\ \forall j, j' \in \{1, \dots, n\}, \quad \forall k \in \{1, \dots, K\} \end{aligned} \quad (13)$$

corresponds to the centralized CSIT configuration [8], [9], while choosing

$$\rho_k^{(j,j')} = 0, \quad \forall j, j' \in \{1, \dots, n\}, j \neq j', \quad \forall k \in \{1, \dots, K\} \quad (14)$$

simplifies to the fully distributed CSIT configuration with uncorrelated estimation errors as previously studied in the literature [14]. □

C. Regularized Zero Forcing with Distributed CSI

We consider in this work the analysis of RZF precoder [15], [26], when faced with CSIT inconsistencies in the large system regime. Hence, the precoder designed at TX j is assumed to take the form

$$\mathbf{T}_{\text{rZF}}^{(j)} \triangleq \left((\hat{\mathbf{H}}^{(j)})^{\text{H}} \hat{\mathbf{H}}^{(j)} + M \alpha^{(j)} \mathbf{I}_M \right)^{-1} (\hat{\mathbf{H}}^{(j)})^{\text{H}} \frac{\sqrt{P}}{\sqrt{\Psi^{(j)}}}. \quad (15)$$

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

$$\Psi^{(j)} \triangleq \left\| \left((\hat{\mathbf{H}}^{(j)})^{\text{H}} \hat{\mathbf{H}}^{(j)} + M \alpha^{(j)} \mathbf{I}_M \right)^{-1} (\hat{\mathbf{H}}^{(j)})^{\text{H}} \right\|_{\text{F}}^2. \quad (16)$$

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

$$\mathbf{C}^{(j)} \triangleq \frac{(\hat{\mathbf{H}}^{(j)})^{\text{H}} \hat{\mathbf{H}}^{(j)}}{M} + \alpha^{(j)} \mathbf{I}_M. \quad (17)$$

Therefore, the precoder at TX j can be rewritten as

$$\mathbf{T}_{\text{rZF}}^{(j)} = \frac{1}{M} (\mathbf{C}^{(j)})^{-1} (\hat{\mathbf{H}}^{(j)})^{\text{H}} \frac{\sqrt{P}}{\sqrt{\Psi^{(j)}}}. \quad (18)$$

Let $\mathbf{E}_j^H \in \mathbb{C}^{M_{TX} \times M}$ denote the block selection matrix defined as

$$\mathbf{E}_j^H \triangleq [\mathbf{0}_{M_{TX} \times (j-1)M_{TX}} \quad \mathbf{I}_{M_{TX}} \quad \mathbf{0}_{M_{TX} \times (n-j)M_{TX}}]. \quad (19)$$

Upon concatenation of all TX's precoding matrices, the effective global precoder denoted by $\mathbf{T}_{rZF}^{\text{DCSI}}$, is written as

$$\mathbf{T}_{rZF}^{\text{DCSI}} \triangleq \begin{bmatrix} \mu_1 \mathbf{E}_1^H \mathbf{T}_{rZF}^{(1)} \\ \mu_2 \mathbf{E}_2^H \mathbf{T}_{rZF}^{(2)} \\ \vdots \\ \mu_n \mathbf{E}_n^H \mathbf{T}_{rZF}^{(n)} \end{bmatrix}, \quad (20)$$

where the scalar $\mu_j > 0$ is the transmit power scaling at TX j . Assume the transmit power allocated at TX j reads

$$P_{\text{TX}_j} = \mu_j^2 \text{tr} \left(\mathbf{E}_j \mathbf{E}_j^H \mathbf{T}_{rZF}^{(j)} (\mathbf{T}_{rZF}^{(j)})^H \right). \quad (21)$$

Based on the sum power constraint,

$$\sum_{j=1}^n P_{\text{TX}_j} = \sum_{j=1}^n \mu_j^2 \text{tr} \left(\mathbf{E}_j \mathbf{E}_j^H \mathbf{T}_{rZF}^{(j)} (\mathbf{T}_{rZF}^{(j)})^H \right) = P. \quad (22)$$

The finite SNR rate analysis under the precoding structure (20) and the D-CSIT model in (9) is challenging due to the dependency of each user performance on all channel estimates. Yet, some useful results can be obtained in the large antenna regime as shown below.

III. DETERMINISTIC EQUIVALENT OF THE SINR

In this section, the analysis of the so-called deterministic equivalent of the SINR under the RZF precoding is presented.

In order to derive a deterministic equivalent, we make the following standard technical assumption on the correlation matrices Θ_k and the Gram matrix $\frac{1}{M}(\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j)}$ [8].

Assumption 1. All correlation matrices $\Theta_k, \forall k = 1, \dots, K$ have uniformly bounded spectral norm on M , i.e.,

$$\limsup_{M, K \rightarrow \infty} \sup_{1 \leq k \leq K} \|\Theta_k\| < \infty. \quad (23)$$

Assumption 2. The random matrices $\frac{1}{M}(\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j)}, \forall j = 1, \dots, n$ have uniformly bounded spectral norm on M with probability one, i.e.,

$$\limsup_{M, K \rightarrow \infty} \left\| \frac{1}{M}(\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j)} \right\| < \infty \quad (24)$$

with probability one.

Our approach will be based on the following fundamental result based on the Stieltjes transform in the analysis of wireless networks [8], [19].

Theorem 1. [19], [27] Let the matrix \mathbf{U} be any matrix with bounded spectral norm and the i th row \mathbf{h}_i^H of \mathbf{H} be $\mathbf{h}_i^H = \sqrt{M} \Theta_i^{\frac{1}{2}} \mathbf{z}_i^H$, where the entries of \mathbf{z}_i are i.i.d of zero mean, variance $\frac{1}{M}$ and have eighth moment of order $O(\frac{1}{M^4})$. Let Assumption 1 holds true. Consider the resolvent matrix $\mathbf{Q} \triangleq \left(\frac{\mathbf{H}^H \mathbf{H}}{M} + \alpha \mathbf{I}_M \right)^{-1}$ with regularization coefficient $\alpha > 0$. Let

$$\mathbf{Q}_o \triangleq \left(\frac{1}{M} \sum_{k=1}^K \frac{\Theta_k}{1 + m_k} + \alpha \mathbf{I}_M \right)^{-1} \quad (25)$$

where m_k satisfies:

$$m_k = \frac{1}{M} \text{tr} \left(\Theta_k \left(\frac{1}{M} \sum_{\ell=1}^K \frac{\Theta_\ell}{1 + m_\ell} + \alpha \mathbf{I}_M \right)^{-1} \right). \quad (26)$$

Then,

$$\frac{1}{M} \text{tr}(\mathbf{U} \mathbf{Q}) - \frac{1}{M} \text{tr}(\mathbf{U} \mathbf{Q}_o) \xrightarrow[K, M \rightarrow \infty]{a.s.} 0. \quad (27)$$

The fixed point m_k can easily be obtained by an iterative fixed-point algorithm described in [8], [20] and recalled in Appendix A for the sake of completeness.

Adopting the shorthand notation used in [8], we introduce

$$c_{0,k}^{(j)} \triangleq 1 - (\sigma_k^{(j)})^2, \quad c_{1,k}^{(j)} \triangleq (\sigma_k^{(j)})^2, \quad c_{2,k}^{(j)} \triangleq \sigma_k^{(j)} \sqrt{1 - (\sigma_k^{(j)})^2}. \quad (28)$$

We can further define the term $\mathbf{Q}_o^{(j)}$ and $m_k^{(j)}$ respectively as \mathbf{Q}_o and m_k in Theorem 1 using instead the local CSI estimate $\hat{\mathbf{H}}^{(j)}$ and regularization coefficient $\alpha^{(j)}$ at TX j . A deterministic equivalent of the SINR under RZF precoding is therefore provided in the following theorem.

Theorem 2. Let the Assumptions 1 and 2 hold true, then the SINR of RX k under RZF precoding satisfies

$$\text{SINR}_k - \text{SINR}_k^o \xrightarrow[K, M_{TX} \rightarrow \infty]{a.s.} 0 \quad (29)$$

with SINR_k^o defined as

$$\text{SINR}_k^o \triangleq \frac{P \left(\sum_{j=1}^n \mu_j \sqrt{\frac{c_{0,k}^{(j)}}{\Gamma_{j,j}^o(\mathbf{I}_M)}} \frac{\Phi_{j,k}^o}{1 + m_k^{(j)}} \right)^2}{1 + I_k^o} \quad (30)$$

with $I_k^o \in \mathbb{R}$ given by

$$I_k^o \triangleq P \sum_{j=1}^n \sum_{j'=1}^n \frac{\mu_j \mu_{j'}}{\sqrt{\Gamma_{j,j}^o(\mathbf{I}_M) \Gamma_{j',j'}^o(\mathbf{I}_M)}} \left(\Gamma_{j,j'}^o(\mathbf{E}_j \mathbf{E}_j^H \Theta_k \mathbf{E}_j \mathbf{E}_j^H) - 2 \Gamma_{j,j'}^o(\Theta_k \mathbf{E}_j \mathbf{E}_j^H) \frac{c_{0,k}^{(j)} \Phi_{j',k}^o}{1 + m_k^{(j')}} + \Phi_{j',k}^o \Phi_{j,k}^o \Gamma_{j,j'}^o(\Theta_k) \frac{c_{0,k}^{(j)} c_{0,k}^{(j')} + \rho_k^{(j,j')} c_{2,k}^{(j)} c_{2,k}^{(j')}}{(1 + m_k^{(j)})(1 + m_k^{(j')})} \right). \quad (31)$$

where $\Phi_{j,k}^o \in \mathbb{R}$ is defined as

$$\Phi_{j,k}^o = \frac{\text{tr} \left(\Theta_k \mathbf{E}_j \mathbf{E}_j^H \mathbf{Q}_o^{(j)} \right)}{M}, \quad (32)$$

and the function $\Gamma_{j,j'}^o(\mathbf{X}) : \mathbb{C}^{M \times M} \mapsto \mathbb{C}$ is defined in Lemma 8. The transmit power scaling μ_j for TX j satisfies

$$\sum_{j=1}^n \mu_j^2 \frac{\Gamma_{j,j}^o(\mathbf{E}_j \mathbf{E}_j^H)}{\Gamma_{j,j}^o(\mathbf{I}_M)} = 1 \quad (33)$$

Proof. The proof of Theorem 2 is given in Appendix C. \square

The theorem demonstrates that in the large system setting, the SINR expression for each RX can be derived as a given function of (i) n, M_{TX}, K that indicate the system dimensions, (ii) $\sigma_k^{(j)}, \rho_k^{(j,j')}, \Theta_k$ which reflect the statistics of the channel

and of CSI estimates at each TX, and (iii) the precoder regularization coefficients $\alpha^{(j)}$ and power scalings $\mu^{(j)}$.

This result is very general and encompasses several important results from the literature.

A. Regularized ZF Precoding for Centralized CSI Isotropic Channel

Choosing $\sigma_k^{(j)} = \sigma_k^{(j')} = \sigma_k$, $\alpha^{(j)} = \alpha^{(j')} = \alpha$, $\rho_k^{(j,j')} = 1, \forall j, j' \in \{1, \dots, n\}, k \in \{1, \dots, K\}$, we obtain the centralized CSIT configuration. Further assuming that $\Theta_k = \mathbf{I}_M$, $m_k^{(j)}$ is obtained in closed form as

$$m_k^{(j)} = m^o = \frac{\beta - 1 - \alpha\beta + \sqrt{(\alpha\beta - \beta + 1)^2 + 4\alpha\beta^2}}{2\alpha\beta}. \quad (34)$$

In this setting, the total power constraint (33) simplifies to

$$\frac{1}{n} \sum_{j=1}^n \mu_j^2 = 1 \quad (35)$$

since

$$\Gamma_{j,j}^o(\mathbf{I}_M) = \frac{(m^o)^2}{\beta(1+m^o)^2 - (m^o)^2}, \quad (36)$$

$$\Gamma_{j,j}^o(\mathbf{E}_j \mathbf{E}_j^H) = \frac{1}{n} \Gamma_{j,j}^o(\mathbf{I}_M). \quad (37)$$

Assume $\mu_j = 1, \forall j = 1, \dots, n$, the transmit power P_{TX_j} at TX j denotes

$$P_{\text{TX}_j} = \mu_j^2 P \frac{\Gamma_{j,j}^o(\mathbf{E}_j \mathbf{E}_j^H)}{\Gamma_{j,j}^o(\mathbf{I}_M)} = \frac{P}{n} \quad (38)$$

This indicates an equal power allocation per TX. Since $\Theta_k = \mathbf{I}_M$ the channel is isotropic, the above setting also indicates that the signal power for RX k satisfies

$$P_{\text{RX}_k} = \frac{P}{K} \quad (39)$$

which is an equal power per RX.

After simple algebraic manipulations, we can obtain the deterministic equivalent of SINR in (30)

$$\text{SINR}_k^o = \frac{(1 - \sigma_k^2)(\beta(1+m^o)^2 - (m^o)^2)}{\left(1 - \sigma_k^2 + (1+m^o)^2 \sigma_k^2 + \frac{(1+m^o)^2}{P}\right)} \quad (40)$$

This coincides with the results in [8, Corollary 2].

B. Regularized ZF Precoding for Fully Distributed CSI Isotropic Channel

Choosing $\rho_k^{(j,j')} = 0, \forall j, j' \in \{1, \dots, n\}, j \neq j', k \in \{1, \dots, K\}$, the fully distributed CSIT configuration with uncorrelated estimation errors is obtained. Let us further assume that the same regularization coefficient is used at each TX, i.e., $\alpha^{(j)} = \alpha^{(j')} = \alpha, \forall j, j' \in \{1, \dots, n\}$, $\Theta_k = \mathbf{I}_M$ and $\mu_j = 1$ indicating equal per TX power allocation.

The deterministic SINR in (30) then becomes

$$\text{SINR}_k^o = \frac{P \left(\frac{1}{n} \sum_{j=1}^n \sqrt{c_{0,k}^{(j)}} \right)^2 \frac{\beta(1+m^o)^2 - (m^o)^2}{(1+m^o)^2}}{I_k^o + 1} \quad (41)$$

with

$$I_k^o = P - P \sum_{j=1}^n \sum_{j'=1}^n \frac{(\beta(1+m^o)^2 - (m^o)^2) \Gamma_{j,j'}^o}{n^2(1+m^o)^2 m^o} \cdot \left[2c_{0,k}^{(j)} + m^o \left(2c_{0,k}^{(j)} - c_{0,k}^{(j)} c_{0,k}^{(j')} \right) \right] \quad (42)$$

$$\Gamma_{j,j'}^o = \frac{\frac{1}{M} \sum_{\ell=1}^K \sqrt{c_{0,\ell}^{(j)} c_{0,\ell}^{(j')}}}{\frac{(1+m^o)^2}{(m^o)^2} - \frac{1}{M} \sum_{\ell=1}^K c_{0,\ell}^{(j)} c_{0,\ell}^{(j')}} \quad (43)$$

This result coincides with [28].

C. Regularized ZF Precoding for D-CSI Isotropic Channel

Assume that $\Theta_k = \mathbf{I}_M, \forall k \in \{1, \dots, K\}$ and $\mu_j = 1, \forall j \in \{1, \dots, n\}$, indicating equal per TX power allocation. In this specific setting, the terms $m_k^{(j)}$ can be obtained in closed form as

$$m_k^{(j)} = m^{(j)} = \frac{\beta - 1 - \alpha^{(j)}\beta + \sqrt{(\alpha^{(j)}\beta - \beta + 1)^2 + 4\alpha^{(j)}\beta^2}}{2\alpha^{(j)}\beta}. \quad (44)$$

After simplification, the deterministic SINR in (30) becomes

$$\text{SINR}_k^o = \frac{P \left(\frac{1}{n} \sum_{j=1}^n \sqrt{\frac{1 - (\sigma_k^{(j)})^2}{\Gamma_{j,j}^o} \frac{m^{(j)}}{1+m^{(j)}}} \right)^2}{1 + I_k^o} \quad (45)$$

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

$$I_k^o = P - P \sum_{j=1}^n \sum_{j'=1}^n \frac{\Gamma_{j,j'}^o}{\sqrt{\Gamma_{j,j}^o \Gamma_{j',j'}^o}} \left[\frac{2c_{0,k}^{(j)} m^{(j)}}{n^2 (1+m^{(j)})} - \frac{\left(\rho_k^{(j,j')} c_{2,k}^{(j)} c_{2,k}^{(j')} + c_{0,k}^{(j)} c_{0,k}^{(j')} \right) m^{(j)} m^{(j')}}{n^2 (1+m^{(j)}) (1+m^{(j')})} \right] \quad (46)$$

where $\Gamma_{j,j'}^o \in \mathbb{R}$ is given by

$$\Gamma_{j,j'}^o = \frac{\frac{1}{M} \sum_{\ell=1}^K \sqrt{c_{0,\ell}^{(j)} c_{0,\ell}^{(j')}} + \sqrt{c_{1,\ell}^{(j)} c_{1,\ell}^{(j')}} \rho_{\ell}^{(j,j')}}{\frac{1+m^{(j)}}{m^{(j)}} \frac{1+m^{(j')}}{m^{(j')}} - \frac{\sum_{\ell=1}^K \left(\sqrt{c_{0,\ell}^{(j)} c_{0,\ell}^{(j')}} + \sqrt{c_{1,\ell}^{(j)} c_{1,\ell}^{(j')}} \rho_{\ell}^{(j,j')} \right)^2}{M}} \quad (47)$$

This result coincides with [29].

IV. APPLICATIONS OF THE THEOREM

The deterministic equivalent of the SINR expression allows to evaluate the performance of RZF precoding. However, there is an added benefit here, which is the possibility to optimize the transmission parameters (i.e., regularization coefficient) so as to obtain some robustness with respect to the D-CSIT configuration. The D-CSI configuration reflects the fact that there are uncertainties in the knowledge of the channel coefficient. With the statistics information known at all TX nodes, it further indicates that the amount of uncertainties varies from one TX to another. A robust transmission parameter design under D-CSI indicates that we can optimize transmission parameters so as to guarantee the average performance, i.e., the expected value of ergodic sum rate over the CSI uncertainties under D-CSI configuration. These robust designs are discussed in the following.

A. Robust Sum Rate Maximizing Regularization

If there exists a predefined per TX power constraint such that the average transmit power for each TX j is given as $P_{TX_j} = p_j$, according to Theorem 2, we can find the power scaling parameter for each TX j as

$$\mu_j = \sqrt{\frac{p_j \Gamma_{j,j}^o(\mathbf{I}_M)}{P \Gamma_{j,j}^o(\mathbf{E}_j \mathbf{E}_j^H)}}. \quad (48)$$

Substituting (48) into Theorem 2, the ergodic sum rate becomes a function only depending on $\alpha^{(j)}$, $j = 1, \dots, n$.

1) *Robust Regularized ZF*: The regularization coefficients tuple $\alpha = [\alpha^{(1)}, \dots, \alpha^{(n)}]$ which maximizes the system sum rate while being robust to the D-CSIT configuration is given by

$$\begin{aligned} \alpha^* &\triangleq \operatorname{argmax}_{\alpha} \sum_{k=1}^K \log(1 + \text{SINR}_k^o), \\ \text{s.t. } \mu_j &= \sqrt{\frac{p_j \Gamma_{j,j}^o(\mathbf{I}_M)}{P \Gamma_{j,j}^o(\mathbf{E}_j \mathbf{E}_j^H)}}, \forall j. \end{aligned} \quad (49)$$

2) *Robust regularized ZF with equal regularization*: The problem (49) is a non-convex optimization. In order to reduce the complexity, we introduce the following optimization assuming that the regularization coefficients are the same at different TXs.

$$\begin{aligned} \alpha_{\text{same}}^* &\triangleq \operatorname{argmax}_{\alpha_{\text{same}}} \sum_{k=1}^K \log(1 + \text{SINR}_k^o), \\ \text{s.t. } \mu_j &= \sqrt{\frac{p_j \Gamma_{j,j}^o(\mathbf{I}_M)}{P \Gamma_{j,j}^o(\mathbf{E}_j \mathbf{E}_j^H)}}, \forall j. \end{aligned} \quad (50)$$

The optimization variable is now a scalar parameter and the global optimal regularization can be easily found using a line search algorithm [30].

3) *Naive Regularized ZF*: We introduced in the following the naive regularization optimization which doesn't take into account the D-CSIT configuration. This is therefore the reference baseline for our improved robust precoding scheme.

When TXs are not aware of the D-CSIT structure, each TX will choose its regularization parameter on the basis of its own CSI quality. This yields a naive (suboptimal) precoding scheme. Specifically, assuming equal power allocation at each TX, each TX j optimizes its regularization coefficient $\alpha^{(j)}$ based on $\hat{\mathbf{H}}^{(j)}$ considering as if $\hat{\mathbf{H}}^{(j)}$ is the centralized CSIT shared among all TXs, i.e.,

$$\alpha_{\text{naive}}^{(j)} \triangleq \operatorname{argmax}_{\alpha^{(j)}} R_{\text{sum}}(\hat{\mathbf{H}}^{(1)}, \dots, \hat{\mathbf{H}}^{(j)}). \quad (51)$$

In the particular case where the CSIT quality is homogeneous across users, i.e., $\sigma_k^{(j)} = \sigma^{(j)}$, $\forall k \in \{1, \dots, K\}$, and the channel is isotropic, i.e., $\Theta_k = \mathbf{I}_M$, the optimal naive regularization coefficient is obtained in closed form [8]

$$\alpha_{\text{naive}}^{(j)} = \frac{1 + (\sigma^{(j)})^2 P}{1 - (\sigma^{(j)})^2} \frac{1}{\beta P}. \quad (52)$$

B. Robust Power Optimization

If the regularization coefficient at each TX is predefined, according to Theorem 2, we can optimize the power scaling tuple $\boldsymbol{\mu} = [\mu_1, \dots, \mu_n]$ that maximizes the system sum rate:

$$\begin{aligned} \boldsymbol{\mu}^* &= \operatorname{argmax}_{\boldsymbol{\mu}} \sum_{k=1}^K \log(1 + \text{SINR}_k^o), \\ \text{s.t. } \sum_{j=1}^n \mu_j^2 \frac{\Gamma_{j,j}^o(\mathbf{E}_j \mathbf{E}_j^H)}{\Gamma_{j,j}^o(\mathbf{I}_M)} &= 1 \end{aligned} \quad (53)$$

Problem (53) can then be reformulated as:

$$\begin{aligned} \boldsymbol{\mu}^* &= \operatorname{argmin}_{\boldsymbol{\mu}} \prod_{k=1}^K \frac{\frac{1}{P} + \boldsymbol{\mu}^T \mathbf{B}_k \boldsymbol{\mu}}{\frac{1}{P} + \boldsymbol{\mu}^T (\mathbf{A}_k + \mathbf{B}_k) \boldsymbol{\mu}}, \\ \text{s.t. } \|\mathbf{C} \boldsymbol{\mu}\|^2 &= 1, \boldsymbol{\mu} \in \mathbb{R}^n, \end{aligned} \quad (\text{P1})$$

where $\mathbf{A}_k, \mathbf{B}_k, \mathbf{C}, \forall k$ are constant matrices defined as

$$[\mathbf{A}_k]_{j,j'} \triangleq \sqrt{\frac{c_{0,k}^{(j)} c_{0,k}^{(j')}}{\Gamma_{j,j}^o(\mathbf{I}_M) \Gamma_{j',j'}^o(\mathbf{I}_M)}} \frac{\Phi_{j,k}^o \Phi_{j',k}^o}{(1 + m_k^{(j)}) (1 + m_k^{(j')})} \quad (54)$$

$$\begin{aligned} [\mathbf{B}_k]_{j,j'} &\triangleq \frac{1}{\sqrt{\Gamma_{j,j}^o(\mathbf{I}_M) \Gamma_{j',j'}^o(\mathbf{I}_M)}} \left(\Gamma_{j,j'}^o(\mathbf{E}_j \mathbf{E}_j^H \Theta_k \mathbf{E}_j \mathbf{E}_j^H) \right. \\ &\quad \left. - 2 \Gamma_{j,j'}^o(\Theta_k \mathbf{E}_j \mathbf{E}_j^H) \frac{c_{0,k}^{(j')} \Phi_{j',k}^o}{1 + m_k^{(j')}} \right. \\ &\quad \left. + \Phi_{j',k}^o \Phi_{j,k}^o \Gamma_{j,j'}^o(\Theta_k) \frac{c_{0,k}^{(j)} c_{0,k}^{(j')} + \rho_k^{(j,j')} c_{2,k}^{(j)} c_{2,k}^{(j')}}{(1 + m_k^{(j)}) (1 + m_k^{(j')})} \right) \end{aligned} \quad (55)$$

$$\mathbf{C} \triangleq \operatorname{diag} \left(\sqrt{\frac{\Gamma_{1,1}^o(\mathbf{E}_1 \mathbf{E}_1^H)}{\Gamma_{1,1}^o(\mathbf{I}_M)}}, \dots, \sqrt{\frac{\Gamma_{n,n}^o(\mathbf{E}_n \mathbf{E}_n^H)}{\Gamma_{n,n}^o(\mathbf{I}_M)}} \right). \quad (56)$$

Let $u_i(\boldsymbol{\mu})$ be denoted as

$$u_i(\boldsymbol{\mu}) \triangleq \frac{\frac{1}{P} + \boldsymbol{\mu}^T \mathbf{B}_i \boldsymbol{\mu}}{\frac{1}{P} + \boldsymbol{\mu}^T (\mathbf{A}_i + \mathbf{B}_i) \boldsymbol{\mu}}. \quad (57)$$

In order to solve problem (P1), we first introduce the following lemma:

Lemma 1 (Lemma 1, [31]). *The optimal point of optimization problem*

$$\begin{aligned} \min_{\boldsymbol{\mu}} \quad & \prod_{i=1}^K u_i(\boldsymbol{\mu}) \\ \text{s.t.} \quad & \|\mathbf{C} \boldsymbol{\mu}\|^2 = 1, \boldsymbol{\mu} \in \mathbb{R}^n \end{aligned}$$

can be obtained by solving the following parametric problem

$$\begin{aligned} \min_{\boldsymbol{\mu}, \{\lambda_i\}_{i=1}^K} \quad & \left(\sum_{i=1}^K \frac{1}{K} \lambda_i u_i(\boldsymbol{\mu}) \right)^K \\ \text{s.t.} \quad & \prod_{i=1}^K \lambda_i = 1, \lambda_i \geq 0 \\ & \|\mathbf{C} \boldsymbol{\mu}\|^2 = 1, \boldsymbol{\mu} \in \mathbb{R}^n \end{aligned}$$

Moreover, for fixed $\boldsymbol{\mu}$, the optimal $\{\lambda_i\}_{i=1}^K$ of this problem is given by

$$\lambda_i^* = \frac{\left[\prod_{\ell=1}^K u_\ell(\boldsymbol{\mu}) \right]^{\frac{1}{K}}}{u_i(\boldsymbol{\mu})}, \forall i. \quad (58)$$

Remark 2. The above lemma is exactly Lemma 1 presented in [31] with a change of notation. \square

According to Lemma 1, noticing that optimization for a positive expression raising to the power of K achieves the same optimal point as the optimization for the expression without the scalar power raise, problem (P1) can be solved by the following parametric problem

$$\begin{aligned} \min_{\boldsymbol{\mu}, \{\lambda_i\}_{i=1}^K} & \sum_{k=1}^K \lambda_k \frac{\frac{1}{P} + \boldsymbol{\mu}^T \mathbf{B}_k \boldsymbol{\mu}}{\frac{1}{P} + \boldsymbol{\mu}^T (\mathbf{A}_k + \mathbf{B}_k) \boldsymbol{\mu}} \\ \text{s.t.} & \prod_{i=1}^K \lambda_i = 1 \\ & \|\mathbf{C}\boldsymbol{\mu}\|^2 = 1, \boldsymbol{\mu} \in \mathbb{R}^{n \times 1}. \end{aligned} \quad (\text{P2})$$

We hereby introduce an iterative procedure to calculate the local optimal solution for problem (P2).

Algorithm 1 Iterative algorithm for problem (P2)

- 1: Initialize $\boldsymbol{\mu}^{[0]}$
 - 2: $t = 0$
 - 3: **while** not converge **do**
 - 4: $\lambda_i^{[t+1]} = \left[\frac{\prod_{\ell=1}^K \frac{\frac{1}{P} + (\boldsymbol{\mu}^{[t]})^T \mathbf{B}_\ell \boldsymbol{\mu}^{[t]}}{\frac{1}{P} + (\boldsymbol{\mu}^{[t]})^T (\mathbf{A}_\ell + \mathbf{B}_\ell) \boldsymbol{\mu}^{[t]}}}{\frac{\frac{1}{P} + (\boldsymbol{\mu}^{[t]})^T (\mathbf{A}_i + \mathbf{B}_i) \boldsymbol{\mu}^{[t]}}{\frac{1}{P} + (\boldsymbol{\mu}^{[t]})^T \mathbf{B}_i \boldsymbol{\mu}^{[t]}}}, \forall i = 1, \dots, K \right]^{\frac{1}{K}}$
 - 5: $\boldsymbol{\mu}^{[t+1]} = \underset{\mathbf{x}}{\operatorname{argmin}} \sum_{k=1}^K \lambda_k^{[t+1]} \cdot \frac{\frac{1}{P} + \mathbf{x}^T \mathbf{B}_k \mathbf{x}}{\frac{1}{P} + \mathbf{x}^T (\mathbf{A}_k + \mathbf{B}_k) \mathbf{x}}$
 - 6: $\text{s.t.} \quad \|\mathbf{C}\mathbf{x}\|^2 \leq 1$
 - 7: $t = t + 1$
 - 8: **end while**
-

The iterative optimization step in Algorithm 1 is equivalent to a maximization for the sum of ratios of two convex functions over a convex set. It can be solved for example by a branch and bound algorithm described in [32].

Theorem 3. *Algorithm 1 converges to a local optimum of the optimization problem (P2).*

Proof. The proof of Theorem 3 is given in Appendix D. \square

Thanks to Theorem 3, the convergence for the iterative optimization in Algorithm 1 is guaranteed. Therefore, we can obtain a local optimal power allocation such that the system sum rate is maximized under the D-CSI configuration.

C. Robust Joint Optimization of Regularization and Power

In subsection IV-A and IV-B, we tackle the problem of finding the regularization coefficient (power scaling factor) which maximizes the system sum rate while the power scaling factor (regularization coefficient) is fixed, respectively. Indeed in the D-CSIT configuration, the regularization tuple $\boldsymbol{\alpha}$ and the power scaling tuple $\boldsymbol{\mu}$ can be jointly optimized according to a predefined power constraint. However, since the joint optimization for $(\boldsymbol{\alpha}, \boldsymbol{\mu})$ is a complicated non-convex problem, we then consider an alternating optimization approach which iterates between the optimization of $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}$ described in subsection IV-A and IV-B. A local optimal point can be reached while applying the alternating optimization.

In this subsection, we mainly consider two categories of joint optimization problems described in the sequel.

1) Robust Joint Optimization:

$$\begin{aligned} (\boldsymbol{\alpha}^*, \boldsymbol{\mu}^*) &= \underset{\boldsymbol{\alpha}, \boldsymbol{\mu}}{\operatorname{argmax}} \sum_{k=1}^K \log(1 + \text{SINR}_k^o), \\ \text{s.t.} & \sum_{j=1}^n \mu_j^2 \frac{\Gamma_{j,j}^o(\mathbf{E}_j \mathbf{E}_j^H)}{\Gamma_{j,j}^o(\mathbf{I}_M)} = 1. \end{aligned} \quad (59)$$

This corresponds to the optimal solution where both parameters are jointly optimized.

2) Robust Joint Optimization with equal regularization:

$$\begin{aligned} (\boldsymbol{\alpha}_{\text{same}}^*, \boldsymbol{\mu}^*) &= \underset{\boldsymbol{\alpha}_{\text{same}}, \boldsymbol{\mu}}{\operatorname{argmax}} \sum_{k=1}^K \log(1 + \text{SINR}_k^o), \\ \text{s.t.} & \sum_{j=1}^n \mu_j^2 \frac{\Gamma_{j,j}^o(\mathbf{E}_j \mathbf{E}_j^H)}{\Gamma_{j,j}^o(\mathbf{I}_M)} = 1. \end{aligned} \quad (60)$$

This corresponds to a jointly optimization for regularization and power scaling, assuming that the regularization coefficient at all TXs are the same.

V. SIMULATION RESULTS

In the following, we provide simulations results to evaluate the accuracy of the deterministic expressions provided and to gain insights into the system design. We also simulate the sum rate performance of the optimal regularization coefficients and power allocation which are robust to the D-CSIT setting.

For the sake of conciseness, the following simulations consider an isotropic channel setting listed in Table I. Similar results can be obtained with cellular setting.

M	K	n	β	Θ_k	P	$\rho_k^{(j,j')}$		
30	30	3	1	\mathbf{I}_M	20dB	fully distributed CSIT	0	
						D-CSIT	0.81	
						centralized CSIT	1	
						$(\sigma_k^{(1)})^2$	$(\sigma_k^{(2)})^2$	$(\sigma_k^{(3)})^2$
asymmetric						0.01	0.16	0.49
symmetric						0.1	0.1	0.1

TABLE I: Simulation parameters for the isotropic channel setting.

In isotropic channel setting, by increasing the value of $\rho_k^{(j,j')}$ from 0 to 1, the CSIT structure for the system gradually changes from fully distributed CSIT to centralized CSIT. For the CSIT discrepancy at different TXs, we consider two cases in the isotropic channel setting: the asymmetric setting where the CSIT accuracy at different TXs are different and the symmetric setting where the CSIT accuracy at different TXs are the same.

In the following simulations of robust regularization and power optimization, we compare the sum rate performance of following algorithms:

- $(\boldsymbol{\alpha}_{\text{naive}}, \boldsymbol{\mu}_{\text{eq}})$: A naive algorithm to obtain the regularization coefficients without considering the D-CSIT configuration, equal power is allocated at each TX (See (51)).
- $(\boldsymbol{\alpha}_{\text{same}}^*, \boldsymbol{\mu}_{\text{eq}})$: A robust optimization of regularization imposing that all TXs have the same regularization coefficient, equal power allocation is assumed at each TX (See (50)).

- (α^*, μ_{eq}) : A robust optimization of regularization with equal power allocation at each TX (See (49)).
- (α_{same}^*, μ^*) : A robust joint optimization of regularization and power at each TX under D-CSIT scenario, with the additional constraint that all TXs have the same regularization coefficient is imposed (See (60)).
- (α^*, μ^*) : A robust joint optimization of regularization and power at each TX under D-CSIT scenario (See (59)).

1) *Monte-Carlo Simulations of Theorem 2*: We verify using Monte-Carlo (MC) simulations the accuracy of the asymptotic expression derived in Theorem 2.

Fig. 2 depicts the absolute error of the deterministic equivalent R_{sum}^0 compared to the ergodic sum rate R_{sum} as a function of the number of users K . The TX nodes number is fixed to $n = 3$. As the K increases, the antennas number at each TX M_{TX} will increase accordingly but the ratio $\beta = \frac{nM_{TX}}{K}$ maintains $\beta = 1$. The ergodic sum rate is averaged over 1000 independent channel realizations. For ease of illustration, we choose the symmetric CSIT configuration and an equal available power per TX. Furthermore, the regularization coefficient at each TX j is chosen as $\alpha^{(j)} = \frac{1}{\beta P}$.

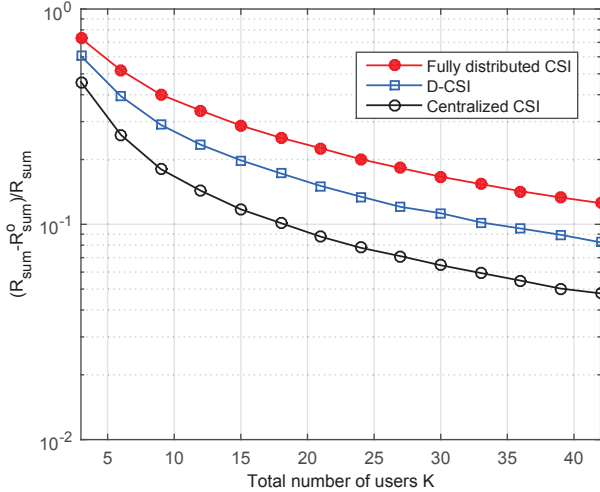


Fig. 2: Relative deviation between the deterministic equivalent and the Monte-Carlo simulations as a function of the number of users K .

It can be seen that the deterministic equivalent converges to the expected sum rate obtained using Monte-Carlo simulations as the system becomes large. It also reveals that the rate of convergence is faster when the CSIT configuration becomes more centralized (i.e., when the CSIT noise becomes more correlated).

2) *Cost of CSIT Distributiveness*: As is mentioned in Section II-B, the CSI estimate noise correlation parameter $\rho_k^{(j,j')}$ reflects the distributiveness of this CoMP network. Let us consider the symmetric accuracy setting, i.e., the CSI estimate noise correlation be $\rho_k^{(j,j')} = \rho, \forall k, \forall j \neq j'$. We plot the ergodic sum rate when the CSI estimate noise correlation ρ varies from 0 to 1, namely, when the CSI structure varies from fully distributed CSI to D-CSI and finally becomes centralized CSI.

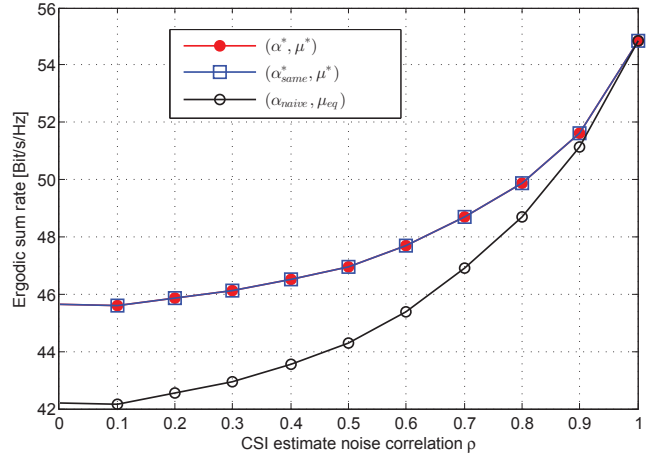


Fig. 3: Ergodic sum rate as a function of estimate noise correlation ρ which indicates the distributiveness for the CSIT, RZF precoding is implemented.

Fig. 3 reveals that the proposed algorithms outperform the naive one in the D-CSIT scenarios. We can also verify that the D-CSI structure introduces a non-vanishing performance degradation compared to the centralized CSI case. We can also observe that the sum rate for (α^*, μ^*) and (α_{same}^*, μ^*) are very close to each other.

3) *Joint Optimization of Regularization and Power for Isotropic Channel*: Let us consider the D-CSIT configuration with asymmetric CSIT accuracy. We then plot the ergodic sum rate as a function of the total transmit power P varies from 0 dB to 30dB.

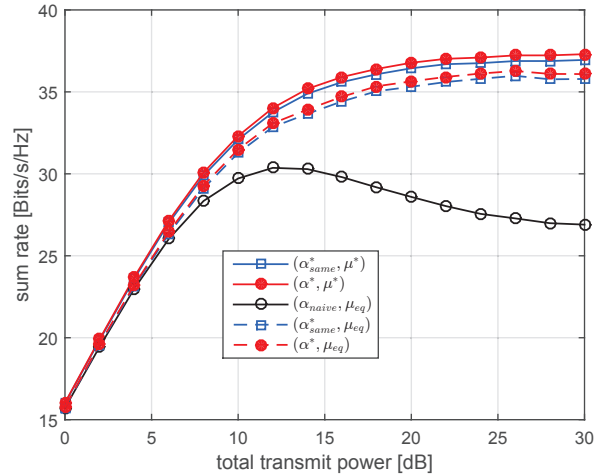


Fig. 4: Ergodic sum rate as a function of total transmit power, comparison between different transmission algorithms, RZF precoding is implemented.

In Fig. 4, the performance of different transmission algorithms are compared. We can clearly observe the improved robustness and the large performance increase for the proposed algorithm. The simulation reveals that the ergodic sum rate performance is not stable if we fail to take into account the D-CSI structure during transmission parameter design. In this

isotropic channel setting, equal power allocation is not a bad strategy since joint optimization $(\boldsymbol{\alpha}^*, \boldsymbol{\mu}^*)$ only brings a 3% sum rate increase compared to $(\boldsymbol{\alpha}^*, \boldsymbol{\mu}_{eq})$. Intriguingly, even if the CSIT accuracy is asymmetric at different TXs, simulation reveals that there is only a negligible performance degradation when imposing identical regularization coefficient at different TXs for isotropic channel setting.

VI. CONCLUSION

In this work, we have studied regularized ZF joint precoding in a distributed CSI configuration. We extend the conventional centralized CSI to distributed CSI scenario by allowing the CSI errors at the different TXs to be arbitrarily correlated. Using RMT tools, an analytical expression is derived to approximate the average rate per user in the large system limit. This deterministic equivalent expression is then used to optimize the regularization coefficients as well as the power allocation at the different TXs in order to reduce the negative impact of the D-CSI configuration.

APPENDIX A

CLASSICAL RANDOM MATRIX THEORY LEMMAS

Lemma 2 (Adapted from [8], [20]). *Let $\alpha^{(j)} > 0, j = 1, \dots, n$ and $m_k^{(j)[t]}, t \geq 0$ be the sequence defined as*

$$\begin{cases} m_k^{(j)[0]} = \frac{1}{\alpha^{(j)}} \\ m_k^{(j)[t]} = \frac{1}{M} \operatorname{tr} \left(\boldsymbol{\Theta}_k \left(\frac{1}{M} \sum_{\ell=1}^K \frac{\boldsymbol{\Theta}_\ell}{1+m_\ell^{(j)[t-1]}} + \alpha^{(j)} \mathbf{I}_M \right)^{-1} \right), \end{cases} \quad (61)$$

$\forall k = 1, \dots, K, t \geq 1$, then $m_k^{(j)[t]} \xrightarrow{t \rightarrow \infty} m_k^{(j)}$, with $m_k^{(j)}$ solved by constructing an iterative algorithm of (61).

Lemma 3 (Resolvent Identities [19], [20]). *Given any matrix $\mathbf{H} \in \mathbb{C}^{K \times M}$, let \mathbf{h}_k^H denote its k th row and $\mathbf{H}_{[k]} \in \mathbb{C}^{(K-1) \times M}$ denote the matrix obtained after removing the k th row from \mathbf{H} . The resolvent matrices of \mathbf{H} and $\mathbf{H}_{[k]}$ are denoted by $\mathbf{Q} = (\mathbf{H}^H \mathbf{H} + \alpha \mathbf{I}_M)^{-1}$ and $\mathbf{Q}_{[k]} = (\mathbf{H}_{[k]}^H \mathbf{H}_{[k]} + \alpha \mathbf{I}_M)^{-1}$, with $\alpha > 0$, respectively. It then holds that*

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

and

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

Lemma 4 ([19], [20]). *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,*

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

Lemma 5 ([19], [20]). *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,

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

Lemma 6 ([8], [20]). *Let \mathbf{Q} and $\mathbf{Q}_{[k]}$ be as given in Lemma 3. Then, for any matrix \mathbf{A} , we have*

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

Lemma 7 ([8], [20]). *Let $\mathbf{U}, \mathbf{V}, \boldsymbol{\Theta}$ be of uniformly bounded spectral norm with respect to N and let \mathbf{V} be invertible. Further, define $\mathbf{x} = \boldsymbol{\Theta}^{\frac{1}{2}} \mathbf{z}$ and $\mathbf{y} = \boldsymbol{\Theta}^{\frac{1}{2}} \mathbf{q}$ where $\mathbf{z}, \mathbf{q} \in \mathbb{C}^N$ have i.i.d. complex entries of zero mean, variance $1/N$ and finite 8th order moment and be mutually independent as well as independent of \mathbf{U}, \mathbf{V} . Define $c_0, c_1, c_2 \in \mathbb{R}^+$ such that $c_0 c_1 - c_2^2 \geq 0$, and let $u = \frac{1}{N} \operatorname{tr}(\boldsymbol{\Theta} \mathbf{V}^{-1})$ and $u' = \frac{1}{N} \operatorname{tr}(\boldsymbol{\Theta} \mathbf{U} \mathbf{V}^{-1})$. Then we have:*

$$\begin{aligned} & \mathbf{x}^H \mathbf{U} (\mathbf{V} + c_0 \mathbf{x} \mathbf{x}^H + c_1 \mathbf{y} \mathbf{y}^H + c_2 \mathbf{x} \mathbf{y}^H + c_2 \mathbf{y} \mathbf{x}^H)^{-1} \mathbf{x} \\ & - \frac{u' (1 + c_1 u)}{(c_0 c_1 - c_2^2) u^2 + (c_0 + c_1) u + 1} \rightarrow 0 \end{aligned}$$

as well as

$$\begin{aligned} & \mathbf{x}^H \mathbf{U} (\mathbf{V} + c_0 \mathbf{x} \mathbf{x}^H + c_1 \mathbf{y} \mathbf{y}^H + c_2 \mathbf{x} \mathbf{y}^H + c_2 \mathbf{y} \mathbf{x}^H)^{-1} \mathbf{y} \\ & - \frac{-c_2 u u'}{(c_0 c_1 - c_2^2) u^2 + (c_0 + c_1) u + 1} \rightarrow 0 \end{aligned}$$

APPENDIX B NEW LEMMAS

Lemma 8. *Consider the channel matrices $\hat{\mathbf{H}}^{(j)}, \hat{\mathbf{H}}^{(j')}$ are distributed according to the D-CSI model in Section II-B. Let*

$$\begin{aligned} \mathbf{Q}^{(j)} &= \left(\frac{(\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j)}}{M} + \alpha^{(j)} \mathbf{I}_M \right)^{-1} \\ \mathbf{Q}^{(j')} &= \left(\frac{(\hat{\mathbf{H}}^{(j')})^H \hat{\mathbf{H}}^{(j')}}{M} + \alpha^{(j')} \mathbf{I}_M \right)^{-1} \end{aligned}$$

with $\alpha^{(j)}, \alpha^{(j')} > 0$. Let $\mathbf{X} \in \mathbb{C}^{M \times M}$ be of uniformly bounded spectral norm with respect to M . Then,

$$\frac{1}{M^2} \operatorname{tr}(\mathbf{X} \mathbf{Q}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}^{(j')}) - \Gamma_{j,j'}^o(\mathbf{X}) \xrightarrow{a.s.} 0$$

where the function $\Gamma_{j,j'}^o(\mathbf{X}) : \mathbb{C}^{M \times M} \mapsto \mathbb{C}$ is defined as

$$\begin{aligned} & \Gamma_{j,j'}^o(\mathbf{X}) \\ &= \frac{1}{M} \sum_{k=1}^K \frac{\left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')}} \rho_k^{(j,j')} \right) \frac{1}{M} \operatorname{tr}(\boldsymbol{\Theta}_k \mathbf{Q}_o^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)})}{(1+m_k^{(j)})(1+m_k^{(j')})} \\ &+ \frac{1}{M} \sum_{k=1}^K \frac{\left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')}} \rho_k^{(j,j')} \right)^2 \Gamma_{j,j'}^o(\boldsymbol{\Theta}_k) \frac{1}{M} \operatorname{tr}(\boldsymbol{\Theta}_k \mathbf{Q}_o^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)})}{(1+m_k^{(j)})(1+m_k^{(j')})}, \end{aligned}$$

with

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

$m_k^{(j)}, \mathbf{Q}_o^{(j)}, m_k^{(j')}, \mathbf{Q}_o^{(j')}$ are defined in Theorem 1 using $\hat{\mathbf{H}}^{(j)}, \alpha^{(j)}, \hat{\mathbf{H}}^{(j')}, \alpha^{(j')}$ respectively. $\Gamma_{j,j'}^o(\boldsymbol{\Theta}_k)$ is the k th entry

of vector $\gamma \in \mathbb{C}^{K \times 1}$. Vector γ is the solution for equation system

$$\mathbf{A}\gamma = \mathbf{b}.$$

$\mathbf{A} \in \mathbb{C}^{K \times K}$ with

$$[\mathbf{A}]_{\ell,t} = 1_{\ell=t} - \frac{\left(\sqrt{c_{0,t}^{(j)} c_{0,t}^{(j')}} + \sqrt{c_{1,t}^{(j)} c_{1,t}^{(j')}} \rho_t^{(j,j')} \right)^2}{M(1+m_t^{(j)})(1+m_t^{(j')})} \cdot \frac{\text{tr} \left(\Theta_t \mathbf{Q}_o^{(j')} \Theta_\ell \mathbf{Q}_o^{(j)} \right)}{M}.$$

$\mathbf{b} \in \mathbb{C}^{K \times 1}$ with

$$[\mathbf{b}]_\ell = \frac{1}{M} \sum_{k=1}^K \frac{\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')}} \rho_k^{(j,j')}}{(1+m_k^{(j)})(1+m_k^{(j')})} \cdot \frac{\text{tr} \left(\Theta_k \mathbf{Q}_o^{(j')} \Theta_\ell \mathbf{Q}_o^{(j)} \right)}{M}.$$

Proof. We start by introducing

$$\mathbf{Q}_{[k]}^{(j)} = \left(\frac{(\hat{\mathbf{H}}_{[k]}^{(j)})^H \hat{\mathbf{H}}_{[k]}^{(j)}}{M} + \alpha^{(j)} \mathbf{I}_M \right)^{-1}$$

with

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

$\mathbf{Q}_{[k]}^{(j')}$ and $\hat{\mathbf{H}}_{[k]}^{(j')}$ are defined respectively in similar manner as $\mathbf{Q}_{[k]}^{(j)}$ and $\hat{\mathbf{H}}_{[k]}^{(j)}$. Let us start by writing the simple equality

$$\begin{aligned} & \mathbf{Q}^{(j)} - \mathbf{Q}_o^{(j)} \\ &= \mathbf{Q}_o^{(j)} \left((\mathbf{Q}_o^{(j)})^{-1} - (\mathbf{Q}^{(j)})^{-1} \right) \mathbf{Q}^{(j)} \\ &= \mathbf{Q}_o^{(j)} \left(\frac{1}{M} \sum_{k=1}^K \frac{\Theta_k}{1+m_k^{(j)}} - \frac{(\hat{\mathbf{H}}^{(j)})^H \mathbf{H}^{(j)}}{M} \right) \mathbf{Q}^{(j)}. \end{aligned} \quad (62)$$

We can then replace $\mathbf{Q}^{(j)}$ using (62) to obtain

$$\begin{aligned} & \frac{1}{M^2} \text{tr} \left(\mathbf{X} \mathbf{Q}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}^{(j')} \right) \\ &= \frac{1}{M^2} \text{tr} \left(\mathbf{X} \mathbf{Q}_o^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}^{(j')} \right) \\ &+ \sum_{k=1}^K \frac{\text{tr} \left(\mathbf{X} \mathbf{Q}_o^{(j)} \Theta_k \mathbf{Q}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}^{(j')} \right)}{M^3 (1+m_k^{(j)})} \\ &- \frac{1}{M^3} \text{tr} \left(\mathbf{X} \mathbf{Q}_o^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j)} \mathbf{Q}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}^{(j')} \right) \\ &= Z_1 + Z_2 + Z_3. \end{aligned}$$

We will now calculate separately each of the term Z_i . Starting with Z_1 gives

$$\begin{aligned} Z_1 &= \frac{1}{M^2} \text{tr} \left(\mathbf{X} \mathbf{Q}_o^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}^{(j')} \right) \\ &= \frac{1}{M} \sum_{k=1}^K \frac{1}{M} (\hat{h}_k^{(j')})^H \mathbf{Q}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \hat{h}_k^{(j)} \\ &\stackrel{(a)}{=} \frac{1}{M} \sum_{k=1}^K \frac{1}{M} \frac{(\hat{h}_k^{(j')})^H \mathbf{Q}_{[k]}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \hat{h}_k^{(j)}}{1 + \frac{1}{M} (\hat{h}_k^{(j')})^H \mathbf{Q}_{[k]}^{(j')} \hat{h}_k^{(j)}} \\ &\stackrel{(b)}{=} \frac{1}{M} \sum_{k=1}^K \frac{\left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')}} \rho_k^{(j,j')} \right) \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_{[k]}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \right)}{1 + \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_{[k]}^{(j')} \right)} \\ &\stackrel{(c)}{=} \frac{1}{M} \sum_{k=1}^K \frac{\left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')}} \rho_k^{(j,j')} \right) \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \right)}{1 + \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}^{(j')} \right)} \\ &\stackrel{(d)}{=} \frac{1}{M} \sum_{k=1}^K \frac{\left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')}} \rho_k^{(j,j')} \right) \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_o^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \right)}{1 + m_k^{(j')}} \end{aligned}$$

where equality (a) follows from Lemma 3, equality (b) from Lemma 4, equality (c) from Lemma 6, and equality (d) from the fundamental Theorem 1. The following calculations are very similar and the same lemmas are used.

We put aside Z_2 for now and turn to Z_3 which gives

$$\begin{aligned} Z_3 &= -\frac{1}{M^3} \text{tr} \left(\mathbf{X} \mathbf{Q}_o^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j)} \mathbf{Q}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}^{(j')} \right) \\ &= -\frac{1}{M^3} \sum_{k=1}^K \text{tr} \left((\hat{h}_k^{(j)})^H \mathbf{Q}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \hat{h}_k^{(j)} \right) \\ &= -\frac{1}{M^3} \sum_{k=1}^K \frac{\text{tr} \left((\hat{h}_k^{(j)})^H \mathbf{Q}_{[k]}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \hat{h}_k^{(j)} \right)}{1 + \frac{1}{M} (\hat{h}_k^{(j)})^H \mathbf{Q}_{[k]}^{(j)} \hat{h}_k^{(j)}} \\ &\stackrel{(e)}{=} -\frac{1}{M^3} \sum_{k=1}^K \frac{\text{tr} \left((\hat{h}_k^{(j)})^H \mathbf{Q}_{[k]}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}_{[k]}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \hat{h}_k^{(j)} \right)}{1 + \frac{1}{M} (\hat{h}_k^{(j)})^H \mathbf{Q}_{[k]}^{(j)} \hat{h}_k^{(j)}} \\ &+ \frac{1}{M^4} \sum_{k=1}^K \frac{\text{tr} \left((\hat{h}_k^{(j)})^H \mathbf{Q}_{[k]}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}_{[k]}^{(j')} \hat{h}_k^{(j')} (\hat{h}_k^{(j')})^H \mathbf{Q}_{[k]}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \hat{h}_k^{(j)} \right)}{\left(1 + \frac{1}{M} (\hat{h}_k^{(j)})^H \mathbf{Q}_{[k]}^{(j)} \hat{h}_k^{(j)} \right) \left(1 + \frac{1}{M} (\hat{h}_k^{(j')})^H \mathbf{Q}_{[k]}^{(j')} \hat{h}_k^{(j')} \right)} \\ &= Z_4 + Z_5, \end{aligned}$$

with equality (e) obtained using Lemma 3 for $\mathbf{Q}^{(j')}$. We also split the calculation in two and start by calculating Z_4 as follows.

$$\begin{aligned} Z_4 &= -\frac{1}{M^3} \sum_{k=1}^K \frac{\text{tr} \left((\hat{h}_k^{(j)})^H \mathbf{Q}_{[k]}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}_{[k]}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \hat{h}_k^{(j)} \right)}{1 + \frac{1}{M} (\hat{h}_k^{(j)})^H \mathbf{Q}_{[k]}^{(j)} \hat{h}_k^{(j)}} \\ &- \frac{1}{M^3} \sum_{k=1}^K \frac{\text{tr} \left((\hat{h}_k^{(j)})^H \mathbf{Q}_{[k]}^{(j)} \hat{h}_k^{(j)} (\hat{h}_k^{(j')})^H \mathbf{Q}_{[k]}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \hat{h}_k^{(j)} \right)}{1 + \frac{1}{M} (\hat{h}_k^{(j)})^H \mathbf{Q}_{[k]}^{(j)} \hat{h}_k^{(j)}} \\ &\asymp -\frac{1}{M^3} \sum_{k=1}^K \frac{\text{tr} \left(\Theta_k \mathbf{Q}_{[k]}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}_{[k]}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \right)}{1 + \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_{[k]}^{(j)} \right)} \\ &- \frac{1}{M} \sum_{k=1}^K \left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')}} \rho_k^{(j,j')} \right) \\ &\cdot \frac{\frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_{[k]}^{(j)} \right) \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_{[k]}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \right)}{1 + \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_{[k]}^{(j)} \right)} \end{aligned} \quad (63)$$

$$\begin{aligned}
 & \stackrel{(f)}{\asymp} \frac{1}{M} \sum_{k=1}^K \frac{\frac{1}{M^2} \text{tr} \left(\Theta_k \mathbf{Q}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \right)}{1 + m_k^{(j)}} \\
 & - \frac{1}{M} \sum_{k=1}^K \left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')} \rho_k^{(j,j')}} \right) \\
 & \cdot \frac{\frac{m_k^{(j)}}{M} \text{tr} \left(\Theta_k \mathbf{Q}_o^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \right)}{1 + m_k^{(j)}} \\
 & \asymp -Z_2 - \frac{1}{M} \sum_{k=1}^K \left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')} \rho_k^{(j,j')}} \right) \\
 & \cdot \frac{\frac{m_k^{(j)}}{M} \text{tr} \left(\Theta_k \mathbf{Q}_o^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \right)}{1 + m_k^{(j)}},
 \end{aligned}$$

where (f) applies multiple times Lemma 6.

Finally, Z_5 is calculated as

$$\begin{aligned}
 Z_5 & \asymp \frac{1}{M^4} \sum_{k=1}^K \frac{\text{tr} \left((\hat{\mathbf{h}}_k^{(j)})^H \mathbf{Q}_{[k]}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}_{[k]}^{(j')} \hat{\mathbf{h}}_k^{(j')} (\hat{\mathbf{h}}_k^{(j')})^H \mathbf{Q}_{[k]}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \hat{\mathbf{h}}_k^{(j)} \right)}{(1 + m_k^{(j)}) (1 + m_k^{(j')})} \\
 & = \frac{1}{M^4} \sum_{k=1}^K \frac{\text{tr} \left((\hat{\mathbf{h}}_k^{(j)})^H \mathbf{Q}_{[k]}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}_{[k]}^{(j')} \hat{\mathbf{h}}_k^{(j')} (\hat{\mathbf{h}}_k^{(j')})^H \mathbf{Q}_{[k]}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \hat{\mathbf{h}}_k^{(j)} \right)}{(1 + m_k^{(j)}) (1 + m_k^{(j')})} \\
 & + \frac{1}{M^4} \sum_{k=1}^K \frac{\text{tr} \left((\hat{\mathbf{h}}_k^{(j)})^H \mathbf{Q}_{[k]}^{(j)} \hat{\mathbf{h}}_k^{(j)} (\hat{\mathbf{h}}_k^{(j')})^H \mathbf{Q}_{[k]}^{(j')} \hat{\mathbf{h}}_k^{(j')} (\hat{\mathbf{h}}_k^{(j')})^H \mathbf{Q}_{[k]}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \hat{\mathbf{h}}_k^{(j)} \right)}{(1 + m_k^{(j)}) (1 + m_k^{(j')})} \\
 & \asymp \frac{1}{M} \sum_{k=1}^K \left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')} \rho_k^{(j,j')}} \right)^2 \\
 & \cdot \frac{\frac{1}{M^2} \text{tr} \left(\Theta_k \mathbf{Q}_{[k]}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}_{[k]}^{(j')} \right) \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_{[k]}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \right)}{(1 + m_k^{(j)}) (1 + m_k^{(j')})} \\
 & + \frac{1}{M} \sum_{k=1}^K \left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')} \rho_k^{(j,j')}} \right) \\
 & \cdot \frac{\frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_{[k]}^{(j)} \right) \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_{[k]}^{(j')} \right) \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_{[k]}^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \right)}{(1 + m_k^{(j)}) (1 + m_k^{(j')})} \\
 & \asymp \frac{1}{M} \sum_{k=1}^K \left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')} \rho_k^{(j,j')}} \right)^2 \\
 & \cdot \frac{\frac{1}{M^2} \text{tr} \left(\Theta_k \mathbf{Q}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}^{(j')} \right) \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_o^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \right)}{(1 + m_k^{(j)}) (1 + m_k^{(j')})} \\
 & + \frac{1}{M} \sum_{k=1}^K \left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')} \rho_k^{(j,j')}} \right) \frac{m_k^{(j)} m_k^{(j')} \text{tr} \left(\Theta_k \mathbf{Q}_o^{(j)} \mathbf{X} \mathbf{Q}_o^{(j')} \right)}{(1 + m_k^{(j)}) (1 + m_k^{(j')})}.
 \end{aligned}$$

Adding all the Z_i gives

$$\begin{aligned}
 & \frac{1}{M^2} \text{tr} \left(\mathbf{X} \mathbf{Q}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}^{(j')} \right) \\
 & \asymp \frac{1}{M} \sum_{k=1}^K \frac{\left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')} \rho_k^{(j,j')}} \right) \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_o^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \right)}{(1 + m_k^{(j)}) (1 + m_k^{(j')})} \\
 & + \frac{1}{M} \sum_{k=1}^K \frac{1}{M^2} \text{tr} \left(\Theta_k \mathbf{Q}^{(j)} (\hat{\mathbf{H}}^{(j)})^H \hat{\mathbf{H}}^{(j')} \mathbf{Q}^{(j')} \right) \\
 & \cdot \frac{\left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')} \rho_k^{(j,j')}} \right)^2 \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_o^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \right)}{(1 + m_k^{(j)}) (1 + m_k^{(j')})},
 \end{aligned}$$

which indicates

$$\begin{aligned}
 & \Gamma_{j,j'}^o(\mathbf{X}) \\
 & = \frac{1}{M} \sum_{k=1}^K \frac{\left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')} \rho_k^{(j,j')}} \right) \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_o^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \right)}{(1 + m_k^{(j)}) (1 + m_k^{(j')})} \\
 & + \frac{1}{M} \sum_{k=1}^K \frac{\left(\sqrt{c_{0,k}^{(j)} c_{0,k}^{(j')}} + \sqrt{c_{1,k}^{(j)} c_{1,k}^{(j')} \rho_k^{(j,j')}} \right)^2 \Gamma_{j,j'}^o(\Theta_k) \frac{1}{M} \text{tr} \left(\Theta_k \mathbf{Q}_o^{(j')} \mathbf{X} \mathbf{Q}_o^{(j)} \right)}{(1 + m_k^{(j)}) (1 + m_k^{(j')})}.
 \end{aligned} \tag{64}$$

It remains then to calculate $\Gamma_{j,j'}^o(\Theta_k)$ to conclude the calculation. Indeed, it is the solution of equation system when asserting $\mathbf{X} = \Theta_k, \forall k = 1, \dots, K$ into (64). \square

Lemma 9. Let $\mathbf{L}, \mathbf{R}, \bar{\mathbf{A}}, \Theta \in \mathbb{C}^{M \times M}$ be of uniformly bounded spectral norm with respect to M and let $\bar{\mathbf{A}}$ be invertible. Further define $\mathbf{x} = \Theta^{\frac{1}{2}} \mathbf{z}$, $\mathbf{x}' = \Theta^{\frac{1}{2}} \mathbf{z}'$ and $\mathbf{y} = \Theta^{\frac{1}{2}} \mathbf{q}$. \mathbf{z}, \mathbf{z}' satisfies $\mathbf{z} = \rho \mathbf{z}' + \sqrt{1 - \rho^2} \mathbf{w}$. \mathbf{z}, \mathbf{q} and $\mathbf{z}', \mathbf{q}, \mathbf{w}$ are mutually independent as well as independent of $\mathbf{L}, \mathbf{R}, \bar{\mathbf{A}}$. $\mathbf{z}, \mathbf{z}', \mathbf{q}, \mathbf{w}$ have i.i.d. complex entries of zero mean, variance $1/M$ and finite 8th order moment. Let us define

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

let $c_0, c_1, c_2 \in \mathbb{R}^+$ with $c_0 + c_1 = 1$ and $c_0 c_1 - c_2^2 = 0$, and

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

Then we have:

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

Proof. Focusing first on the first equality gives

$$\begin{aligned}
 & \mathbf{x}^H \mathbf{L} \mathbf{A}^{-1} \mathbf{R} \mathbf{x} - \mathbf{x}^H \mathbf{L} \bar{\mathbf{A}}^{-1} \mathbf{R} \mathbf{x} \\
 & = \mathbf{x}^H \mathbf{L} \mathbf{A}^{-1} (\bar{\mathbf{A}} - \mathbf{A}) \bar{\mathbf{A}}^{-1} \mathbf{R} \mathbf{x} \\
 & = -\mathbf{x}^H \mathbf{L} \mathbf{A}^{-1} (c_0 \mathbf{x} \mathbf{x}^H + c_1 \mathbf{y} \mathbf{y}^H + c_2 \mathbf{x} \mathbf{y}^H + c_2 \mathbf{y} \mathbf{x}^H) \bar{\mathbf{A}}^{-1} \mathbf{R} \mathbf{x} \\
 & \stackrel{(a)}{\asymp} - (c_0 \mathbf{x}^H \mathbf{L} \mathbf{A}^{-1} \mathbf{x} + c_2 \mathbf{x}^H \mathbf{L} \mathbf{A}^{-1} \mathbf{y}) \text{tr}(\Theta \bar{\mathbf{A}}^{-1} \mathbf{R}) \\
 & \stackrel{(b)}{\asymp} -c_0 \frac{\text{tr}(\Theta \bar{\mathbf{L}} \bar{\mathbf{A}}^{-1})}{M} \frac{\text{tr}(\Theta \bar{\mathbf{A}}^{-1} \mathbf{R})}{M} \frac{1 + c_1 \frac{\text{tr}(\Theta \bar{\mathbf{A}}^{-1})}{M}}{1 + \frac{\text{tr}(\Theta \bar{\mathbf{A}}^{-1})}{M}} \\
 & + c_2^2 \frac{\text{tr}(\Theta \bar{\mathbf{L}} \bar{\mathbf{A}}^{-1})}{M} \frac{\text{tr}(\Theta \bar{\mathbf{A}}^{-1} \mathbf{R})}{M} \frac{\frac{\text{tr}(\Theta \bar{\mathbf{A}}^{-1})}{M}}{1 + \frac{\text{tr}(\Theta \bar{\mathbf{A}}^{-1})}{M}},
 \end{aligned}$$

where equality (a) is obtained from using Lemma 5 and Lemma 4 and equality (b) follows from Lemma 7. Similarly, we turn to the second equality to write

$$\begin{aligned}
 & \mathbf{x}^H \mathbf{L} \mathbf{A}^{-1} \mathbf{R} \mathbf{y} - \mathbf{x}^H \mathbf{L} \bar{\mathbf{A}}^{-1} \mathbf{R} \mathbf{y} \\
 &= \mathbf{x}^H \mathbf{L} \mathbf{A}^{-1} (\bar{\mathbf{A}} - \mathbf{A}) \bar{\mathbf{A}}^{-1} \mathbf{R} \mathbf{y} \\
 &= -\mathbf{x}^H \mathbf{L} \mathbf{A}^{-1} (c_0 \mathbf{x} \mathbf{x}^H + c_1 \mathbf{y} \mathbf{y}^H + c_2 \mathbf{y} \mathbf{x}^H + c_2 \mathbf{x} \mathbf{y}^H) \bar{\mathbf{A}}^{-1} \mathbf{R} \mathbf{y} \\
 &\stackrel{(c)}{\asymp} - (c_1 \mathbf{x}^H \mathbf{L} \mathbf{A}^{-1} \mathbf{y} + c_2 \mathbf{x}^H \mathbf{L} \mathbf{A}^{-1} \mathbf{x}) \frac{\text{tr}(\Theta \bar{\mathbf{A}}^{-1} \mathbf{R})}{M} \\
 &\stackrel{(d)}{\asymp} c_1 c_2 \frac{\text{tr}(\Theta \mathbf{L} \bar{\mathbf{A}}^{-1})}{M} \frac{\text{tr}(\Theta \bar{\mathbf{A}}^{-1} \mathbf{R})}{M} \frac{\text{tr}(\Theta \bar{\mathbf{A}}^{-1})}{1 + \frac{\text{tr}(\Theta \bar{\mathbf{A}}^{-1})}{M}} \\
 &\quad - c_2 \frac{\text{tr}(\Theta \mathbf{L} \bar{\mathbf{A}}^{-1})}{M} \frac{\text{tr}(\Theta \bar{\mathbf{A}}^{-1} \mathbf{R})}{M} \frac{1 + c_1 \frac{\text{tr}(\Theta \bar{\mathbf{A}}^{-1})}{M}}{1 + \frac{\text{tr}(\Theta \bar{\mathbf{A}}^{-1})}{M}},
 \end{aligned}$$

where equality (c) is obtained from using Lemma 5 and Lemma 4 and equality (d) follows from Lemma 7. For the third equality,

$$\begin{aligned}
 & \mathbf{x}^H \mathbf{L} \mathbf{A}'^{-1} \mathbf{R} \mathbf{y} \\
 &= \rho \mathbf{x}'^H \mathbf{L} \mathbf{A}'^{-1} \mathbf{R} \mathbf{y} + \sqrt{1 - \rho^2} \Theta^{\frac{1}{2}} \mathbf{w} \mathbf{L} \mathbf{A}'^{-1} \mathbf{R} \mathbf{y} \\
 &\stackrel{(e)}{\asymp} \rho \mathbf{x}'^H \mathbf{L} \mathbf{A}'^{-1} \mathbf{R} \mathbf{y} \\
 &\asymp -\rho \frac{c_2 u_L u_R}{1 + u},
 \end{aligned}$$

where equality (e) is obtained from using Lemma 5. \square

APPENDIX C

PROOF OF DETERMINISTIC EQUIVALENT THEOREM 2

The proof is built upon results from both [8] and [19] and novel lemmas Lemma 8 and Lemma 9. We also make extensive use of classical RMT lemmas recalled in Appendix A. In particular, Lemma 8 extends [19, Lemma 15] and is an interesting result in itself.

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

We start by finding a deterministic equivalent for $\Psi^{(j)}$. Apply Lemma 8 with $\hat{\mathbf{H}}^{(j')} = \hat{\mathbf{H}}^{(j)}$, $\mathbf{A} = \mathbf{I}_M$, which gives

$$\begin{aligned}
 \Psi^{(j)} &\asymp \Gamma_{j,j}^o(\mathbf{I}_M) \\
 &= \frac{1}{M} \sum_{\ell=1}^K \frac{\frac{1}{M} \text{tr}(\Theta_\ell \mathbf{Q}_o^{(j)} \mathbf{Q}_o^{(j)})}{(1 + m_\ell^{(j)})^2} \\
 &\quad + \frac{1}{M} \sum_{\ell=1}^K \frac{\Gamma_{j,j}^o(\Theta_\ell) \frac{1}{M} \text{tr}(\Theta_\ell \mathbf{Q}_o^{(j)} \mathbf{Q}_o^{(j)})}{(1 + m_\ell^{(j)})^2}. \quad (65)
 \end{aligned}$$

From (65), it can be noted that, as expected, this deterministic equivalent does not depend on $\sigma_\ell^{(j)}$. The total power constraint for large scale system reads

$$\begin{aligned}
 & \|\mathbf{T}_{\text{rZF}}^{\text{DCSI}}\|_{\text{F}}^2 \\
 &= \sum_{j=1}^n \mu_j^2 \text{tr}(\mathbf{E}_j^H \mathbf{T}_{\text{rZF}}^{(j)} (\mathbf{T}_{\text{rZF}}^{(j)})^H \mathbf{E}_j) \\
 &\stackrel{(a)}{\asymp} \sum_{j=1}^n \mu_j^2 \frac{P}{\Gamma_{j,j}^o(\mathbf{I}_M)} \Gamma_{j,j}^o(\mathbf{E}_j \mathbf{E}_j^H) \\
 &= P,
 \end{aligned}$$

where (a) follows from Lemma 8. Therefore, there is a constraint for the power scaling factors μ_j :

$$\sum_{j=1}^n \mu_j^2 \frac{\Gamma_{j,j}^o(\mathbf{E}_j \mathbf{E}_j^H)}{\Gamma_{j,j}^o(\mathbf{I}_M)} = 1.$$

B. Deterministic equivalent for $\mathbf{h}_k^H \mathbf{t}_{\text{rZF},k}^{\text{DCSI}}$

Turning to the desired signal at RX k , we can write

$$\begin{aligned}
 \mathbf{h}_k^H \mathbf{t}_{\text{rZF},k}^{\text{DCSI}} &= \sum_{j=1}^n \frac{1}{M} \frac{\mu_j \sqrt{P}}{\sqrt{\Psi^{(j)}}} \mathbf{h}_k^H \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}^{(j)})^{-1} \hat{\mathbf{h}}_k^{(j)} \\
 &\stackrel{(a)}{\asymp} \sqrt{P} \sum_{j=1}^n \mu_j \sqrt{\frac{1}{\Gamma_{j,j}^o(\mathbf{I}_M)} \frac{\frac{1}{M} \mathbf{h}_k^H \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}_{[k]}^{(j)})^{-1} \hat{\mathbf{h}}_k^{(j)}}{1 + \frac{1}{M} (\hat{\mathbf{h}}_k^{(j)})^H (\mathbf{C}_{[k]}^{(j)})^{-1} \hat{\mathbf{h}}_k^{(j)}}}} \\
 &\stackrel{(b)}{\asymp} \sqrt{P} \sum_{j=1}^n \mu_j \sqrt{\frac{1 - (\sigma_k^{(j)})^2}{\Gamma_{j,j}^o(\mathbf{I}_M)} \frac{\frac{1}{M} \mathbf{h}_k^H \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}_{[k]}^{(j)})^{-1} \mathbf{h}_k}{1 + \frac{1}{M} (\hat{\mathbf{h}}_k^{(j)})^H (\mathbf{C}_{[k]}^{(j)})^{-1} \hat{\mathbf{h}}_k^{(j)}}}} \\
 &\stackrel{(c)}{\asymp} \sqrt{P} \sum_{j=1}^n \mu_j \sqrt{\frac{1 - (\sigma_k^{(j)})^2}{\Gamma_{j,j}^o(\mathbf{I}_M)} \frac{\frac{1}{M} \text{tr}(\Theta_k \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}_{[k]}^{(j)})^{-1})}{1 + \frac{1}{M} \text{tr}(\Theta_k (\mathbf{C}_{[k]}^{(j)})^{-1})}} \\
 &\stackrel{(d)}{\asymp} \sqrt{P} \sum_{j=1}^n \mu_j \sqrt{\frac{1 - (\sigma_k^{(j)})^2}{\Gamma_{j,j}^o(\mathbf{I}_M)} \frac{\frac{1}{M} \text{tr}(\Theta_k \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}^{(j)})^{-1})}{1 + \frac{1}{M} \text{tr}(\Theta_k (\mathbf{C}^{(j)})^{-1})}} \\
 &\stackrel{(e)}{\asymp} \sqrt{P} \sum_{j=1}^n \mu_j \sqrt{\frac{1 - (\sigma_k^{(j)})^2}{\Gamma_{j,j}^o(\mathbf{I}_M)} \frac{\frac{1}{M} \text{tr}(\Theta_k \mathbf{E}_j \mathbf{E}_j^H \mathbf{Q}_o^{(j)})}{1 + \frac{1}{M} \text{tr}(\Theta_k \mathbf{Q}_o^{(j)})}},
 \end{aligned}$$

where we have defined

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

with

$$(\hat{\mathbf{H}}_{[k]}^{(j)})^H = \left[\hat{\mathbf{h}}_1^{(j)} \quad \dots \quad \hat{\mathbf{h}}_{k-1}^{(j)} \quad \hat{\mathbf{h}}_{k+1}^{(j)} \quad \dots \quad \hat{\mathbf{h}}_K^{(j)} \right], \quad \forall j.$$

Equality (a) follows then from Lemma 3 and the use of the deterministic equivalent derived for $\Psi^{(j)}$, (b) from Lemma 5, (c) from Lemma 4, (d) from Lemma 6 and (e) from the fundamental Theorem 1.

It follows then directly that

$$|\mathbf{h}_k^H \mathbf{t}_{\text{rZF},k}^{\text{DCSI}}|^2 \asymp P \left(\sum_{j=1}^n \mu_j \sqrt{\frac{1 - (\sigma_k^{(j)})^2}{\Gamma_{j,j}^o(\mathbf{I}_M)} \frac{\frac{1}{M} \text{tr}(\Theta_k \mathbf{E}_j \mathbf{E}_j^H \mathbf{Q}_o^{(j)})}{1 + \frac{1}{M} \text{tr}(\Theta_k \mathbf{Q}_o^{(j)})}} \right)^2.$$

C. Deterministic Equivalent for \mathcal{I}_k

Our first step is to write explicitly the interference term using the definition of \mathbf{T}^{DCSI} and replace $\Psi^{(j)}$ by its deter-

ministic equivalent.

$$\begin{aligned}
 \mathcal{I}_k &= \sum_{\ell=1, \ell \neq k}^K |\mathbf{h}_k^H \mathbf{t}_{\text{rZF}, \ell}^{\text{DCSI}}|^2 \\
 &= \mathbf{h}_k^H \mathbf{T}_{\text{rZF}}^{\text{DCSI}} (\mathbf{T}_{\text{rZF}}^{\text{DCSI}})^H \mathbf{h}_k - \mathbf{h}_k^H \mathbf{t}_{\text{rZF}, k}^{\text{DCSI}} (\mathbf{t}_{\text{rZF}, k}^{\text{DCSI}})^H \mathbf{h}_k \\
 &= \frac{P}{M^2} \sum_{j=1}^n \sum_{j'=1}^n \frac{\mu_j \mu_{j'} \mathbf{h}_k^H \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}^{(j)})^{-1} (\hat{\mathbf{H}}_{[k]}^{(j)})^H \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}^{(j')})^{-1} \mathbf{E}_{j'} \mathbf{E}_{j'}^H \mathbf{h}_k}{\sqrt{\Psi(j)\Psi(j')}} \\
 &\asymp \frac{P}{M^2} \sum_{j=1}^n \sum_{j'=1}^n \frac{\mu_j \mu_{j'}}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \mathbf{h}_k^H \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}^{(j)})^{-1} \\
 &\quad \cdot (\hat{\mathbf{H}}_{[k]}^{(j)})^H \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}^{(j')})^{-1} \mathbf{E}_{j'} \mathbf{E}_{j'}^H \mathbf{h}_k \\
 &+ \frac{P}{M^2} \sum_{j=1}^n \sum_{j'=1}^n \frac{\mu_j \mu_{j'}}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \mathbf{h}_k^H \mathbf{E}_j \mathbf{E}_j^H \\
 &\quad \cdot ((\mathbf{C}^{(j)})^{-1} - (\mathbf{C}_{[k]}^{(j)})^{-1}) (\hat{\mathbf{H}}_{[k]}^{(j)})^H \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}^{(j')})^{-1} \mathbf{E}_{j'} \mathbf{E}_{j'}^H \mathbf{h}_k. \quad (66)
 \end{aligned}$$

To obtain a deterministic equivalent for the second summation in (66) we use the following relation

$$\begin{aligned}
 &(\mathbf{C}^{(j)})^{-1} - (\mathbf{C}_{[k]}^{(j)})^{-1} \\
 &= (\mathbf{C}^{(j)})^{-1} (\mathbf{C}_{[k]}^{(j)} - \mathbf{C}^{(j)}) (\mathbf{C}_{[k]}^{(j)})^{-1} \\
 &= -\frac{(\mathbf{C}^{(j)})^{-1}}{M} (c_{0,k}^{(j)} \mathbf{h}_k \mathbf{h}_k^H + c_{1,k}^{(j)} \boldsymbol{\delta}_k^{(j)} (\boldsymbol{\delta}_k^{(j)})^H \\
 &\quad + c_{2,k}^{(j)} \boldsymbol{\delta}_k^{(j)} \mathbf{h}_k^H + c_{2,k}^{(j)} \mathbf{h}_k (\boldsymbol{\delta}_k^{(j)})^H) (\mathbf{C}_{[k]}^{(j)})^{-1}, \quad (67)
 \end{aligned}$$

where $c_{0,k}^{(j)}, c_{1,k}^{(j)}, c_{2,k}^{(j)}$ is defined in (28). It is important to note that

$$c_{0,k}^{(j)} c_{1,k}^{(j)} = (c_{2,k}^{(j)})^2, c_{0,k}^{(j)} + c_{1,k}^{(j)} = 1,$$

as these relations will be used several times through the proof. Inserting (67) into (66), the interference term can be denoted as

$$\begin{aligned}
 \tilde{\mathcal{I}}_k &\asymp \frac{P}{M^2} \sum_{j=1}^n \sum_{j'=1}^n \frac{\mu_j \mu_{j'}}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \mathbf{h}_k^H \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}_{[k]}^{(j)})^{-1} \\
 &\quad \cdot (\hat{\mathbf{H}}_{[k]}^{(j)})^H \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}^{(j')})^{-1} \mathbf{E}_{j'} \mathbf{E}_{j'}^H \mathbf{h}_k \\
 &- \frac{P}{M^3} \sum_{j=1}^n \sum_{j'=1}^n \frac{\mu_j \mu_{j'}}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \mathbf{h}_k^H \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}^{(j)})^{-1} \\
 &\quad \cdot [\mathbf{h}_k c_{0,k}^{(j)} \mathbf{h}_k^H] (\mathbf{C}_{[k]}^{(j)})^{-1} (\hat{\mathbf{H}}_{[k]}^{(j)})^H \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}^{(j')})^{-1} \mathbf{E}_{j'} \mathbf{E}_{j'}^H \mathbf{h}_k \\
 &- \frac{P}{M^3} \sum_{j=1}^n \sum_{j'=1}^n \frac{\mu_j \mu_{j'}}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \mathbf{h}_k^H \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}^{(j)})^{-1} \\
 &\quad \cdot [\boldsymbol{\delta}_k^{(j)} c_{1,k}^{(j)} (\boldsymbol{\delta}_k^{(j)})^H] (\mathbf{C}_{[k]}^{(j)})^{-1} (\hat{\mathbf{H}}_{[k]}^{(j)})^H \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}^{(j')})^{-1} \mathbf{E}_{j'} \mathbf{E}_{j'}^H \mathbf{h}_k \\
 &- \frac{P}{M^3} \sum_{j=1}^n \sum_{j'=1}^n \frac{\mu_j \mu_{j'}}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \mathbf{h}_k^H \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}^{(j)})^{-1} \\
 &\quad \cdot [\boldsymbol{\delta}_k^{(j)} c_{2,k}^{(j)} \mathbf{h}_k^H] (\mathbf{C}_{[k]}^{(j)})^{-1} (\hat{\mathbf{H}}_{[k]}^{(j)})^H \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}^{(j')})^{-1} \mathbf{E}_{j'} \mathbf{E}_{j'}^H \mathbf{h}_k \\
 &- \frac{P}{M^3} \sum_{j=1}^n \sum_{j'=1}^n \frac{\mu_j \mu_{j'}}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \mathbf{h}_k^H \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}^{(j)})^{-1} \\
 &\quad \cdot [\mathbf{h}_k c_{2,k}^{(j)} (\boldsymbol{\delta}_k^{(j)})^H] (\mathbf{C}_{[k]}^{(j)})^{-1} (\hat{\mathbf{H}}_{[k]}^{(j)})^H \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}^{(j')})^{-1} \mathbf{E}_{j'} \mathbf{E}_{j'}^H \mathbf{h}_k \\
 &= A + B + C + D + E. \quad (68)
 \end{aligned}$$

We proceed by calculating terms A to E in (68) successively, using Lemma 9. For the sake of simplicity, we only proceed

the calculation of term A and the rest terms can be calculated in similar manner.

$$\begin{aligned}
 A &= \frac{P}{M^2} \sum_{j=1}^n \sum_{j'=1}^n \frac{\mu_j \mu_{j'}}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \mathbf{h}_k^H \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}_{[k]}^{(j)})^{-1} \\
 &\quad \cdot (\hat{\mathbf{H}}_{[k]}^{(j)})^H \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}^{(j')})^{-1} \mathbf{E}_{j'} \mathbf{E}_{j'}^H \mathbf{h}_k \\
 &\asymp P \sum_{j=1}^n \sum_{j'=1}^n \frac{\mu_j \mu_{j'}}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \\
 &\quad \cdot \left[\frac{\text{tr}(\mathbf{E}_{j'} \mathbf{E}_{j'}^H \boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}_{[k]}^{(j)})^{-1} (\hat{\mathbf{H}}_{[k]}^{(j)})^H \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}_{[k]}^{(j')})^{-1})}{M^2} \right. \\
 &\quad - c_{0,k}^{(j')} \frac{\text{tr}(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}_{[k]}^{(j)})^{-1} (\hat{\mathbf{H}}_{[k]}^{(j)})^H \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}_{[k]}^{(j')})^{-1})}{M^2} \\
 &\quad \cdot \frac{\text{tr}(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}_{[k]}^{(j')})^{-1})}{M} \frac{1 + c_{1,k}^{(j')} \frac{\text{tr}(\boldsymbol{\Theta}_k (\mathbf{C}_{[k]}^{(j')})^{-1})}{M}}{1 + \frac{\text{tr}(\boldsymbol{\Theta}_k (\mathbf{C}_{[k]}^{(j')})^{-1})}{M}} \\
 &\quad \left. + (c_{2,k}^{(j')})^2 \frac{\text{tr}(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}_{[k]}^{(j)})^{-1} (\hat{\mathbf{H}}_{[k]}^{(j)})^H \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}_{[k]}^{(j')})^{-1})}{M^2} \right. \\
 &\quad \left. \cdot \frac{\text{tr}(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}_{[k]}^{(j')})^{-1})}{M} \frac{\text{tr}(\boldsymbol{\Theta}_k (\mathbf{C}_{[k]}^{(j')})^{-1})}{1 + \frac{\text{tr}(\boldsymbol{\Theta}_k (\mathbf{C}_{[k]}^{(j')})^{-1})}{M}} \right]. \quad (69)
 \end{aligned}$$

According to Lemma 6 and Lemma 8, we can have:

$$\begin{aligned}
 &\frac{1}{M^2} \text{tr}(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}_{[k]}^{(j)})^{-1} (\hat{\mathbf{H}}_{[k]}^{(j)})^H \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}_{[k]}^{(j')})^{-1}) \\
 &\asymp \Gamma_{j,j'}^{\circ}(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^H) \\
 &\frac{1}{M^2} \text{tr}(\mathbf{E}_{j'} \mathbf{E}_{j'}^H \boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^H (\mathbf{C}_{[k]}^{(j)})^{-1} (\hat{\mathbf{H}}_{[k]}^{(j)})^H \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}_{[k]}^{(j')})^{-1}) \\
 &\asymp \Gamma_{j,j'}^{\circ}(\mathbf{E}_{j'} \mathbf{E}_{j'}^H \boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^H). \quad (70)
 \end{aligned}$$

Inserting (70) in (69) and using the fundamental Theorem 1 yields

$$\begin{aligned}
 A &\asymp P \sum_{j=1}^n \sum_{j'=1}^n \mu_j \mu_{j'} \frac{\Gamma_{j,j'}^{\circ}(\mathbf{E}_{j'} \mathbf{E}_{j'}^H \boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^H)}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \\
 &- \mu_j \mu_{j'} c_{0,k}^{(j')} \frac{\Gamma_{j,j'}^{\circ}(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^H)}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \frac{\text{tr}(\boldsymbol{\Theta}_k \mathbf{E}_{j'} \mathbf{E}_{j'}^H \mathbf{Q}_o^{(j')})}{M} \frac{1 + c_{1,k}^{(j')} m_k^{(j')}}{1 + m_k^{(j')}} \\
 &+ \mu_j \mu_{j'} (c_{2,k}^{(j')})^2 \frac{\Gamma_{j,j'}^{\circ}(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^H)}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \frac{\text{tr}(\boldsymbol{\Theta}_k \mathbf{E}_{j'} \mathbf{E}_{j'}^H \mathbf{Q}_o^{(j')})}{M} \frac{m_k^{(j')}}{1 + m_k^{(j')}} \\
 &\stackrel{(a)}{\asymp} P \sum_{j=1}^n \sum_{j'=1}^n \mu_j \mu_{j'} \frac{\Gamma_{j,j'}^{\circ}(\mathbf{E}_{j'} \mathbf{E}_{j'}^H \boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^H)}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \\
 &- \mu_j \mu_{j'} c_{0,k}^{(j')} \frac{\Gamma_{j,j'}^{\circ}(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^H)}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \frac{\text{tr}(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^H \mathbf{Q}_o^{(j')})}{M} \frac{1}{1 + m_k^{(j')}},
 \end{aligned}$$

where equality (a) follows from $c_{0,k}^{(j')} c_{1,k}^{(j')} = (c_{2,k}^{(j')})^2$. Proceed similarly for the remaining 4 terms and add term A, B, C, D

and \mathbf{E} together, we can get

$$\begin{aligned} \mathcal{I}_k &= A + B + C + D + E \\ &= P \sum_{j=1}^n \sum_{j'=1}^n \frac{\mu_j \mu_{j'} \Gamma_{j,j'}^{\circ}(\mathbf{E}_{j'} \mathbf{E}_j^H \mathbf{Q}_k \mathbf{E}_j \mathbf{E}_{j'}^H)}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \\ &\quad - 2P \sum_{j=1}^n \sum_{j'=1}^n \frac{\mu_j \mu_{j'} \Gamma_{j,j'}^{\circ}(\mathbf{Q}_k \mathbf{E}_j \mathbf{E}_j^H)}{\sqrt{\Gamma_{j,j}^{\circ}(\mathbf{I}_M) \Gamma_{j',j'}^{\circ}(\mathbf{I}_M)}} \frac{\text{tr}(\mathbf{Q}_k \mathbf{E}_{j'} \mathbf{E}_{j'}^H \mathbf{Q}_o^{(j')})}{M} \frac{c_{0,k}^{(j')}}{1 + m_k^{(j')}} \\ &\quad + P \sum_{j=1}^n \sum_{j'=1}^n \mu_j \mu_{j'} \frac{\text{tr}(\mathbf{Q}_k \mathbf{E}_j \mathbf{E}_j^H \mathbf{Q}_o^{(j)})}{M} \frac{\text{tr}(\mathbf{Q}_k \mathbf{E}_{j'} \mathbf{E}_{j'}^H \mathbf{Q}_o^{(j')})}{M} \Gamma_{j,j'}^{\circ}(\mathbf{Q}_k) \\ &\quad \cdot \frac{c_{0,k}^{(j)} c_{0,k}^{(j')} + \rho_k^{(j,j')} c_{2,k}^{(j)} c_{2,k}^{(j')}}{(1 + m_k^{(j)})(1 + m_k^{(j')})}. \end{aligned}$$

This concludes the proof.

APPENDIX D

PROOF OF POWER ALLOCATION THEOREM 3

Since $\boldsymbol{\mu}^{[t+1]}$ minimize the optimization problem in step 5 in Algorithm 1, we can have

$$\begin{aligned} &\sum_{k=1}^K \lambda_k^{[t]} \cdot \frac{\frac{1}{P} + (\boldsymbol{\mu}^{[t+1]})^T \mathbf{B}_k \boldsymbol{\mu}^{[t+1]}}{\frac{1}{P} + (\boldsymbol{\mu}^{[t+1]})^T (\mathbf{A}_k + \mathbf{B}_k) \boldsymbol{\mu}^{[t+1]}} \\ &\leq \sum_{k=1}^K \lambda_k^{[t]} \cdot \frac{\frac{1}{P} + (\boldsymbol{\mu}^{[t]})^T \mathbf{B}_k \boldsymbol{\mu}^{[t]}}{\frac{1}{P} + (\boldsymbol{\mu}^{[t]})^T (\mathbf{A}_k + \mathbf{B}_k) \boldsymbol{\mu}^{[t]}} \end{aligned}$$

Insert the expression for $\lambda_k^{[t]}$ in (58), use the notation for u_k defined in (57), the above expression simplifies as

$$\sum_{k=1}^K \frac{u_k(\boldsymbol{\mu}^{[t+1]})}{u_k(\boldsymbol{\mu}^{[t]})} \leq K.$$

According to AM-GM inequality

$$K \sqrt[K]{\frac{\prod_{k=1}^K u_k(\boldsymbol{\mu}^{[t+1]})}{\prod_{k=1}^K u_k(\boldsymbol{\mu}^{[t]})}} \leq \sum_{k=1}^K \frac{u_k(\boldsymbol{\mu}^{[t+1]})}{u_k(\boldsymbol{\mu}^{[t]})},$$

we can obtain

$$\prod_{k=1}^K u_k(\boldsymbol{\mu}^{[t+1]}) \leq \prod_{k=1}^K u_k(\boldsymbol{\mu}^{[t]}).$$

This shows that the value $\prod_{k=1}^K u_k(\boldsymbol{\mu})$ decreases during the iteration for updating $\boldsymbol{\mu}$. Since the physical meaning for $\prod_{k=1}^K u_k(\boldsymbol{\mu})$ is the sequence product of the MSE at each RX and therefore $\prod_{k=1}^K u_k(\boldsymbol{\mu}) > 0$. According to monotone convergence theorem, the iterative algorithm will produce a decreasing and lower bound series of MSE sequence product while updating $\boldsymbol{\mu}$, therefore the iterative procedure is surely to converge to a local optimum. This completes the proof.

ACKNOWLEDGMENT

The work of David Gesbert and Paul de Kerret has been supported by the ERC PERFUME under the European Unions Horizon 2020 research and innovation program (Agreement no. 670896).

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] A. Sanderovich, O. Somekh, H. V. Poor, and S. Shamai (Shitz), "Uplink macro diversity of limited backhaul cellular network," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3457–3478, 2009.
- [4] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai (Shitz), "Downlink multicell processing with limited-backhaul capacity," *EURASIP Journal on Advances in Signal Processing*, May 2009.
- [5] Y. Zhou, Y. Xu, W. Yu, and J. Chen, "On the optimal fronthaul compression and decoding strategies for uplink cloud radio access networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7402–7418, 2016.
- [6] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "Multihop backhaul compression for the uplink of cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3185–3199, 2016.
- [7] W. Yu, P. Patil, B. Dai, and Y. Zhou, "Cooperative beamforming and resource optimization in C-RAN," in *Cloud radio access networks: Principles, technologies, and applications*, T. Q. Quek, M. Peng, O. Simeone, and W. Yu, Eds. Cambridge University Press, 2017, ch. 4.
- [8] 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.
- [9] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.
- [10] 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.
- [11] P. Marsch and G. Fettweis, "On multicell cooperative transmission in backhaul-constrained cellular systems," *Annals of Telecommunications*, vol. 63, no. 5, 2008.
- [12] J. Zhao, T. Quek, and Z. Lei, "Coordinated Multipoint Transmission with Limited Backhaul Data Transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2762–2775, Jun. 2013.
- [13] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Drone small cells in the clouds: Design, deployment and performance analysis," in *Proc. Global Communications Conference (GLOBECOM)*, 2015.
- [14] 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.
- [15] 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.
- [16] P. de Kerret and D. Gesbert, "Network MIMO: Transmitters with no csi can still be very useful," in *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016, pp. 1466–1470.
- [17] 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.
- [18] A. Tulino and S. Verdú, *Random matrix theory and wireless communications*. Now Publisher Inc., 2004.
- [19] 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>
- [20] R. Couillet and M. Debbah, *Random matrix methods for wireless Communications*. Cambridge University Press, 2011.
- [21] T. Cover and A. Thomas, *Elements of information theory*. Wiley-Interscience, Jul. 2006.
- [22] 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, August 2011.
- [23] 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, October 2010.
- [24] J. Wang and D. P. Palomar, "Robust MMSE precoding in MIMO channels with pre-fixed receivers," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5802–5818, 2010.

- [25] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152–1159, Jul 1996.
- [26] 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.
- [27] 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.
- [28] P. de Kerret, D. Gesbert, and U. Salim, "Large system analysis of joint regularized Zero Forcing precoding with distributed CSIT," in *Proc. IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2015.
- [29] 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.
- [30] A. Wächter and L. T. Biegler, "Line search filter methods for nonlinear programming: Motivation and global convergence," *SIAM Journal on Optimization*, vol. 16, no. 1, pp. 1–31, 2005.
- [31] T. E. Bogale and L. Vandendorpe, "Weighted sum rate optimization for downlink multiuser MIMO coordinated base station systems: Centralized and distributed algorithms," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1876–1889, 2012.
- [32] P. Shen, W. Li, and X. Bai, "Maximizing for the sum of ratios of two convex functions over a convex set," *Computers & Operations Research*, vol. 40, no. 10, pp. 2301–2307, 2013.



David Gesbert David Gesbert (IEEE Fellow) is Professor and Head of the Communication Systems Department, EURECOM. He obtained the Ph.D degree from Ecole Nationale Supérieure des Télécommunications, France, in 1997. From 1997 to 1999 he has been with the Information Systems Laboratory, Stanford University. He was then a founding engineer of Iospan Wireless Inc, a Stanford spin off pioneering MIMO-OFDM (now Intel). Before joining EURECOM in 2004, he has been with the Department of Informatics, University of Oslo as an adjunct professor. D. Gesbert has published about 300 papers and 25 patents, some of them winning 2019 ICC Best Paper Award, 2015 IEEE Best Tutorial Paper Award (Communications Society), 2012 SPS Signal Processing Magazine Best Paper Award, 2004 IEEE Best Tutorial Paper Award (Communications Society), 2005 Young Author Best Paper Award for Signal Proc. Society journals, and paper awards at conferences 2011 IEEE SPAWC, 2004 ACM MSWiM. He has been a Technical Program Co-chair for ICC2017. He was named a Thomson-Reuters Highly Cited Researchers in Computer Science. Since 2015, he holds the ERC Advanced grant "PERFUME" on the topic of smart device Communications in future wireless networks. He is a Board member for the OpenAirInterface (OAI) Software Alliance. Since early 2019, he heads the Huawei-funded Chair on Advanced Wireless Systems Towards 6G Networks. He sits on the Advisory Board of HUAWAI European Research Institute. He held visiting professor positions in KTH (2014) and TU Munich (2016). Since 2017 he is also a visiting Academic Master within the Program 111 at the Beijing University of Posts and Telecommunications as well as as a member in the Joint BUPT-EURECOM Open5G Lab.



Qianrui Li Qianrui Li received the B.Sc degree in information engineering from Shanghai Jiaotong University, Shanghai, China, and the M.Eng. and Ph.D. degree in electronics and communications from Telecom ParisTech, France, in 2010, 2012, and 2016, respectively. He joined Mitsubishi Electric R&D Center Europe as a researcher since 2016. His research interests include signal processing, cooperative transmission and 5G wireless communication network and its vertical domain applications.



Nicolas Gresset Nicolas Gresset was born in France, in 1978. He has received the Ph.D. degree from Telecom Paristech, Paris, France, in collaboration with Mitsubishi Electric R&D Center Europe, Rennes, France, in 2004. He was a postdoctoral Researcher with Orange Labs, Issy les Moulineaux, France, in 2005. He was senior engineer in Wipro-Newlogic and delegate to the IEEE 802.11n standardization, in 2006. In 2007, he joined the Mitsubishi Electric R&D Center Europe, where he is now research manager of the Wireless Communication Systems group. The domains addressed by his research studies are the 5G cellular networks, wireless mesh networks, and train communication systems.



Paul de Kerret Paul de Kerret received the Engineering degree from the French Graduate School, IMT Atlantique, in 2010, the Diploma degree in electrical engineering and information technology from the Munich University of Technology through a double-degree program, and the Ph.D. degree from the French Graduate School, Telecom Paris, in 2013. From 2015 to 2019, he has been a Senior Researcher with EURECOM and he joined in September 2019 the Mantu Artificial Intelligence lab to lead the Research Group. He has been involved in several European collaborative projects on mobile communications, co-presented several tutorials at major IEEE international conferences, and authored over 30 papers in IEEE flagship conferences.

European collaborative projects on mobile communications, co-presented several tutorials at major IEEE international conferences, and authored over 30 papers in IEEE flagship conferences.