



HAL
open science

Asymptotically Achieving Centralized Rate on the Decentralized Network MISO Channel

Antonio Bazco-Nogueras, Paul de Kerret, David Gesbert, Nicolas Gresset

► **To cite this version:**

Antonio Bazco-Nogueras, Paul de Kerret, David Gesbert, Nicolas Gresset. Asymptotically Achieving Centralized Rate on the Decentralized Network MISO Channel. *IEEE Transactions on Information Theory*, 2022, 68 (1), pp.248-271. 10.1109/TIT.2021.3121752 . hal-04138853

HAL Id: hal-04138853

<https://hal.science/hal-04138853>

Submitted on 23 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Asymptotically Achieving Centralized Rate on the Decentralized Network MISO Channel

Antonio Bazco-Nogueras, *Member, IEEE*, Paul de Kerret, *Member, IEEE*,
David Gesbert, *Fellow, IEEE*, and Nicolas Gresset, *Senior Member, IEEE*

Abstract—In this paper, we analyze the high-SNR regime of the $M \times K$ Network MISO channel in which each transmitter has access to a different channel estimate, possibly with different precision. It has been recently shown that, for some regimes, this setting attains the same Degrees-of-Freedom as the ideal centralized setting with perfect Channel State Information (CSI) sharing, in which all the transmitters are endowed with the best estimate available at any transmitter. This result is restricted by the limitations of the Degrees-of-Freedom metric, as it only provides information about the slope of growth of the capacity as a function of the SNR, without any insight about the possible performance at a given SNR. In order to overcome this limitation, we analyze the affine approximation of the rate on the high-SNR regime for this decentralized Network MISO setting for the antenna configurations in which it achieves the Degrees-of-Freedom of the centralized setting. We show that, for a regime of antenna configurations, it is possible to asymptotically attain the same achievable rate as in the ideal centralized scenario. Consequently, it is possible to achieve the beamforming gain of the ideal perfect-CSI-sharing setting even if only a subset of transmitters is endowed with precise CSI, which can be exploited in scenarios such as distributed massive MIMO where the number of transmit antennas is much bigger than the number of served users. This outcome is a consequence of the synergistic compromise between CSI precision at the transmitters and consistency between the locally-computed precoders, which is an inherent trade-off of decentralized settings that does not exist in the centralized CSI configuration. We propose a precoding scheme achieving the previous result, which is built on an uneven structure in which some transmitters reduce the precision of their own precoding vector for the sake of using transmission parameters that can be more easily predicted by the other transmitters.

Index Terms—Network MIMO, Cooperative transmission, Distributed Broadcast Channel, decentralized zero-forcing.

I. INTRODUCTION

A. Cooperative Transmission

Multi-user cooperative networks and the extend of their theoretical capabilities have been thoroughly analyzed in the literature [1]–[4]. Initially, it was shown that cooperation can

This work was in part supported by the European Research Council under the European Union’s Horizon 2020 research and innovation program (Agreement no. 670896). (Corresponding author: Antonio Bazco-Nogueras).

A. Bazco-Nogueras was with the Mitsubishi Electric R&D Centre Europe, Rennes, France, and also with the Communication Systems Department, EURECOM, Sophia-Antipolis, France. He is now with the IMDEA Networks Institute, Madrid, Spain (e-mail:antonio.bazco@imdea.org).

P. de Kerret was with the Communication Systems Department, EURECOM, Sophia-Antipolis, France; he is now with Greenly.

D. Gesbert is with the Communication Systems Department, EURECOM, Sophia-Antipolis, France (e-mail: gesbert@eurecom.fr).

N. Gresset is with the Mitsubishi Electric R&D Centre Europe, Rennes, France (e-mail: n.gresset@fr.mercede.mee.com).

bring multiplicative gains under ideal assumptions on the channel knowledge available at the communicating nodes. As a matter of example, it is known that, under the assumption of perfect Channel State Information (CSI), the setting in which M transmitters (TXs) jointly serve K single-antenna users (the so-called Network MISO channel) achieves a rate that scales as $\min(N_T, K)$ times the rate of the single-antenna point-to-point channel [5], where N_T is the total number of transmit antennas. Conversely, the $K \times K$ interfering channel can attain a multiplexing gain of $K/2$ [6]. This perfect CSI scenario has been profoundly studied [5], [7]–[9]. Unfortunately, the assumption of perfect information is not practical in most of the current wireless networks. Motivated by the infeasibility of the previous assumption, the community has investigated settings in which the information available at the communicating nodes does not meet the perfect CSI assumption, such as scenarios where the information available is imperfect [10]–[14], partial [15], [16], or delayed [17]–[20].

Even though the aforementioned works considered an imperfect acquisition or estimation of the CSI, all the cooperating nodes are assumed to *perfectly* share the same *imperfect* information. Yet, current and upcoming wireless networks characteristics make this assumption of *perfect sharing* also impractical for many applications. This is due to, for example, the proliferation of heterogeneous networks for which some of the nodes have a wireless, fluctuating, or limited backhaul [21], [22], hierarchical distributed networks (with remote radio-heads of different capabilities), or Ultra-Reliable Low-Latency Communication (URLLC) applications [23], [24], in which the perfect sharing of the information would result in an intolerable delay. Settings in which simple devices with low capabilities aim to communicate in a dense environment, as in IoT applications, also fall into the use cases in which the sharing of channel information is indispensable but challenging. This evolution of different use cases boosts the interest of distributed information settings, in which the information available at the communicating nodes is not only imperfect but different from one node to another. This type of settings can be formalized as the so-called Team Decision problems [25], in which different agents aiming for the same goal attempt to cooperate in the absence of perfect communication among them.

B. Distributed CSIT Setting

There exists a great number of different distributed settings [26]–[31]. In particular, we focus on the scenario where M TXs jointly serve K different single-antenna users (RXs), which

is often referred to as the Network MISO setting, under the assumption of distributed CSI at the TXs (CSIT). In such scenario, every TX has access to the information symbols of every RX, but it does not share the same CSI with the other TXs. This setting arises in situations in which the data can be buffered or cached, but the CSI needs to be available with very small delay; for example, for the transmission of high-popularity content that can be cached at the network edge before the transmission occurs [31]–[33], or in high mobility scenarios and IoT (or V2X) networks with fast varying channels but low data rate [34], [35].

The Network MISO with distributed CSIT has been analyzed in recent works. Initially, it was shown in [36] that the 2×2 single-antenna scenario in which one TX has better knowledge of the full channel matrix than the other TX achieves the same Degrees-of-Freedom (DoF) as the ideal centralized case in which both TXs are endowed with the best CSI. The DoF metric, which will be presented in the following section, is defined as the pre-logarithmic factor of the capacity as function of the SNR [2], [6], [9], and it is also known as multiplexing gain. The outcome of [36] derives from a precoding with a master-slave structure, named Active-Passive Zero-Forcing (AP-ZF), in which the TX whose CSI is less precise transmits with a fixed precoder. Although this result could depend on the asymmetric structure of the setting, it has been extended to more general settings. Indeed, it was shown in [37] that the Generalized DoF¹ of the centralized setting with perfect CSIT sharing are attained in the distributed setting, no matter which TX has the best estimation of every single channel coefficient. By way of explanation, the pre-logarithmic factor of the centralized 2×2 Network MISO is preserved as long as the estimate of a certain link is available at one of the TXs. This comprises for example cases in which TXs have only local CSI or in which each TX knows better the channel towards a certain RX. The DoF analysis has been also applied to the $K \times K$ scenario, in which two main insights can be outlined: First, the optimal DoF of the centralized setting with perfect CSIT sharing is also achieved for some regimes of the $K \times K$ scenario [38]. Second, the quantization of the information available at a certain TX can be beneficial in distributed settings, inasmuch as it helps to transform the setting into a hierarchical configuration in which the structure of the CSI allocation can be used to increase the DoF [39].

These results provide some understanding on the resilience of cooperating settings under information mismatches between different nodes. However, the DoF metric is a limited metric because it only provides information about the pre-log factor, not offering any information about the achievable rate at any given SNR. For that reason, it is necessary to take a step beyond and analyze the affine approximation of the rate at the high-SNR regime.

C. Linear Approximation of Rate at High-SNR

Finding the fundamental limits of communication in distributed settings is a challenging problem. Indeed, these limits

¹Generalized DoF refers to the DoF analysis under the assumption that the difference of channel strengths between the links does not vanish at high SNR. See [9] for more details.

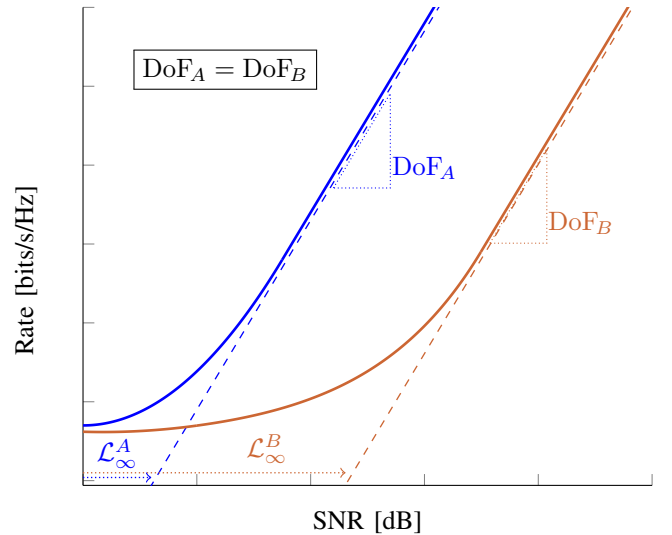


Fig. 1: Qualitative illustration of the affine approximation of two different settings with the same DoF (slope) but different rate offsets $\mathcal{R}_\infty = \text{DoF} \mathcal{L}_\infty$, and hence different achieved rates.

have remained open even for several important centralized cases. Nevertheless, it is possible to find significant insights on those settings through rate approximations, which help us to move towards a better understanding of the behavior of complex wireless networks.

A very useful metric that has been applied in the literature is the affine approximation of the rate at high SNR, introduced in [40]. Following this linear approximation, the rate can be written as

$$R = \text{DoF} \log_2(P) - \mathcal{R}_\infty + o(1), \quad (1)$$

where P denotes the SNR, DoF is the pre-logarithmic factor (or Degrees-of-Freedom), and \mathcal{R}_∞ is the *rate offset* (or vertical offset). The approximation in (1) can also be written in terms of the *power offset* (horizontal offset) \mathcal{L}_∞ , where $\mathcal{R}_\infty = \text{DoF} \mathcal{L}_\infty$. An illustrative representation is shown in Fig. 1. The term \mathcal{L}_∞ represents the zero-order term with respect to a reference setting with the same slope but whose affine approximation intersects the origin. These terms are defined as

$$\text{DoF} \triangleq \lim_{P \rightarrow \infty} \frac{R(P)}{\log_2(P)} \quad (2)$$

$$\mathcal{R}_\infty \triangleq \lim_{P \rightarrow \infty} \text{DoF} \log_2(P) - R(P), \quad (3)$$

where $R(P)$ represents the rate as function of the SNR P . This measure has already proved instrumental in several findings. In [3], Lozano *et al.* analyzed the multiple-antenna point-to-point scenario, revealing that some system features that do not impact the DoF (as antenna correlation, fading, etc.) do considerably impact the zero-order term, affecting the performance of the system at any possible SNR. In addition to exposing some limitations of the DoF metric, [3] also revealed that the affine expansion offers appreciably tight approximations also at medium-to-low SNR. This characterization has been also established for the Broadcast Channel (BC) with perfect CSIT using Dirty-Paper Coding and linear precoding [41], and

for the BC with imperfect CSIT [42]. In [42], the BC setting with quantized feedback was studied under the assumption of Zero-Forcing (ZF) schemes, showing that the CSIT error variance must be proportional to $\text{SNR}^{-\alpha}$ in order to attain a DoF per user of $\text{DoF}_{\text{RX } i} = \alpha$. Furthermore, having a CSIT error variance scaling with $\text{SNR}^{-\alpha}$ was shown to be equivalent to obtaining a quantized feedback of $\alpha \log_2(\text{SNR})$ bits from the RX, which could be attained if the feedback resources scale proportionally to $\log_2(\text{SNR})$.

The aforementioned works are yet focused on the centralized scenario. For the best of our knowledge, this affine approximation was analyzed in the Distributed CSIT setting for first time in [43]. In that work, it was shown that the 2×2 Network MISO with Distributed CSIT and quantized feedback achieves the same rate offset as the ideal centralized scenario of [42], in which the best CSIT is available at both TXs. Therefore, both settings asymptotically achieve the same rate. This result strengthens the previous results on DoF because it shows that there is no fundamental gap between both centralized and distributed CSIT settings, which could not be inferred from DoF analysis. Besides that, it was also shown that the loss of performance at practical SNR values can be dramatically reduced with the correct design of the transmission schemes.

D. Main Contributions

Motivated by the result of [43] for the 2×2 setting, we aim to characterize the high-SNR achievable throughput of ZF precoding techniques for the $M \times K$ Network MISO setting with Distributed CSIT, and we compare it with the ideal centralized CSIT setting. This comparison is only meaningful for the antenna configurations that allow the distributed setting to achieve the same DoF as its centralized counterpart, since, otherwise, the rate gap between both settings will become unboundedly big as the SNR increases. Therefore, we restrict ourselves to such regime. Our main contributions write as follows:

- We obtain the affine expansion of the rate achieved with ZF precoding on the $M \times K$ Network MISO setting with Distributed CSIT for the antenna configurations in which the centralized DoF can be achieved in the Distributed CSIT setting. We prove that it is possible to asymptotically reach the same rate as in the ideal centralized setting in which the best estimate available in the network is shared by all the TXs.

This result is especially relevant when the number of transmit antennas (N_t) is considerably greater than the number of RXs (K), such as in distributed massive MIMO settings [44], since we can achieve the beamforming gain of the centralized CSIT setting even if we only have precise CSIT at K transmit antennas.

- We present a discussion on a dilemma that is intrinsic to distributed settings and does not exist in the conventional centralized case: the interplay between local precision and global consistency. In particular, we present an achievable scheme that achieves the previous result by means of capitalizing on the compromise between precoder precision and consistency among transmitters. We demonstrate

that decreasing the precision of the precoding at certain transmitters improves the average performance as it helps to enhance the predictability of the transmission from the other TXs.

In addition to that, the techniques and approaches employed for the design of the transmission scheme are believed to be worthwhile by themselves for general decentralized problems, since they deal with the interplay between local precision and global consistency, which is an inherent aspect of decentralized and Team Decision problems.

Notations: \mathbb{N}_N stands for $\mathbb{N}_N \triangleq \{1, 2, \dots, N\}$. We follow the asymptotic notation presented in [45], based on the prevalent Bachmann–Landau notation [46]. In consequence, $f(x) = o(g(x))$ implies that $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0$, $f(x) = \mathcal{O}(g(x))$ implies that $\limsup_{x \rightarrow \infty} \frac{|f(x)|}{g(x)} < \infty$, and $f(x) = \Theta(g(x))$ implies that $\lim_{x \rightarrow \infty} \frac{|f(x)|}{g(x)} = c$, $0 < c < \infty$. For any expected value E and event X , $E|_X$ denotes the conditional expectation given X . $\Pr(X)$ stands for the probability of the event X and X^c represents the complementary event to X . $\mathbf{A}_{i,k}$ or $(\mathbf{A})_{i,k}$ represent the element of the matrix \mathbf{A} located in the i -th row and the k -th column. \mathbf{I}_n writes for the identity matrix of size $n \times n$. $\mathbf{1}_{M \times N}$ (resp. $\mathbf{0}_{M \times N}$) represents the all-ones (resp. all-zeros) matrix of size $M \times N$. $\|\mathbf{A}\|$ denotes the Frobenius norm of the matrix \mathbf{A} .

II. PROBLEM FORMULATION

A. Transmission and System Model

We consider the Network MISO setting in which M multi-antenna TXs jointly serve K single-antenna RXs. We suppose that TX j has N_j antennas, and we denote the total number of transmit antennas as $N_T = \sum_{j=1}^M N_j$. The received signal is given by

$$\mathbf{y} \triangleq \sqrt{P} \mathbf{H} \mathbf{W} \mathbf{s} + \mathbf{n}, \quad (4)$$

where P is the transmit power, $\mathbf{y} \triangleq [y_1, \dots, y_K]^T$ is the received signal vector, and y_i is the received signal at RX i . \mathbf{n} stands for the Additive White Gaussian Noise (AWGN) distributed as $\mathcal{N}_{\mathbb{C}}(0, 1)$, where $\mathcal{N}_{\mathbb{C}}(0, \Gamma)$ stands for the circularly-symmetric complex normal distribution with covariance matrix (or variance) Γ . The vector $\mathbf{s} \triangleq [s_1, \dots, s_K]^T$ is the vector of independent and identically distributed (i.i.d.) information symbols, where s_i is the message to RX i . The vector of information symbols satisfies $\mathbb{E}[\mathbf{s} \mathbf{s}^H] = \mathbf{I}$. The channel matrix is given by

$$\mathbf{H} \triangleq \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_K \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{h}_{1,1} & \dots & \mathbf{h}_{1,M} \\ \vdots & \ddots & \vdots \\ \mathbf{h}_{K,1} & \dots & \mathbf{h}_{K,M} \end{bmatrix} \in \mathbb{C}^{K \times N_T}. \quad (5)$$

Hence, $\mathbf{h}_i \in \mathbb{C}^{1 \times N_T}$ denotes the global channel vector towards RX i and $\mathbf{h}_{i,j} \in \mathbb{C}^{1 \times N_j}$ is the channel vector from TX j to RX i . Note that we have defined the row vectors as \mathbf{h}_i and $\mathbf{h}_{i,j}$ in place of the usual Hermitian notation \mathbf{h}_i^H and $\mathbf{h}_{i,j}^H$. This is done to ease the notation for the remainder of the document. The channel coefficients are assumed to be i.i.d. as

$\mathcal{N}_{\mathbb{C}}(0, 1)$ such that all the channel sub-matrices are full rank with probability one. The precoding matrix is given by

$$\mathbf{W} \triangleq \begin{bmatrix} \mathbf{T}_1 \\ \vdots \\ \mathbf{T}_M \end{bmatrix} \triangleq \mu [\mathbf{w}_1 \quad \dots \quad \mathbf{w}_K] \quad (6)$$

$$\triangleq \mu \begin{bmatrix} \mathbf{w}_{1,1} & \dots & \mathbf{w}_{K,1} \\ \vdots & \ddots & \vdots \\ \mathbf{w}_{1,M} & \dots & \mathbf{w}_{K,M} \end{bmatrix} \in \mathbb{C}^{N_T \times K}. \quad (7)$$

Hence, $\mathbf{T}_j \in \mathbb{C}^{N_j \times K}$ represents the precoding matrix applied at TX j , $\mu \mathbf{w}_i \in \mathbb{C}^{N_T \times 1}$ denotes the global precoding vector for the information symbol of RX i (s_i), and $\mu \mathbf{w}_{i,j} \in \mathbb{C}^{N_j \times 1}$ is the precoding vector applied at TX j for s_i . We further define $\mathbf{w}_{i,j,n}$ as the coefficient at the n -th antenna of TX j . The parameter $0 < \mu \leq 1$ is a power correction value that will be detailed afterwards. We define $\mathbf{T}_{j,n} \in \mathbb{C}^{1 \times K}$ as the precoding vector applied at the n -th antenna of TX j for the vector \mathbf{s} , with $n \in \mathbb{N}_{N_j}$. We assume that the precoder has a per-antenna instantaneous unit-norm constraint, such that

$$\|\mathbf{T}_{j,n}\| \leq 1. \quad (8)$$

The results presented here also hold under the assumption of per-TX instantaneous constraint ($\|\mathbf{T}_j\| \leq 1$). Note that, even if we set $\|\mathbf{T}_j\| = 1$, the transmit power varies over the time as the power of the information symbols s_i varies. For sake of concision, we refer hereinafter to the unit-norm constraint of (8) as *instantaneous power constraint*, although strictly speaking it is an *instantaneous power constraint on the precoding vector averaged over the information symbols*.

This is done in opposition to the average power constraint on the precoder ($\mathbb{E}[\|\mathbf{T}_j\|^2] \leq 1$) which has been assumed in other works [37], [38]. This average power constraint, which results in a power normalization factor of $1/\sqrt{\mathbb{E}[\|\mathbf{w}_i\|^2]}$ for each \mathbf{w}_i , it is known to be less harmful for distributed CSIT settings. This fact follows because, under the instantaneous power constraint, the precoding must be carefully designed to avoid exceeding the power budget at every time, and this power normalization must be locally done at each TX without an exact knowledge of the actions of the other TXs. On the other hand, under the average power constraint the normalization is based on statistics and it is constant.

Thus, we consider the instantaneous power constraint because it is the most challenging case for a distributed setting where each TX must independently compute and agree on the applied power normalization. As it will be shown in the following, the constraint on the instantaneous power is one of the key aspects that could diminish the performance in the distributed setting.

B. Network MISO Setting with Distributed CSIT

The Distributed CSIT (D-CSIT) model is characterized by the consideration that each TX is endowed with a possibly different estimate. The key particularity of this setting is that, for any channel coefficient, there exist as many estimates as TXs, each of them *locally available at a single TX*. By way of example, we can think of a practical scenario in which there

are some TXs with more precise knowledge for some channel coefficients and some other TXs with more precise information about other part of the channel matrix, while some sort of coarse information could be shared between them.

Although the assumption of Distributed CSIT may seem contradictory with the assumption of perfect sharing of user data symbols, which is inherent to the Network MISO setting, both aspects co-exist in many scenarios of interest. More specifically, this model is inspired by the different timescales of latency that information data and CSI may experience in a range of emerging applications. Indeed, CSI sharing is constrained by the channel coherence time (which can be very short in high-mobility scenarios) and by the possibly latency-limited backhaul connection; on the other hand, many data applications have delivery time restrictions which are orders of magnitude weaker, such that the data can be pre-fetched or cached at the TXs and ready to be synchronously transmitted. We refer to [38] for a detailed discussion and motivation on the joint transmission with distributed CSIT and the practical scenarios in which it arises.

In this work, we assume that a limited cooperation between TXs occurred before the transmission phase, leading to a certain CSIT precision configuration. Hence, we assume hereinafter that the average CSIT precision at each TX remains fixed for a period of time that is long enough to be considered constant in our analysis. The problem of studying the best strategy of CSIT sharing with constrained links and/or delays is a very interesting research problem; yet, it is also a complex problem which would require another work on its own, and thus we do not discuss the exact CSIT acquisition and sharing mechanism.

We focus on a particular CSIT configuration where each TX knows the whole channel matrix with the same precision. This setting is denoted as the *Sorted CSIT setting*, as the TXs can be sorted by level of average precision. Let us denote the estimate of the channel matrix \mathbf{H} available at TX j as $\hat{\mathbf{H}}^{(j)} \in \mathbb{C}^{K \times N_T}$. Then, we model the D-CSIT configuration such that the estimate $\hat{\mathbf{H}}^{(j)}$ is given by

$$\hat{\mathbf{H}}^{(j)} \triangleq \sqrt{1 - Z^{(j)}} \mathbf{H} + \sqrt{Z^{(j)}} \mathbf{\Delta}^{(j)}, \quad (9)$$

where $\mathbf{\Delta}^{(j)} \triangleq [\boldsymbol{\delta}_1^{(j)}, \dots, \boldsymbol{\delta}_K^{(j)}]^T$ is a random matrix that contains the additive estimation noise and whose elements are i.i.d. as $\mathcal{N}_{\mathbb{C}}(0, 1)$. The matrix $\mathbf{\Delta}^{(j)}$ is independent of \mathbf{H} . $Z^{(j)}$ is a deterministic value that represents the variance scaling of the estimation noise with respect to the SNR. Importantly, $Z^{(j)}$ characterizes the average precision of the estimate at TX j , which is assumed to remain constant.

From (9), we can see that the variance scaling $Z^{(j)}$ at a given TX is the same for all the channel coefficients. This model encloses e.g. a scenario in which a main multi- or massive-antenna base station serves a set of users with the help of some single or multi-antenna remote radio-head or simple TXs, as depicted in Fig. 2.

We would like to remark that this configuration is the only one known to achieve the same DoF as the centralized setting where the most precise CSIT is shared. The analysis can be however extended to the case in which the CSIT precision order is preserved but the channel towards each user is known

with different precision at a particular TX. As we will see later, the precoding vectors are independently computed for each RX, and only the normalization parameter depends on all the user's vectors.² Because of this last point, we can restrict ourselves to the case in which the precision of each TX is characterized by a single parameter $Z^{(j)}$ for the sake of readability and concision, as the extension to the setting where each RX channel vector is obtained with different precision at the same TX (and the order of TXs is preserved) would follow in a similar way.

Furthermore, this Sorted CSIT setting can model other CSIT configurations. For example, a setting where each TX obtains precise CSIT from a subset of the network (e.g., the full channel vector of some users that are associated with that specific TX, or the direct channel from all users to that specific TX), which is the most usual case in practice. Then, we could apply *perfect* CSIT sharing only unidirectionally towards the main TX, whereas a coarse global CSIT estimate could be broadcast back to all the TXs from this main TX. This configuration could be enforced by tight delay constraints, and it has been already considered in practical analysis of cell-free Massive MIMO and *radio stripes* systems [47]–[49]. Then, the only difference w.r.t. our Sorted CSIT setting would be that each TX has also the best estimate for the locally acquired CSIT, but this would provide no *high-SNR* gain because the performance will be limited by the other TXs.

Remark 1. *In contrast to the prior work in [43] for the 2×2 setting, in which a distributed adaptation of the Random Vector Quantization (RVQ) feedback model of [42] was supposed, we assume in this work an additive Gaussian model for the estimation noise. This modification allows us to enlarge the contribution and verify that the previous result is not dependent on the feedback model. Indeed, the asymptotic results are expected to hold for a broad family of estimation models, since the key parameter that characterizes the asymptotic analysis is the scaling of the variance of the CSIT error $Z^{(j)}$, which is assumed to be constant, and not in the unit-variance random matrix $\Delta^{(j)}$.*

It is known that, for the centralized CSIT case in which all the transmit antennas share the same CSI, the variance of the estimation noise should scale as $P^{-\alpha}$ (provided that the channel estimate has unit variance) in order to obtain a multiplexing gain of α [42], [50]. As our focus is on the high-SNR regime, we consider a high-SNR modeling in which the estimation error scales such that

$$Z^{(j)} = \bar{P}^{-\alpha^{(j)}}, \quad (10)$$

where $\bar{P} \triangleq \sqrt{P}$ and $0 \leq \alpha^{(j)} \leq 1$. $\alpha^{(j)}$ is the *precision scaling* parameter that measures the average quality of estimation of the channel matrix at TX j . Hence, we can order the TXs w.l.o.g. such that

$$1 \geq \alpha^{(1)} > \alpha^{(2)} \geq \dots \geq \alpha^{(M)} \geq 0, \quad (11)$$

which implies that TX 1 is the *best-informed* TX; in other words, the TX whose CSIT has the highest average precision.

²Moreover, it was shown in [43] that this normalization parameter can be calculated with a less restricting precision without affecting the asymptotic performance.

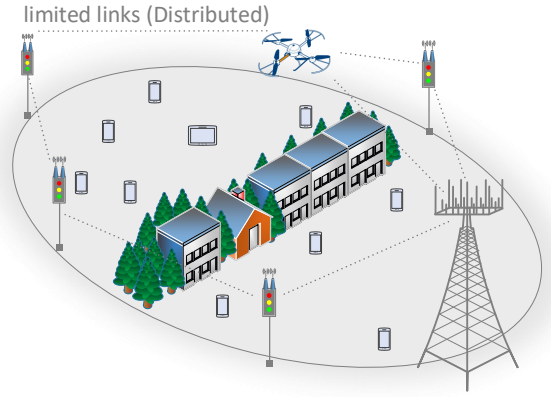


Fig. 2: Master Base Station with remote radio-heads. The Base Station obtains an estimate of the whole channel matrix, then it transmits noisy or compressed CSI to the auxiliary TXs.

The reason for which the inequality $\alpha^{(1)} > \alpha^{(2)}$ is strict will be explained in Section IV. We define the set of precision parameters of the Distributed CSIT setting as

$$\alpha_M = \{\alpha^{(j)}\}_{j \in \mathbb{N}_M}. \quad (12)$$

For further use, we define the estimate at TX j for the channel of RX i as

$$\hat{\mathbf{h}}_i^{(j)} \triangleq \check{z}^{(j)} \mathbf{h}_i + z^{(j)} \boldsymbol{\delta}_i^{(j)}, \quad (13)$$

where $z^{(j)} \triangleq \bar{P}^{-\alpha^{(j)}}$, $\check{z}^{(j)} \triangleq \sqrt{1 - (z^{(j)})^2}$, and $\hat{\mathbf{h}}_i^{(j)}$, $\boldsymbol{\delta}_i^{(j)}$ are the i -th rows of the matrices $\hat{\mathbf{H}}^{(j)}$, $\boldsymbol{\Delta}^{(j)}$, respectively. As stated before, the precision parameters α_M are assumed to be long-term coefficients that vary slowly. Based on that, it is assumed that every TX knows the full set α_M , as it only requires a sharing of few bits over a long period of time.

C. Ideal Centralized Setting

Finding purely distributed upper-bounds is a challenging subject that remains open, although some first results have been shown in [51]. However, any decentralized scenario with distributed estimates has an *ideal* centralized counterpart in which a genie provides the best estimate of each parameter to every node. Based on that, we define an ideal centralized scenario as follows.

Definition 1 (Ideal Centralized Setting). *Consider the distributed setting introduced in Section II-B. The Ideal Centralized Setting is defined as the setting in which all the TXs are endowed with the estimate of best average precision for every channel coefficient.*

Hereinafter, we denote the centralized channel estimate as $\hat{\mathbf{H}}$ and the estimate for the channel vector of RX i as $\hat{\mathbf{h}}_i$. We further denote the CSIT precision for the ideal centralized case as α^* . Note that in the sorted setting, where $\alpha^{(1)} > \dots > \alpha^{(M)}$, the ideal centralized setting is equivalent to a MISO Broadcast Channel with N_T transmit antennas, CSIT $\hat{\mathbf{H}}$ equal to $\hat{\mathbf{H}}^{(1)}$, and $\alpha^* = \alpha^{(1)}$.

We will compare the rate achieved in the distributed scenario described in the previous section with the respective ideal

centralized MISO BC counterpart. This provides us with a benchmark for the performance of ZF schemes in the Distributed CSIT setting. In this way, we are able to analyze which is the impact of having distributed information or, in other words, *the cost of not sharing the best CSI*.

Remark 2. *It is important to observe that the ideal centralized scenario is such that every TX owns the best estimate among the available estimates at any TX, instead of its own estimate which by definition would have less precision. This is in opposition to another genie-aided scenario, also assumed in the literature [37], [38], in which each TX shares its CSIT with any other TX, such that every TX owns the set of M estimates of the M TXs. The former model, here assumed, permits to compare the distributed scenario with the centralized counterpart. The later, although it was shown in [38] that does not attain a greater DoF, would benefit from the fact that the knowledge of M estimates allows to reduce the noise variance by a factor proportional to M .* \square

D. Figure of Merit

Our figure of merit is the ergodic sum rate, i.e., the expected sum rate over the joint process of fading realizations and estimates $\{\mathbf{H}, \hat{\mathbf{H}}^{(1)}, \dots, \hat{\mathbf{H}}^{(M)}\}$ [52]. For the sake of readability, we omit hereinafter the reference to the joint process $\{\mathbf{H}, \hat{\mathbf{H}}^{(1)}, \dots, \hat{\mathbf{H}}^{(M)}\}$ when denoting the expectation $\mathbb{E}[\cdot]$.

Let us define the expected rate of RX i as $R_i \triangleq \mathbb{E}[r_i]$, where r_i is the instantaneous rate of RX i . In our K -user setting, r_i writes as

$$r_i \triangleq \log_2 \left(1 + \frac{P|\mathbf{h}_i^H \mathbf{t}_i|^2}{1 + \sum_{j \neq i} P|\mathbf{h}_i^H \mathbf{t}_j|^2} \right), \quad (14)$$

where \mathbf{t}_i denotes the precoder vector for the symbols of RX i . Then, the expected sum rate is given by $R \triangleq \sum_{i=1}^K R_i$. Specifically, we study the linear approximation presented in (1) for the rate of the Distributed CSIT setting. Let us denote the rate achieved in a Distributed CSIT setting characterized by α_M as $R(\alpha_M)$. Hence, we aim to find the values DoF_d , \mathcal{R}_∞^d , such that

$$R(\alpha_M) = \text{DoF}_d \log_2(P) - \mathcal{R}_\infty^d + o(1). \quad (15)$$

III. ZERO-FORCING PRECODING

A. Centralized Zero-Forcing Schemes (with ideal CSIT sharing)

We restrict in this work to a general type of ZF precoders that we rigorously characterize in the following. First, in order to distinguish when the precoding vectors refer to the ideal centralized CSIT setting or the D-CSIT one, we denote the centralized precoding coefficients as $\mathbf{v}_{i,k}$; note that the counterpart vector for the D-CSIT setting in (7) is denoted by $\mathbf{w}_{i,k}$. In addition, \mathbf{V} and \mathbf{v}_i are defined as the centralized counterpart of \mathbf{W} and \mathbf{w}_i , respectively. Hence, the vectors \mathbf{v}_i are computed from any ZF precoding algorithm satisfying

$$1) \hat{\mathbf{h}}_i \mathbf{v}_\ell = 0, \quad \forall \ell \neq i \quad (\text{Zero-Forcing condition}) \quad (\text{ZF1})$$

$$2) \mathbb{E}[\|\mathbf{v}_{i,j,n}\|^{-1}] = \Theta(1) \quad (\text{ZF2})$$

$$3) f_{\|\mathbf{v}_i\|} \leq f_{\|\mathbf{v}_i\|}^{\max} < \infty \quad (\text{ZF3})$$

Note that (ZF1) is nothing but the condition that defines ZF schemes, (ZF2) implies that the probability of precoding with a vanishing power is negligible, and (ZF3) that the precoding vector has a bounded probability density function (pdf), i.e., that it is neither predetermined nor constant. This last property follows from the assumption that \mathbf{h}_i and $\hat{\mathbf{h}}_i$ are drawn from Gaussian distributions and that the ZF precoding is a continuous map from estimates $\hat{\mathbf{h}}_i$ to vectors \mathbf{v}_ℓ . Hereinafter, we assume that the centralized precoding scheme satisfies (ZF1), (ZF2), (ZF3). Furthermore, we assume that the precoding vectors and matrices can be expressed as a combination of summations, products, and generalized inverses³ of the channel estimates. As an example, we can use the typical choice of the projection of the matched filters onto the null spaces of the interfered users, i.e.,

$$\mathbf{v}_i = \frac{\mathbf{P}_{\hat{\mathbf{h}}_i^\perp} \hat{\mathbf{h}}_i^H}{\|\mathbf{P}_{\hat{\mathbf{h}}_i^\perp} \hat{\mathbf{h}}_i^H\|} \quad (16)$$

where $\mathbf{P}_{\hat{\mathbf{h}}_i^\perp}$ is defined as

$$\mathbf{P}_{\hat{\mathbf{h}}_i^\perp} \triangleq \left(\mathbf{I}_{N_T} - \hat{\mathbf{H}}_i^H (\hat{\mathbf{H}}_i \hat{\mathbf{H}}_i^H)^{-1} \hat{\mathbf{H}}_i \right), \quad (17)$$

and where the matrix $\hat{\mathbf{H}}_i$ stands for the global channel matrix with the i -th row removed. Note that the inversion in (17) can be regularized in order to avoid degenerate cases and increase the performance at low SNR. However, as conventional regularized schemes converge to their non-regularized counterpart at high SNR, we omit any reference to regularized inverses. We further model the precoding scheme as a function of the CSIT, such that \mathcal{V} denotes the function applied to the channel estimate:

$$\mathcal{V} : \mathbb{C}^{K \times N_T} \rightarrow \mathbb{C}^{N_T \times K} \quad \text{and} \quad \mathbf{V} = \mathcal{V}(\hat{\mathbf{H}}). \quad (18)$$

B. Zero-Forcing on Distributed CSIT Settings

When we consider a distributed CSIT setting, we face a new challenge that does not exist in the centralized CSIT counterpart. This challenge is whether the TXs should use their CSIT even if they are among the TXs with worst CSIT precision, and to which extend they should use it. Hereinafter, we will use the term “naive precoder” at a certain transmitter to refer to any precoding strategy that considers that such transmitter (naively) assumes that all the other TXs have obtained exactly its same CSIT, as if the TXs were in a centralized CSIT setting. Let us mention two simple and intuitive ZF approaches that can be applied in the distributed CSIT setting and their shortcomings.

- 1) *Naive ZF precoder:* In this strategy, each TX assumes that all the TXs have obtained its same (imperfect) CSIT, i.e., the TXs consider that they are in a centralized CSIT setting. This approach could prove efficient at low-to-intermediate SNR for the considered CSIT error model, because the exponential relation of the CSIT error with P (cf. (10)) implies that the noise variance at most transmitters will not be significantly different at such SNR ranges. Yet, this strategy fails at high SNR. Indeed,

³The *generalized inverse* of an $M \times N$ matrix \mathbf{X} is any matrix \mathbf{X}^- satisfying that $\mathbf{X}\mathbf{X}^-\mathbf{X} = \mathbf{X}$ [53], [54]. For any full-rank matrix \mathbf{X} , the generalized inverse is any matrix \mathbf{X}^- that satisfies the fact that $\mathbf{X}^-\mathbf{X} = \mathbf{I}$.

it is known that the high-SNR performance of centralized ZF schemes collapses when they are naively applied on Distributed CSIT settings [36], [38]. The main reason is that the interference cancellation achieved through (ZF1) is proportional to the worst CSIT precision among the TXs, $\alpha^{(M)}$ in the sorted case.

Thus, the question is how to prevent the least precise TXs from harming the transmission. Intuitively, the TXs should not act naively, but rather they should take into account the action of the other TXs.

- 2) *Active-Passive ZF (AP-ZF)*: The second intuitive idea is that the TXs whose low CSIT precision is harming the transmission should not exploit their instantaneous CSI for precoding. Instead, they could transmit with a fixed or known precoder based on statistical information. This solution, coined *Active-Passive ZF (AP-ZF)*, achieves the centralized DoF under the less restrictive average power constraint $\mathbb{E}[\|\mathbf{T}_j\|^2] \leq 1$ for the 2×2 setting [37] and for some regimes of the $K \times K$ setting [38]. Under the instantaneous power constraint $\|\mathbf{T}_{j,n}\|^2 \leq 1$ that we consider in this work, this approach also achieves the centralized DoF, but at the expense of a power back-off that allows the most precise TXs to have enough power to realign with high-enough probability the interference generated by the TXs that apply a *fixed* precoder [36], [43]. The required power back-off scales as $\log_2(P)$, and thus it does not vanish at high SNR. Consequently, it incurs an important rate penalty that prevents the distributed scheme from achieving centralized performance. This case is a illustrative example of the limitation of DoF analysis, as it does not take into account such power back-off (because the penalty is $o(\log_2(P))$), and thus the DoF analysis is oblivious to the loss of performance that the power back-off creates.

These two approaches are the opposed and extreme solutions for the question of whether the TXs should use their instantaneous CSIT: While in the first case the TXs naively consider their CSIT to be perfect, in the second case these TXs completely disregard their CSIT. In this work, we bridge the gap between the two strategies, which will prove instrumental to achieve our main results.

We present in the following several definitions that help to emphasize the two main limitations of distributed precoding.

Definition 2 (Consistency). *Consider two TXs, each endowed with a different CSI. Suppose that the precoder of TX 1 depends on the decision at TX 2 such that we can write \mathbf{T}_1 as $\mathbf{T}_1 = f(\hat{\mathbf{H}}^{(1)}, g(\hat{\mathbf{H}}^{(2)}))$, where f, g are generic functions. In a D-CSIT setting, TX 1 has no access to $\hat{\mathbf{H}}^{(2)}$. Let us assume that TX 1 estimates $g(\hat{\mathbf{H}}^{(2)})$ from $\hat{\mathbf{H}}^{(1)}$ as $\hat{g}(\hat{\mathbf{H}}^{(1)})$. Then, the computation is said to be Consistent if and only if $\hat{g}(\hat{\mathbf{H}}^{(1)}) = g(\hat{\mathbf{H}}^{(2)})$. Otherwise, it is said to be Inconsistent.*

Definition 3 (Power Outage). *Let \mathbf{T}_j denote the linear precoding matrix computed at TX j . For any precoder \mathbf{T}_j , a power outage occurs if the instantaneous power constraint is violated.*

In other words, the concept of *consistency* in Definition 2

stresses the fact that the TXs should precoder in a coherent manner with respect to the actions of the other TXs. In turn, the concept of *Power Outage* reflects that even if a certain TX acquires a perfect knowledge of the precoder applied at all the other TXs, it may not be able to reduce the interference if the power required for that action overpasses the instantaneous power constraint. Throughout this document, we will be interested in the cases in which the precoding is *both consistent and feasible*:

Definition 4 (Feasible Consistency). *Several TXs apply a Feasible Consistent precoder if the precoding coefficients are Consistent and there is no Power Outage.*

One of our contributions is to show that these limitations can be overcome by encouraging *consistency* among the different TXs, at the cost of reducing the precision of precoding at some TXs.

IV. MAIN RESULTS

Our main contributions rely on a novel ZF-type precoding scheme coined *Consistent Decentralized ZF (CD-ZF)*, which is presented in detail in Section IV-B. In short, this scheme is an adaptation to distributed scenarios of the aforementioned centralized ZF precoding, such that the precoding applied at each TX is different if the TX is the best informed one or not. Let $R(\alpha_M)$ be the expected sum rate for our D-CSIT setting. Similarly, let $R^*(\alpha^{(1)})$ be the expected sum rate achieved by a Zero-Forcing scheme on the ideal centralized CSIT setting as described in Section III-A. Accordingly, the rate gap between those settings is defined as

$$\Delta R \triangleq R^*(\alpha^{(1)}) - R(\alpha_M). \quad (19)$$

We can now state our main result.

Theorem 1. *In the Network MISO setting with Distributed CSIT, with $N_1 \geq K - 1$ and $\alpha^{(M)} > 0$, the expected sum rate achieved by ZF-type schemes in the ideal centralized CSIT setting is asymptotically achieved, i.e.,*

$$\lim_{P \rightarrow \infty} R^*(\alpha^{(1)}) - R(\alpha_M) = 0. \quad (20)$$

Proof. The proof is relegated to Section V, and it builds on the proposed CD-ZF precoding scheme, which is presented in Section IV-B. \square

Corollary 1 (Rate-Offset under Distributed CSIT). *It holds from Theorem 1 that the rate offset \mathcal{R}_∞^d (defined in (1)) of ZF with Distributed CSIT is the same as for the ideal centralized setting, whose rate offset was shown in [42] to be constant for the case of $\alpha^* = 1$ with respect to Perfect CSIT ZF—and thus with respect to the capacity-achieving Dirty Paper Coding.*

Remarkably, Theorem 1 implies⁴ that it is possible to achieve not only the multiplexing gain but also the beamforming gain

⁴We assume (see (11)) that the inequality $\alpha^{(1)} > \alpha^{(2)}$ is strict. This follows because, if $\alpha^{(1)} = \alpha^{(2)}$, since both channel realizations and noise are Gaussian, we can write $\hat{\mathbf{H}}^{(2)}$ as a noisy version of $\hat{\mathbf{H}}^{(1)}$, with a noise variance $\sqrt{2}P^{-\alpha^{(1)}}$. Hence, the error generated when zero-forcing the interference will have $\sqrt{2}$ times more variance than in the centralized case, which precludes the convergence presented in Theorem 1.

achieved by the ideal N_T -antenna MISO BC, even if only N_1 antennas are endowed with the maximum precision. We would like to remark that the constraint $N_1 \geq K - 1$, i.e., that the TX with the most precise CSI has a number of antennas at least equal to the number of interfered RXs, comes from the fact that, if $N_1 < K - 1$, the use of only ZF is not enough to achieve the DoF of the centralized setting [38], and thus $\lim_{P \rightarrow \infty} R^*(\alpha^{(1)}) - R(\alpha_M) = \infty$. Despite that, it was shown in [38] that for certain regimes of the parameters $\alpha^{(j)}$ it is possible to reach the ideal centralized DoF. This is achieved by means of a transmission scheme which comprises interference quantization and retransmission, superposition coding at the TXs, and successive decoding at the RXs. Since in this work we focus on a simple ZF transmission, we restrict the analysis to the DoF-achieving regime $N_1 \geq K - 1$, because it is the only regime in which the DoF (and thus the rate gap) of the centralized CSIT setting can be achieved. For the other cases, the distributed CSIT setting does not achieve the centralized DoF, and thus the analysis here considered is not meaningful.

It is known that the optimal DoF of the centralized CSIT setting with precision $\alpha^{(1)}$ is $1 + (K - 1)\alpha^{(1)}$ [50], which is attained by means of superposition coding where a common message is broadcast and intended to be decoded before the zero-forced messages. It is remarkable that, in the regime of interest, $N_1 \geq K - 1$, the distributed CSIT setting performance still converges asymptotically to the centralized performance even if superposition coding is applied. This comes from the fact that the instantaneous power applied converges to the one used in the centralized setting (as we prove at a later stage), such that the common symbol broadcast can be sent with the same rate.

A. Achievability: Some Insights

Theorem 1 evidences that the issues associated with *feasible consistency* between the TXs (which are enunciated in Section III-B) can be overcome. Intuitively, the strategies mentioned in Section III-B are extreme cases of consistency. Particularly, the Naive ZF represents the extreme in which consistency is not considered, whereas the Active-Passive ZF embodies the extreme with perfect consistency but limited precision and possible Power Outage, such that there may not be *feasible* consistency. The block diagrams of Naive ZF and Active-Passive ZF are shown in Fig. 3a and Fig. 3b, respectively. The main question is whether a good compromise can be found.

We can build on the idea first presented in [43] for the simple 2×2 setting, i.e., that discretizing the decision space of the TXs helps to enforce consistency. Yet, the application of this idea is not straightforward, as in [43] the only source of inconsistency was a single scalar power parameter, and no beamforming was possible. Specifically, the strategy in [43] was to design the ZF precoder such that the TX having worse channel estimate for a certain user (i.e., smaller $\alpha^{(j)}$) precodes with a single real value, and thus it does not consider the possible phase of the coefficient. Then, the other TX (the *best-informed* one) fully controls the phase tuning necessary to cancel out the interference, such that both TXs need to agree only on a single value of transmit power—which must yet be agreed upon with

high precision. In the general $M \times K$ setting here considered, however, beamforming and phase precoding must be applied at every TX equipped with more than one antenna, which rules out the aforementioned strategy.

Nevertheless, the main insight is still valid: By means of discretizing the decision space of the TXs that do not have the best CSI, we construct a probabilistic hierarchical setting in which the best informed TX is able to estimate correctly the action taken by the other TXs with a certain probability. Interestingly, this discretization can be applied to either the available information (the channel matrix) or the output parameters (the precoding vector). Both cases are illustrated in Fig. 4. This flexibility is due to the asymptotic nature of our analysis and the properties of linear systems with respect to error propagation.⁵ It is however clear that the performance at low-to-medium SNR can importantly differ for each of the cases. In this document, we focus on the scheme that quantizes the channel matrix for the sake of a better understanding, as the proof is less devious, and because the proof for the other case (quantizing the precoder) follows the same approach and steps.

Actually, we will see that this quantization is crucial for the results. Indeed, if TX 1 attempts to estimate the information used at the other TXs without any quantization, the estimation error will scale as the error variance of the other TX. Thus, TX 1 would be unable to compensate and cancel out the interference generated by the other TXs.

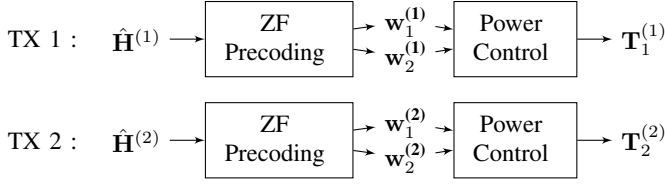
The key for attaining the surprising result of Theorem 1 is the proposed precoding scheme, whose rigorous description is presented in the following section. The proof of Theorem 1 relies on a simple idea: Let A be a set enclosing the *feasible consistent* cases in which the precoders transmit coordinately, and A^c its complementary event. Hence, the rate gap $\Delta R \triangleq R^*(\alpha^{(1)}) - R(\alpha_M)$ can be expressed as

$$\Delta R = \Delta R_A \Pr(A) + \Delta R_{A^c} \Pr(A^c). \quad (21)$$

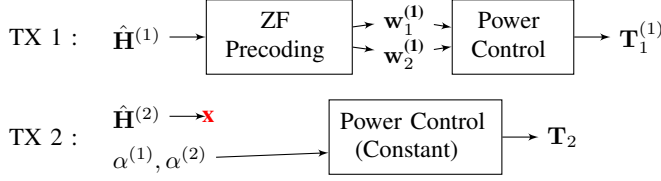
The transmission scheme has to be both feasible consistent ($\Pr(A^c) \rightarrow 0$) and, for the consistent cases, it has to be precise ($\Delta R_A \rightarrow 0$). Withal, it turns out that these two conditions are not enough to obtain Theorem 1. In particular, we need not only that the scheme is feasible consistent ($\Pr(A^c) \rightarrow 0$) but also that it attains consistency *fast enough* with respect to the CSI scaling, i.e., that $\Delta R_{A^c} \Pr(A^c) \rightarrow 0$.

As a matter of example, consider that the rate gap for the inconsistent cases (ΔR_{A^c}) scales proportionally to $\log(P)$ (as it will proven later on) such that $\Delta R_{A^c} = \Theta(\log(P))$. Then, if the probability of inconsistency $\Pr(A^c)$ approached to 0 at a convergence rate proportional to $\frac{1}{\log(P)}$ ($\Pr(A^c) = \Theta(\frac{1}{\log(P)})$), the term $\Delta R_{A^c} \Pr(A^c)$ would not vanish and Theorem 1 would not hold. Fortunately, we can design the achievable scheme so as to reduce the probability of cases without consistency *and* without power outage faster than $\Theta(\frac{1}{\log(P)})$, *while providing a precise precoding and vanishing gap ΔR_A in the consistent cases*. This last condition is critical: As P increases, the

⁵These properties will be detailed in Appendix VI. There, it is shown that the error scaling of the output of a system that applies additions, products, and matrix pseudo-inverses is the same as the error scaling of the input.

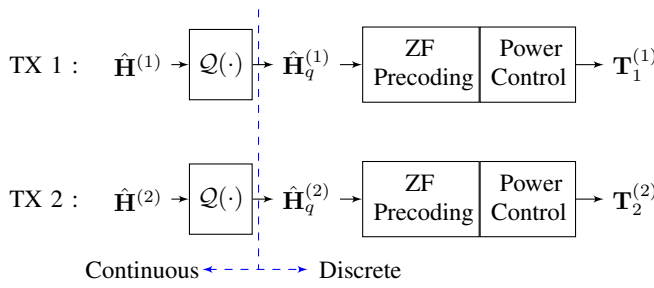


(a) Block diagram of conventional ZF applied naively in the 2×2 Distributed CSIT scenario (No Consistency).

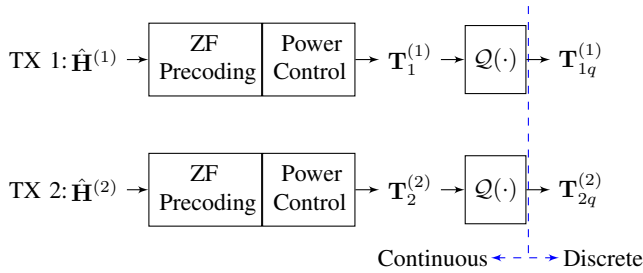


(b) Block diagram of AP-ZF applied in the 2×2 Distributed CSIT scenario (Full Consistency but possible infeasibility).

Fig. 3: Simple strategies for distributed precoding.



(a) Quantizing input (CSIT).



(b) Quantizing output (precoder).

Fig. 4: Two manners of discretizing decision space: At the input (information) or at the output (action).

precision of the precoder in the consistent cases must increase faster than $\mathcal{O}(P)$. Otherwise, ΔR_A would not vanish. The key to attaining sufficient precision for the consistent cases while reducing at the same time the probability of inconsistent cases is that they depend on different scaling of P ($\Theta(\log(P))$ versus $\Theta(P)$). This is rigorously shown in Section V.

B. Proposed Transmission Scheme: Consistent Distributed ZF

We present in the following the Consistent Distributed ZF (CD-ZF) precoding scheme, where the CSIT of the TXs that do not have access to the most precise estimate is quantized to improve the consistency of the scheme. Later, we analyze the feasibility of the proposed precoder. The CD-ZF scheme

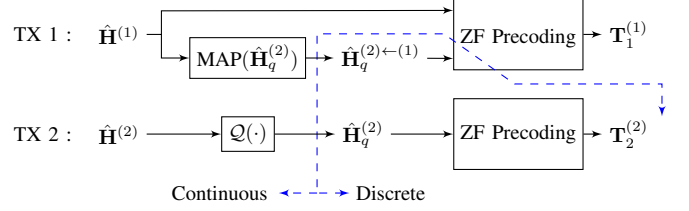


Fig. 5: Block diagram of Consistent Distributed ZF applied in the 2×2 Distributed CSIT scenario

presents an uneven structure, such that each TX applies a different strategy depending on who has higher average precision. Furthermore, the proposed scheme independently computes the precoder for the symbols of different RXs, except for the final power normalization. We recall that the precoding vector for the message intended at RX i is defined in (7) as $\mu \mathbf{w}_i \in \mathbb{C}^{N_T \times 1}$, and the segment of $\mu \mathbf{w}_i$ applied at TX j is given by $\mu \mathbf{w}_{i,j} \in \mathbb{C}^{N_j \times 1}$.

1) *Quantizing the CSIT*: The block diagram of this precoding scheme is depicted in Fig. 5. Because of the uneven structure of the precoder, we separately describe the precoding strategy at the best-informed TX (TX 1) and at any other TX (TX 2 to TX M). Let us consider first the later. The main limitation of the distributed precoding is not the error variance at the restricting TXs, but the impossibility at TX 1 of knowing what the other TXs are going to transmit. In order to overcome this problem, all the TXs but TX 1 quantize their estimate of the channel matrix with a known quantizer \mathcal{Q} . Hence, for any $j > 1$, TX j does not directly use its CSIT $\hat{\mathbf{H}}^{(j)}$ to precode, but first pre-processes it. In other words, TX j applies

$$\hat{\mathbf{H}}_q^{(j)} = \mathcal{Q}(\hat{\mathbf{H}}^{(j)}). \quad (22)$$

The characteristics of the quantizer \mathcal{Q} will be detailed later. Then, TX j naively applies a centralized ZF scheme as described in Section III-A but in a distributed manner (based on $\hat{\mathbf{H}}_q^{(j)}$). Since the quantization transforms the continuous variable $\hat{\mathbf{H}}^{(j)}$ into a discrete one, it facilitates that the setting becomes a hierarchical setting in which the information available at other TXs is estimated without explicit communication.

We focus now on the precoder at the best-informed TX (TX 1), which attempts to correct the error of the previous TXs. Let TX 1 estimate $\hat{\mathbf{H}}_q^{(j)}$ based on its own information $\hat{\mathbf{H}}^{(1)}$, e.g. by computing the Maximum A Posteriori estimator (MAP) of $\hat{\mathbf{H}}_q^{(j)}$:

$$\hat{\mathbf{H}}_q^{(j) \leftarrow (1)} = \underset{\hat{\mathbf{H}}_q^{(j)} \in \mathcal{Q}(\mathbb{C}^{K \times N_T})}{\operatorname{argmax}} \Pr \left(\hat{\mathbf{H}}_q^{(j)} \mid \hat{\mathbf{H}}^{(1)} \right). \quad (23)$$

It is important to notice that the quantized value $\hat{\mathbf{H}}_q^{(j)}$ is not intended to be transmitted, but it is aimed at helping TX 1 to estimate the CSIT used at TX j , *without any explicit communication between them*.

2) *Canceling the interference*: The CD-ZF scheme naively assumes that TX 1 correctly estimates the CSIT at TX j , for any j in \mathbb{N}_M , such that

$$\hat{\mathbf{H}}_q^{(j) \leftarrow (1)} = \hat{\mathbf{H}}_q^{(j)}. \quad (24)$$

The probability that (24) holds true will be evaluated in the following section. Therefore, the fact that (24) is fulfilled implies that the setting is *consistent*—although it does not guarantee feasibility, which will be ensured by the parameter μ . The goal of TX 1 is to imitate the interference cancellation performance that the centralized ZF scheme would achieve if every other TX also had access to $\hat{\mathbf{H}}^{(1)}$.⁶ In order to provide some insights, we first describe the single-antenna case with 2 TXs and 2 RXs.

a) 2×2 case: Let $\bar{i} \triangleq i \pmod{2} + 1$. Mathematically, the goal is to have $|\hat{\mathbf{h}}_i^{(1)} \mathbf{w}_{\bar{i}}| = |\hat{\mathbf{h}}_i^{(1)} \mathbf{v}_{\bar{i}}|$, which can be rewritten as

$$|\hat{\mathbf{h}}_{i,1}^{(1)} \mathbf{w}_{\bar{i},1} + \hat{\mathbf{h}}_{i,2}^{(1)} \mathbf{w}_{\bar{i},2}| = |\hat{\mathbf{h}}_{i,1}^{(1)} \mathbf{v}_{\bar{i},1} + \hat{\mathbf{h}}_{i,2}^{(1)} \mathbf{v}_{\bar{i},2}|. \quad (25)$$

We remind the reader that \mathbf{w} stands for the distributed precoder whereas \mathbf{v} stands for the centralized precoder. Under the assumption that TX 1 correctly estimates $\hat{\mathbf{H}}_q^{(2)}$, it knows $\mathbf{w}_{\bar{i},2}$. Then, TX 1 computes its precoder such that

$$\mathbf{w}_{\bar{i},1} = \mathbf{v}_{\bar{i},1} + \underbrace{(\hat{\mathbf{h}}_{i,1}^{(1)})^\dagger \hat{\mathbf{h}}_{i,2}^{(1)} (\mathbf{v}_{\bar{i},2} - \mathbf{w}_{\bar{i},2})}_{\phi_{\bar{i}}}, \quad (26)$$

where $(\mathbf{x})^\dagger$ denotes the pseudo-inverse⁷ of \mathbf{x} , which is known to have minimal Frobenius norm among all the generalized inverses [53]. The term $\phi_{\bar{i}}$ represents the correction term that TX 1 has to apply in order to compensate the error introduced by TX 2; note that (26) satisfies (25).

b) $M \times K$ case: The generalization from the 2×2 case to the general $M \times K$ case needs one more step. Let TX 1 have $N_1 \geq K - 1$ antennas. The goal is again to obtain the same interference cancellation as for the centralized precoder, such that, for any RX $i \in \mathbb{N}_K$,

$$\sum_{\ell \in \mathbb{N}_K \setminus i} |\hat{\mathbf{h}}_i^{(1)} \mathbf{w}_\ell|^2 = \sum_{\ell \in \mathbb{N}_K \setminus i} |\hat{\mathbf{h}}_i^{(1)} \mathbf{v}_\ell|^2. \quad (27)$$

The equality in (27) is attained if (but not only if) $\hat{\mathbf{h}}_i^{(1)} \mathbf{w}_\ell = \hat{\mathbf{h}}_i^{(1)} \mathbf{v}_\ell$, for any $i, \ell \in \mathbb{N}_K$ such that $i \neq \ell$. Let us split the precoding and channel vectors in two parts: $\mathbf{v}_{\ell,1}$, $\mathbf{w}_{\ell,1}$ and $\hat{\mathbf{h}}_{i,1}^{(1)}$ denote the sub-vector corresponding to the antennas of TX 1, whereas $\mathbf{v}_{\ell,\bar{1}}$, $\mathbf{w}_{\ell,\bar{1}}$ and $\hat{\mathbf{h}}_{i,\bar{1}}^{(1)}$ represent the sub-vector corresponding to the antennas from TX 2 to TX M . The sub-matrices $\hat{\mathbf{H}}_{\bar{\ell},1}$ and $\hat{\mathbf{H}}_{\bar{\ell},\bar{1}}$ are defined in the same manner, and they are illustrated in Fig. 6 for the sake of better comprehension. Note that we have introduced the notation $N_{\bar{1}} \triangleq N_T - N_1$. We can expand the condition $\hat{\mathbf{h}}_i^{(1)} \mathbf{w}_\ell = \hat{\mathbf{h}}_i^{(1)} \mathbf{v}_\ell$ as a matrix equation in which $\mathbf{w}_{\ell,1}$ has to satisfy

$$\hat{\mathbf{H}}_{\bar{\ell},1} \mathbf{w}_{\ell,1} = \hat{\mathbf{H}}_{\bar{\ell},1} \mathbf{v}_{\ell,1} + \hat{\mathbf{H}}_{\bar{\ell},\bar{1}} (\mathbf{v}_{\ell,\bar{1}} - \mathbf{w}_{\ell,\bar{1}}). \quad (28)$$

⁶We could develop more convoluted schemes in order to increase the achievable rate at low or medium SNR regimes. As a matter of example, we could analyze the minimum mean-square error (MMSE) precoding as in [49], or design several layers of quantization such that, for every j , TX $(j-1)$ tries to correct the interference generated by TX j , in a similar manner to the algorithm presented in [39]. Nevertheless, in this work we focus on the asymptotic regime, and thus we maintain the scheme in its simplest expression.

⁷ $(\mathbf{x})^\dagger$ could also represent the regularized pseudo-inverse.

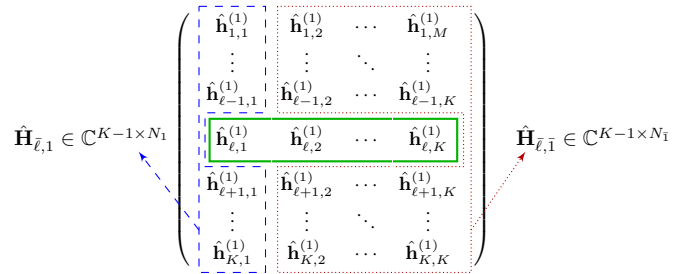


Fig. 6: Illustration of the channel sub-matrices $\hat{\mathbf{H}}_{\bar{\ell},1}$ and $\hat{\mathbf{H}}_{\bar{\ell},\bar{1}}$.

The precoding vector at TX 1 is then selected as

$$\mathbf{w}_{\ell,1} = \mathbf{v}_{\ell,1} + \underbrace{\hat{\mathbf{H}}_{\bar{\ell},1}^\dagger \hat{\mathbf{H}}_{\bar{\ell},\bar{1}} (\mathbf{v}_{\ell,\bar{1}} - \mathbf{w}_{\ell,\bar{1}})}_{\phi_\ell}. \quad (29)$$

Remark 3. We can see from the dimensionality of the linear system in (28) that we need $N_1 \geq K - 1$ in order to make (28) feasible with high-probability. This is another way of understanding why we need the restriction that $N_1 \geq K - 1$. However, the main reason to have this constraint comes from the limitations of the Distributed CSIT setting. Specifically, it is known that the use of ZF precoding (or other linear precodings) in the Distributed CSIT setting can only achieve the DoF of its centralized counterpart when this constraint is satisfied [38], and thus $\lim_{P \rightarrow \infty} R^*(\alpha^{(1)}) - R(\alpha_M) = \infty$ otherwise.

In (28), it is ensured that the interference received is the same as for the centralized setting. It is possible also to ensure that the receive signal $\hat{\mathbf{h}}_i^{(1)} \mathbf{w}_i$ is equal to the one of the centralized setting. This would add an extra equation to the linear system that would require an extra antenna at TX 1. However, it is not necessary as the received intended signal turns out to be statistically equivalent in both distributed and centralized settings.

3) *Feasibility and consistency:* In the previous description of the scheme, it has been assumed that TX 1 obtains a feasible consistent precoder, such that it correctly estimates the CSI at the other TXs, and that the obtained precoding vector can be used for transmission. In fact, the CD-ZF scheme naively considers that the precoder is *always* feasible and consistent.

However, the transmission scheme will suffer from the two main issues described in Section III-B: Power outage, since it has to satisfy that $\|\mathbf{T}_{1,n}\| \leq 1$ for any $n \in \mathbb{N}_{N_1}$, and *Consistency*, as the quantization of the CSI at TX j allows TX 1 to obtain the quantized CSI only with a certain probability. Further, as explained in Section IV-A, we also need to demonstrate that the interference cancellation is precise enough in the feasible consistent cases. In the following we present some properties that will prove instrumental in dealing with those limitations. For that, we introduce a set of quantizers that are essential in the proof of Theorem 1. We first tackle the probability of inconsistency.

Definition 5 (Asymptotically Consistent Quantizers). A quantizer \mathcal{Q} is said to be Asymptotically Consistent if the probability of correct estimation of the MAP estimator at TX 1 satisfies

$$\Pr \left(\hat{\mathbf{H}}_q^{(j) \leftarrow (1)} \neq \hat{\mathbf{H}}_q^{(j)} \right) = o \left(\frac{1}{\log_2(P)} \right), \quad \forall j \in \mathbb{N}_M. \quad (\text{P1})$$

Property (P1) implies that it is possible to induce that the probability of having *inconsistent* precoding among TXs vanishes faster than $1/\log_2(P)$. This fact implies that the rate impact of inconsistent precoding events vanishes asymptotically, as mentioned in Section IV-A. Clearly, it remains to prove that there exists some quantizer \mathcal{Q}_c that satisfies (P1) while also ensuring both feasibility and an adequate precision for the consistent cases. Surprisingly, very simple quantizers as the scalar uniform quantizer satisfy the requirements. To prove this statement, we first show that this quantizer is an Asymptotically Consistent Quantizer.

Lemma 1. *Let $\mathcal{Q}_u(\mathbf{X})$ be a scalar uniform quantizer with quantization step $q = \bar{P}^{-\alpha_q}$, where α_q is such that $\alpha^{(j)} > \alpha_q > 0$, for all $j \in \mathbb{N}_M$. Then, \mathcal{Q}_u is an Asymptotically Consistent Quantizer and*

$$\Pr\left(\hat{\mathbf{H}}_q^{(j)\leftarrow(1)} \neq \hat{\mathbf{H}}_q^{(j)}\right) = o\left(\frac{1}{\log_2(P)}\right), \quad \forall j \in \mathbb{N}_M. \quad (30)$$

Proof. The proof is relegated to Appendix V. \square

Note that \mathcal{Q}_u is a scalar quantizer. Thus, for a matrix \mathbf{A} , the notation $\mathbf{A}_q = \mathcal{Q}_u(\mathbf{A})$ denotes (with an abuse of notation) that \mathbf{A}_q is composed of the independent scalar quantization of the real and imaginary part of each element in \mathbf{A} . Obviously, using vector quantization would improve the performance. However, as the proof of Theorem 1 is constructive, we are interested in an intuitive example, and thus we use \mathcal{Q}_u for the sake of simplicity. The analysis at medium and low SNR, which requires an optimization on the applied quantizer, is an interesting research topic that is however out of the scope of this work.

We focus now on the probability of *power outage*. Let us denote the event of power outage as \mathcal{P}_o and recall that the precoder at the n -th antenna of TX j is expressed by $\mathbf{T}_{j,n}$. In that case,

$$\mathcal{P}_o \triangleq \bigcup_{\substack{n \in \mathbb{N}_{N_j} \\ j \in \mathbb{N}_M}} \{\|\mathbf{T}_{j,n}\| > 1\}. \quad (31)$$

The power outage probability is handled by the parameter μ of the precoding vector $\mu \mathbf{w}_i \in \mathbb{C}^{N_T \times 1}$. The value of μ acts as power back-off that can be tuned to achieve the required scaling, as stated in the following lemma.

Lemma 2. *Let $\mu = 1 - \varpi$, where $\varpi > 0$, $\varpi = \Theta(\bar{P}^{-\alpha_\mu})$, and $\alpha_\mu < \alpha_q$. Then,*

$$\Pr(\mathcal{P}_o) = o\left(\frac{1}{\log_2(P)}\right). \quad (P2)$$

Proof. See Appendix III. \square

Similarly to property (P1), property (P2) in Lemma 2 implies that *power outage* events are negligible in terms of asymptotic rate. The only TX that may incur in power outage is TX 1, as the other TXs apply the naive centralized precoder and hence they will always satisfy the power constraint.

To this extent, we have shown that the uniform scalar quantizer \mathcal{Q}_u enables us to reach the requirements regarding the probability of the cases that are not *feasible consistent*.

Hence, it remains to prove that it also attains high enough precision in the feasible consistent cases. This will be proven in Section V. Hereinafter, we assume that the uniform quantizer of Lemma 1 is applied in the CD-ZF.

C. Hierarchical CSIT Setting

Theorem 1 shows that it is possible to asymptotically attain the rate of the centralized setting. Its performance at low-to-medium SNR is however limited by the probability of obtaining a *feasible consistent* precoder. This probability depends on the quantizer applied, the power back-off considered, and the values of $\alpha^{(j)}$, and hence it is challenging to obtain. As shown in Section IV-B, the precoder is computed assuming a correct estimation of the CSI at the other TXs. Consequently, if the probability of *consistency* is low, the scheme does not perform properly. Moreover, this probability decreases as the network size increases, since TX 1 needs to correctly estimate more parameters.

This limitation is inherent to the D-CSIT setting here assumed, in which each TX only knows its own CSI. However, there exist another practical setting with distributed CSI but in which there is more structure in the network CSI: The Hierarchical CSIT setting (H-CSIT). In this setting, each TX is endowed with its own multi-user CSI $\hat{\mathbf{H}}^{(j)}$, as in the D-CSIT setting, but it is also endowed with the CSI of the TXs having less precision than itself. Namely, in the sorted CSI scenario with $\alpha^{(1)} > \dots \geq \alpha^{(M)}$, TX j has access to $\{\hat{\mathbf{H}}^{(j)}, \hat{\mathbf{H}}^{(j+1)}, \dots, \hat{\mathbf{H}}^{(M)}\}$.

This scenario, although it may seem less practical, may arise in many heterogeneous networks. Fig. 2 depicts an example: Suppose that the RXs are all connected to the same main TX (e.g. TX 1), and the other TXs are remote radio-heads that receive a coarse version of the CSI by means of a wireless link from TX 1. In this use case, TX 1 will know the CSI available at each other TX. If the CSI sharing is done through dedicated links for each TX, each TX would receive CSI with precision proportional to its own link. If the CSI is broadcast, they may obtain an estimate with different precision if layered encoding [55] or analog feedback [56] is used.

Corollary 2. *Theorem 1 also holds in the Hierarchical CSIT setting, and hence $\lim_{P \rightarrow \infty} R^*(\alpha^{(1)}) - R(\alpha_M) = 0$.*

Proof. The proof follows directly from the proof of Theorem 1 in Section V. \square

In this setting, TX 1 already knows $\hat{\mathbf{H}}^{(j)}$ for any $j \in \mathbb{N}_M$. Hence, the discretization of the variables at the other TXs is not needed, and the precoders are *consistent* with probability 1. In fact, the idea of quantizing the CSI in the D-CSIT setting boils down to making the CSIT setting asymptotically hierarchical with a high enough probability. The only effect that may restrain TX 1 to achieve the centralized performance is the power outage. Therefore, the performance at medium SNR will improve with respect to the general Distributed CSIT case, and moreover, this performance is not affected by the size of the network, as we will see in the numerical examples of Section VI.

V. PROOF OF THEOREM 1

In order to prove Theorem 1, we need to demonstrate that the user rate gap $\Delta R_i = R_i^*(\alpha^{(1)}) - R_i(\alpha_M)$ vanishes, which directly yields that $\Delta R = \sum_{i \in \mathbb{N}_K} \Delta R_i$ will also vanish. The proof is divided in several steps: First, armed with Lemma 1 and Lemma 2, we show that both *power outage* and *inconsistent precoding* events can be made negligible in terms of rate loss. Then, we prove that the rate gap also vanishes in the feasible consistent cases thanks to the fine precision of the CD-ZF scheme in these cases. We demonstrate this by showing that both the interference received and the total power received in the distributed setting converge to their counterparts of the centralized setting.

A. Neglecting Non-consistent Events

The proof of Theorem 1 builds on Lemma 1 and Lemma 2. Indeed, the proposed scheme will perform poorly in the cases in which the precoder is not feasible consistent, as the scheme is built on the naive assumption that it is *always* feasible and consistent. Nevertheless, both Lemma 1 and Lemma 2 illustrate that those events can be made very unlikely (in particular, the probability of these events is $o(\frac{1}{\log(P)})$). Let \mathcal{H}_\neq denote the set of *inconsistent* events, i.e., $\mathcal{H}_\neq \triangleq \bigcup_{2 \leq j \leq M} \{\hat{\mathbf{H}}_q^{(j) \leftarrow (1)} \neq \hat{\mathbf{H}}_q^{(j)}\}$. Hence, the probability of having *feasible consistent* precoding is $\Pr(\mathcal{P}_o^c \cap \mathcal{H}_\neq^c)$. By means of the law of total expectation, we can split the expected rate gap for RX i (ΔR_i) as⁸

$$\begin{aligned} \Delta R_i &= \Pr(\mathcal{P}_o \cup \mathcal{H}_\neq) \Delta R_{i|\mathcal{P}_o \cup \mathcal{H}_\neq} \\ &\quad + \Pr(\mathcal{P}_o^c \cap \mathcal{H}_\neq^c) \Delta R_{i|\mathcal{P}_o^c \cap \mathcal{H}_\neq^c}. \end{aligned} \quad (32)$$

Let us focus on $\Delta R_{i|\mathcal{P}_o \cup \mathcal{H}_\neq}$. The rate gap can be upper-bounded by setting the rate of the D-CSIT setting to 0, such that $\Delta R_{i|\mathcal{P}_o \cup \mathcal{H}_\neq} \leq R_{i|\mathcal{P}_o \cup \mathcal{H}_\neq}^*(\alpha^{(1)})$. By neglecting the received interference, we can write that

$$\begin{aligned} R_{i|\mathcal{P}_o \cup \mathcal{H}_\neq}^*(\alpha^{(1)}) &\leq \mathbb{E}_{|\mathcal{P}_o \cup \mathcal{H}_\neq} \left[\log_2 \left(1 + P |\mathbf{h}_i \mathbf{v}_i|^2 \right) \right] \\ &\leq \log_2(P) + \mathbb{E}_{|\mathcal{P}_o \cup \mathcal{H}_\neq} \left[\log_2 \left(1 + |\mathbf{h}_i \mathbf{v}_i|^2 \right) \right]. \end{aligned} \quad (33)$$

Moreover, the set $\mathcal{P}_o \cup \mathcal{H}_\neq$ depends on the different estimation noise at each TX, which is absent in the centralized setting. Accordingly, (33) implies that $R_{i|\mathcal{P}_o \cup \mathcal{H}_\neq}^*(\alpha^{(1)}) = \Theta(\log_2(P))$. Hence, it follows from Lemma 1 and Lemma 2 that

$$\Pr(\mathcal{P}_o \cup \mathcal{H}_\neq) \Delta R_{i|\mathcal{P}_o \cup \mathcal{H}_\neq} = o\left(\frac{1}{\log_2(P)}\right) \Theta(\log_2(P))$$

and consequently

$$\Delta R_i = \Delta R_{i|\mathcal{P}_o^c \cap \mathcal{H}_\neq^c} + o(1). \quad (34)$$

Thus, in the remainder of the proof we assume w.l.o.g. that TX 1 knows $\hat{\mathbf{H}}_q^{(j)}$ for any $j \in \mathbb{N}_M$, and that there is no power outage, as both cases become negligible at high SNR. This assumption implies that the setting becomes hierarchical, because TX 1 correctly estimates the quantized CSI of the other

⁸Given a certain feasible event A and its complementary event A^c , the law of total expectation states that $\mathbb{E}[X] = \Pr(A) \mathbb{E}[X|A] + \Pr(A^c) \mathbb{E}[X|A^c]$. Furthermore, for any two events A_1, A_2 , and union event $A = A_1 \cup A_2$, it follows that $A^c = (A_1 \cup A_2)^c = (A_1^c \cap A_2^c)$.

TXs. It is important to remark that this simplification is only possible because of the proposed scheme, in which we apply a correct power back-off and quantization step. Indeed, the more important outcome of this work—and the main purpose of the careful design of the scheme—is not (34) but the fact that $\Delta R_{i|\mathcal{P}_o^c \cap \mathcal{H}_\neq^c}$ converges to the centralized setting rate.

B. Reformulating the Rate Gap

We can rewrite the rate gap for RX i as

$$\begin{aligned} \Delta R_i &= \mathbb{E} \left[\log_2 \left(1 + \frac{P |\mathbf{h}_i \mathbf{v}_i|^2}{1 + P \sum_{\ell \neq i} |\mathbf{h}_i \mathbf{v}_\ell|^2} \right) \right] \\ &\quad - \mathbb{E} \left[\log_2 \left(1 + \frac{P |\mu \mathbf{h}_i \mathbf{w}_i|^2}{1 + P \sum_{\ell \neq i} |\mu \mathbf{h}_i \mathbf{w}_\ell|^2} \right) \right] \\ &= \mathbb{E} \left[\log_2 \left(\underbrace{\frac{1 + P \sum_{\ell \in \mathbb{N}_K} |\mathbf{h}_i \mathbf{v}_\ell|^2}{1 + P \sum_{\ell \in \mathbb{N}_K} |\mu \mathbf{h}_i \mathbf{w}_\ell|^2}}_{\mathcal{F}_D} \right) \right] \\ &\quad + \mathbb{E} \left[\log_2 \left(\underbrace{\frac{1 + P \sum_{\ell \neq i} |\mu \mathbf{h}_i \mathbf{w}_\ell|^2}{1 + P \sum_{\ell \neq i} |\mathbf{h}_i \mathbf{v}_\ell|^2}}_{\mathcal{F}_I} \right) \right]. \end{aligned} \quad (35)$$

This rewriting of ΔR_i allows us to separate the ratio of received interference power (\mathcal{F}_I) and the ratio of total received power (\mathcal{F}_D). In the following, we will prove that $\lim_{P \rightarrow \infty} \Delta R_i = 0$ by showing that $\lim_{P \rightarrow \infty} \mathbb{E}[\log_2(\mathcal{F}_i)] = 0$ for both \mathcal{F}_D and \mathcal{F}_I . We start with \mathcal{F}_I for simplicity, and later we apply a similar argument to \mathcal{F}_D .

C. Analysis of the Interference Ratio (\mathcal{F}_I)

We prove the convergence by upper and lower-bounding \mathcal{F}_I , and then showing that both bounds converge to 0. We recall that we assume that TX 1 is able to transmit the desired precoding vector of (26) since the opposite case only yields an $o(1)$ rate contribution.

1) *Upper-bounding* $\mathbb{E}[\log_2(\mathcal{F}_I)]$: Note that, since $\mu \leq 1$,

$$\begin{aligned} \mathbb{E} \left[\log_2 \left(\frac{1 + P \sum_{\ell \neq i} |\mu \mathbf{h}_i \mathbf{w}_\ell|^2}{1 + P \sum_{\ell \neq i} |\mathbf{h}_i \mathbf{v}_\ell|^2} \right) \right] \\ \leq \mathbb{E} \left[\log_2 \left(\underbrace{\frac{1 + P \sum_{\ell \neq i} |\mathbf{h}_i \mathbf{w}_\ell|^2}{1 + P \sum_{\ell \neq i} |\mathbf{h}_i \mathbf{v}_\ell|^2}}_{\mathcal{F}'_I} \right) \right], \end{aligned} \quad (36)$$

where we have introduced the notation \mathcal{F}'_I for the sake of readability. Let η be a scalar satisfying $0 \leq \eta \leq 1$. We can split the expectation based on whether the term \mathcal{F}'_I is smaller than $1 + \eta$ or not. Therefore,

$$\begin{aligned} \mathbb{E}[\log_2(\mathcal{F}'_I)] &= \Pr(\mathcal{F}'_I < 1 + \eta) \mathbb{E}_{\mathcal{F}'_I < 1 + \eta}[\log_2(\mathcal{F}'_I)] \\ &\quad + \Pr(\mathcal{F}'_I \geq 1 + \eta) \mathbb{E}_{\mathcal{F}'_I \geq 1 + \eta}[\log_2(\mathcal{F}'_I)]. \end{aligned} \quad (37)$$

Now we present a useful lemma.

Lemma 3. *Let $\eta = \bar{P}^{-\varepsilon}$, with $\alpha_q > \varepsilon > 0$ and ε arbitrarily small. Then,*

$$\Pr(\mathcal{F}'_I \geq 1 + \eta) = o\left(\frac{1}{\log_2(P)}\right) \quad (38)$$

and

$$\Pr\left(\frac{1}{\mathcal{F}'_{\mathcal{I}}} \geq 1 + \eta\right) = o\left(\frac{1}{\log_2(P)}\right). \quad (39)$$

Proof. The proof is relegated to Appendix II-A. \square

Let $\eta = \bar{P}^{-\varepsilon}$, with $\alpha_q > \varepsilon > 0$ and ε arbitrarily small. Then, (37) becomes

$$\begin{aligned} \mathbb{E}[\log_2(\mathcal{F}'_{\mathcal{I}})] &\leq \mathbb{E}_{\mathcal{F}'_{\mathcal{I}} < 1+\eta}[\log_2(\mathcal{F}'_{\mathcal{I}})] \\ &\quad + o\left(\frac{1}{\log_2(P)}\right) \mathbb{E}_{\mathcal{F}'_{\mathcal{I}} \geq 1+\eta}[\log_2(\mathcal{F}'_{\mathcal{I}})] \quad (40) \\ &\leq \log_2(1 + \eta) + o(1) \end{aligned}$$

since $\mathbb{E}_{\mathcal{F}'_{\mathcal{I}} \geq 1+\eta}[\log_2(\mathcal{F}'_{\mathcal{I}})] = \mathcal{O}(\log_2(P))$. In order to prove that $\mathbb{E}_{\mathcal{F}'_{\mathcal{I}} \geq 1+\eta}[\log_2(\mathcal{F}'_{\mathcal{I}})] = \mathcal{O}(\log_2(P))$, note that

$$\begin{aligned} \mathbb{E}_{\mathcal{F}'_{\mathcal{I}} \geq 1+\eta}[\log_2(\mathcal{F}'_{\mathcal{I}})] &= \frac{1}{\Pr(\mathcal{F}'_{\mathcal{I}} \geq 1 + \eta)} \left(\mathbb{E}[\log_2(\mathcal{F}'_{\mathcal{I}})] \right. \\ &\quad \left. - \Pr(\mathcal{F}'_{\mathcal{I}} < 1 + \eta) \mathbb{E}_{\mathcal{F}'_{\mathcal{I}} < 1+\eta}[\log_2(\mathcal{F}'_{\mathcal{I}})] \right) \quad (41) \\ &\leq \frac{1}{\Pr(\mathcal{F}'_{\mathcal{I}} \geq 1 + \eta)} \mathbb{E}[\log_2(\mathcal{F}'_{\mathcal{I}})]. \end{aligned}$$

Furthermore,

$$\begin{aligned} \mathbb{E}[\log_2(\mathcal{F}'_{\mathcal{I}})] &\leq \mathbb{E}\left[\log_2\left(1 + P \sum_{\ell \neq i} |\mathbf{h}_i \mathbf{w}_\ell|^2\right)\right] \\ &\leq \log_2(P) + \mathbb{E}\left[\log_2\left(1 + \sum_{\ell \neq i} |\mathbf{h}_i \mathbf{w}_\ell|^2\right)\right]. \quad (42) \end{aligned}$$

From (41), (42), and the fact that $\Pr(\mathcal{F}'_{\mathcal{I}} \geq 1 + \eta) = \Theta(1)$, we obtain that $\mathbb{E}_{\mathcal{F}'_{\mathcal{I}} \geq 1+\eta}[\log_2(\mathcal{F}'_{\mathcal{I}})] = \mathcal{O}(\log_2(P))$.

2) *Lower-bounding* $\mathbb{E}[\log_2(\mathcal{F}'_{\mathcal{I}})]$: Let us now lower-bound the expectation. Note that

$$\mathbb{E}[\log_2(\mathcal{F}'_{\mathcal{I}})] \geq \log_2(\mu^2) + \mathbb{E}[\log_2(\mathcal{F}'_{\mathcal{I}})]. \quad (43)$$

Furthermore, lower-bounding (43) is equivalent to upper-bounding $\mathbb{E}\left[\log_2\left(\frac{1}{\mathcal{F}'_{\mathcal{I}}}\right)\right]$. By applying Lemma 3 and in a similar way as in (40), we obtain that $\mathbb{E}\left[\log_2\left(\frac{1}{\mathcal{F}'_{\mathcal{I}}}\right)\right] \leq \log_2(1 + \eta) + o(1)$ and hence

$$\mathbb{E}[\log_2(\mathcal{F}'_{\mathcal{I}})] \geq \log_2(\mu^2) - \log_2(1 + \eta) + o(1). \quad (44)$$

Consequently, the term $\mathbb{E}[\log_2(\mathcal{F}'_{\mathcal{I}})]$ can be bounded as

$$\begin{aligned} \log_2(\mu^2) - \log_2(1 + \eta) + o(1) \\ \leq \mathbb{E}[\log_2(\mathcal{F}'_{\mathcal{I}})] \leq \log_2(1 + \eta) + o(1). \quad (45) \end{aligned}$$

Since $\lim_{P \rightarrow \infty} \mu = 1$ and $\lim_{P \rightarrow \infty} \eta = 0$, it follows that

$$\lim_{P \rightarrow \infty} \mathbb{E}[\log_2(\mathcal{F}'_{\mathcal{I}})] = 0. \quad (46)$$

D. Analysis of the Received Signal Ratio ($\mathcal{F}_{\mathcal{D}}$)

It remains to prove that the first expectation in (35) also converges to zero. As for $\mathcal{F}_{\mathcal{I}}$, we can write

$$\begin{aligned} \mathbb{E}[\log_2(\mathcal{F}_{\mathcal{D}})] &\leq \log_2\left(\frac{1}{\mu^2}\right) \\ &\quad + \mathbb{E}\left[\log_2\left(\underbrace{\frac{1 + P \sum_{\ell \in \mathbb{N}_K} |\mathbf{h}_i \mathbf{v}_\ell|^2}{1 + P \sum_{\ell \in \mathbb{N}_K} |\mathbf{h}_i \mathbf{w}_\ell|^2}}_{\mathcal{F}'_{\mathcal{D}}}\right)\right]. \quad (47) \end{aligned}$$

Moreover, the equivalent to Lemma 3 also holds for $\mathcal{F}'_{\mathcal{D}}$.

Lemma 4. Let $\eta = \bar{P}^{-\varepsilon}$, with $\alpha_q > \varepsilon > 0$ and ε arbitrarily small. Then,

$$\Pr(\mathcal{F}'_{\mathcal{D}} \geq 1 + \eta) = o\left(\frac{1}{\log_2(P)}\right) \quad (48)$$

and

$$\Pr\left(\frac{1}{\mathcal{F}'_{\mathcal{D}}} \geq 1 + \eta\right) = o\left(\frac{1}{\log_2(P)}\right). \quad (49)$$

Proof. The proof is relegated to Appendix II-B. \square

Thus, applying the same step as in (40) yields

$$\mathbb{E}[\log_2(\mathcal{F}'_{\mathcal{D}})] \leq \log_2(1 + \eta) + o(1). \quad (50)$$

We can similarly lower-bound $\mathbb{E}[\log_2(\mathcal{F}_{\mathcal{D}})]$ to obtain that

$$\begin{aligned} -\log_2(1 + \eta) + o(1) &\leq \mathbb{E}[\log_2(\mathcal{F}_{\mathcal{D}})] \\ &\leq \log_2(1/\mu^2) + \log_2(1 + \eta) + o(1). \quad (51) \end{aligned}$$

The fact that $\lim_{P \rightarrow \infty} \mu = 1$ and $\lim_{P \rightarrow \infty} \eta = 0$ leads to

$$\lim_{P \rightarrow \infty} \mathbb{E}[\log_2(\mathcal{F}_{\mathcal{D}})] = 0. \quad (52)$$

E. Merging Previous Sections

Given that $\lim_{P \rightarrow \infty} \Delta R = \lim_{P \rightarrow \infty} \sum_{i=1}^K \Delta R_i$, we obtain that

$$\begin{aligned} \lim_{P \rightarrow \infty} \Delta R &= \lim_{P \rightarrow \infty} K(\mathbb{E}[\log_2(\mathcal{F}_{\mathcal{D}})] + \mathbb{E}[\log_2(\mathcal{F}_{\mathcal{I}})]) \\ &= 0, \quad (53) \end{aligned}$$

which concludes the proof of Theorem 1. \square

VI. NUMERICAL RESULTS

In this section, we provide some performance analysis for the previous asymptotic results. We consider a scenario in which the most-informed TX has a CSI precision scaling parameter $\alpha^{(1)} = 1$ for the whole channel matrix, and the rest of TXs have a CSI precision scaling parameter $\alpha^{(j)} = 0.6$, for any $j > 1$. Intuitively, this configuration can model a setting in which a main TX receives a quantized CSI feedback from all the RXs, and then it shares a compressed version of the CSI to the other auxiliary transmit antennas. We present the performance of several schemes:

- The ideal centralized CSIT setting, in which all the TXs are endowed with the CSI of TX 1.
- The CD-ZF scheme with Hierarchical CSIT (TX 1 knows the other TXs' CSI).
- The CD-ZF, AP-ZF, and Naive ZF schemes when the CSIT is non hierarchical (general D-CSIT setting).
- The performance of transmitting only from TX 1 and turning off the other TXs.

In Fig. 7, we show the rate performance for a setting with 2 single-antenna TXs and 2 RXs under the assumption of instantaneous power constraint for the precoder. Several insights emerge from the figure.

First, we observe how the proposed CD-ZF scheme performs almost as good as the ideal centralized CSIT setting for the Hierarchical CSIT configuration. This fact holds for any setting

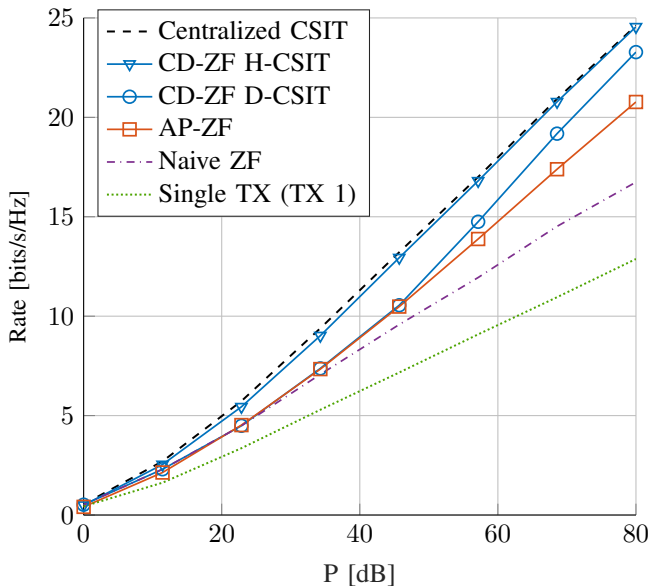


Fig. 7: Performance of different precoding schemes for a setting with 2 single-antenna TXs and 2 RXs with instantaneous power constraint and $\alpha^{(1)} = 1$, $\alpha^{(2)} = 0.6$.

configuration and size, yet considering that $N_1 \geq K - 1$. Besides this, the CD-ZF scheme is shown to tend towards the centralized rate also for the general D-CSIT setting, where the CSI at TX 2 is not available at TX 1. However, we can see how the convergence is slow, and at low SNR the CD-ZF scheme outperforms the single-TX transmission or the Naive ZF only by a slight gap. This is an aftermath of the scheme definition, as it is tailored for the asymptotic high-SNR regime. Indeed, the CD-ZF scheme performs in an almost optimal manner if TX 1 correctly estimates the CSI at the other TXs; however, the probability of correct estimation increases slowly. Thus, the performance at medium SNR is limited.

It is important to note that the CD-ZF scheme here presented is not optimized, as our objective was to show the asymptotic behavior. For example, we assume a scalar quantizer that independently quantizes every real and imaginary part of each channel coefficient. Considerably higher probabilities of consistency would be obtained if the quantization phase is optimized, as it can be seen in [57]. We could, for example, use vector quantization or consider more complex schemes such as the ones proposed in [39], [58]. Nevertheless, the aforementioned points show how important it is to provide the CSI with structure (or hierarchy), as it has been proven indispensable to boost the performance. Moreover, this CSI structure is sometimes given by the network configuration, such that it does not imply an extra aspect to develop.

Another point to be considered is that CD-ZF allows to obtain centralized performance with one informed antenna less than the single-TX transmission. This consideration can be seen in Fig. 7, as the single-TX transmission does not even achieve the centralized DoF.

VII. CONCLUSIONS

We have analyzed the achievable rate through linear precoding in a Network MISO setting with Distributed CSIT. We have developed an achievable scheme that asymptotically attains the achievable rate of an ideal centralized setting where every TX is endowed with the best estimate among all the TXs. For the case in which the CSIT is distributed following a hierarchical structure among TXs, the previous asymptotic insight is shown to be valid also at the moderate-SNR regime, where the performance obtained at both distributed and centralized settings are alike. This result implies that we are not only able to reach the DoF performance of the ideal centralized setting but also its *beamforming gain* when there is precise CSIT only at a subset of the transmit antennas. Thus, the performance degradation on account of the CSI mismatches between TXs can be overcome by a properly designed precoding scheme which is aware of the distributed nature of the setting. Besides this, it has been shown that the quantization of the information available at certain nodes is helpful as it facilitates the consistency of the decision at all the transmitters. This last result could be applied to a broad set of distributed problems, in which the trade-off between global consistency and local precision has not been deeply analyzed yet. Furthermore, this work fosters the question of whether we should *enforce* heterogeneous CSIT allocation when designing the CSI sharing mechanism.

APPENDIX I A USEFUL LEMMA

We present in the following a lemma that is key for the next proofs. The impact of the channel quantization on the precoder design can be asymptotically computed as shown below.

Lemma 5. *Let TX j , $2 \leq j \leq M$, quantize its CSIT with a scalar uniform quantizer with quantization step $q = \bar{P}^{-\alpha_q}$, $\alpha^{(M)} > \alpha_q > 0$. The naive precoder of Section IV-B at TX j for any $i \in \mathbb{N}_K$ satisfies*

$$\mathbb{E} [\| \mathbf{v}_{i,j} - \mathbf{w}_{i,j} \|] = \mathcal{O} (\bar{P}^{-\alpha_q}) \quad (54)$$

$$\mathbb{E} [\| \mathbf{v}_{i,j} - \mathbf{w}_{i,j} \|^2] = \mathcal{O} (P^{-\alpha_q}). \quad (55)$$

Proof. The proof is provided in Appendix VI. \square

Lemma 5 is based on error propagation properties of linear systems. Thus, it is expected to hold for a broad set of noisy estimation models whose error variance scales as P^{-a} for any $a > 0$. For example, it holds for the quantized feedback model of [42], in which random vector quantization is assumed and the number of quantization bits scales with P , as shown in [43]. Furthermore, Lemma 5 leads to the following corollary.

Corollary 3. *Let TX j , $2 \leq j \leq M$, quantize its CSIT with a scalar uniform quantizer with quantization step $q = \bar{P}^{-\alpha_q}$, $\alpha^{(M)} > \alpha_q > 0$. The global precoder of Section IV-B satisfies*

$$\mathbb{E} [\| \mathbf{v}_i - \mathbf{w}_i \|] = \mathcal{O} (\bar{P}^{-\alpha_q}) \quad (56)$$

$$\mathbb{E} [\| \mathbf{v}_i - \mathbf{w}_i \|^2] = \mathcal{O} (P^{-\alpha_q}). \quad (57)$$

Proof. The proof is relegated to Appendix VI. \square

APPENDIX II
PROOF OF LEMMA 3 AND LEMMA 4

In this section we prove Lemma 3 and Lemma 4, which are instrumental for the proof of Theorem 1.

A. Proof of Lemma 3

We aim to prove that, for any $\eta = \bar{P}^{-\varepsilon}$, with $\alpha_q > \varepsilon > 0$ and ε arbitrarily small, it holds that

$$\Pr(\mathcal{F}'_{\mathcal{I}} \geq 1 + \eta) = o\left(\frac{1}{\log_2(P)}\right) \quad (58)$$

and

$$\Pr\left(\frac{1}{\mathcal{F}'_{\mathcal{I}}} \geq 1 + \eta\right) = o\left(\frac{1}{\log_2(P)}\right). \quad (59)$$

We start by noting that \mathbf{h}_i can be written as $\mathbf{h}_i = \frac{1}{z^{(j)}}(\hat{\mathbf{h}}_i^{(j)} - z^{(j)} \delta_i^{(j)})$ from the definition of the estimate in (13). Let us introduce the notations $\hat{z}_{\text{inv}}^{(j)} = \frac{1}{z^{(j)}}$ and $z_n^{(j)} = \frac{z^{(j)}}{z^{(j)}}$. Hence, for any $i \neq \ell$, it follows that

$$\begin{aligned} |\mathbf{h}_i \mathbf{w}_\ell| &\stackrel{(a)}{=} |\hat{z}_{\text{inv}}^{(1)} \hat{\mathbf{h}}_i^{(1)} \mathbf{w}_\ell - z_n^{(1)} \delta_i^{(1)} \mathbf{w}_\ell| \\ &\stackrel{(b)}{=} |\hat{z}_{\text{inv}}^{(1)} \hat{\mathbf{h}}_i^{(1)} \mathbf{v}_\ell - z_n^{(1)} \delta_i^{(1)} \mathbf{w}_\ell + z_n^{(1)} \delta_i^{(1)} \mathbf{v}_\ell - z_n^{(1)} \delta_i^{(1)} \mathbf{v}_\ell| \\ &\stackrel{(c)}{=} |\mathbf{h}_i \mathbf{v}_\ell - z_n^{(1)} \delta_i^{(1)} (\mathbf{w}_\ell - \mathbf{v}_\ell)|, \end{aligned} \quad (60)$$

where (a) and (c) come from the D-CSIT model of (13) and (b) from the precoder definition in (25) since $\hat{\mathbf{h}}_i^{(1)} \mathbf{v}_\ell = \hat{\mathbf{h}}_i^{(1)} \mathbf{w}_\ell$. Let us define $\mathbf{d}_\ell^{\mathbf{w}, \mathbf{v}} \triangleq \mathbf{w}_\ell - \mathbf{v}_\ell$ for the sake of readability and space. Hence, from the triangular inequality it follows that

$$\begin{aligned} &\frac{1 + P \sum_{\ell \neq i} |\mathbf{h}_i \mathbf{w}_\ell|^2}{1 + P \sum_{\ell \neq i} |\mathbf{h}_i \mathbf{v}_\ell|^2} \\ &\leq 1 + \frac{P}{1 + P \sum_{\ell \neq i} |\mathbf{h}_i \mathbf{v}_\ell|^2} \sum_{\ell \neq i} \left(|z_n^{(1)} \delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{w}, \mathbf{v}}|^2 \right. \\ &\quad \left. + 2P |\mathbf{h}_i \mathbf{v}_\ell| |z_n^{(1)} \delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{w}, \mathbf{v}}| \right). \end{aligned} \quad (61)$$

Let us recall that

$$\Pr\left(\sum_{k=1}^K A_k \geq c\right) \leq \sum_{k=1}^K \Pr\left(A_k \geq \frac{c}{K}\right). \quad (62)$$

From (61) and (62), it follows that

$$\begin{aligned} &\Pr(\mathcal{F}'_{\mathcal{I}} \geq 1 + \eta) \\ &\leq \sum_{\ell \neq i} \Pr\left(\frac{P(|z_n^{(1)} \delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{w}, \mathbf{v}}|^2 + 2P |\mathbf{h}_i \mathbf{v}_\ell| |z_n^{(1)} \delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{w}, \mathbf{v}}|)}{1 + P \sum_{\ell \neq i} |\mathbf{h}_i \mathbf{v}_\ell|^2} \geq \frac{\eta}{K}\right) \\ &\leq \sum_{\ell \neq i} \Pr\left(\frac{|z_n^{(1)} \delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{w}, \mathbf{v}}|^2 + 2|\mathbf{h}_i \mathbf{v}_\ell| |z_n^{(1)} \delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{w}, \mathbf{v}}|}{|\mathbf{h}_i \mathbf{v}_\ell|^2} \geq \frac{\eta}{K}\right) \\ &\stackrel{(a)}{=} (K-1) \Pr\left(\frac{|z_n^{(1)} \delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{w}, \mathbf{v}}|^2 + 2|\mathbf{h}_i \mathbf{v}_\ell| |z_n^{(1)} \delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{w}, \mathbf{v}}|}{|\mathbf{h}_i \mathbf{v}_\ell|^2} \geq \frac{\eta}{K}\right) \\ &\stackrel{(b)}{\leq} (K-1) \left(\Pr\left(\frac{|z_n^{(1)} \delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{w}, \mathbf{v}}|^2}{|\mathbf{h}_i \mathbf{v}_\ell|^2} \geq \frac{\eta}{2K}\right) \right. \\ &\quad \left. + \Pr\left(\frac{2|\mathbf{h}_i \mathbf{v}_\ell| |z_n^{(1)} \delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{w}, \mathbf{v}}|}{|\mathbf{h}_i \mathbf{v}_\ell|^2} \geq \frac{\eta}{2K}\right) \right) \end{aligned}$$

$$\stackrel{(c)}{\leq} 2(K-1) \Pr\left(\frac{|z_n^{(1)} \delta_i^{(1)} (\mathbf{w}_\ell - \mathbf{v}_\ell)|}{|\mathbf{h}_i \mathbf{v}_\ell|} \geq \frac{\eta}{4K}\right) \quad (63)$$

where (a) comes from symmetry, (b) from (62), and (c) because $\eta < 1$ and $\mathbf{d}_\ell^{\mathbf{w}, \mathbf{v}} \triangleq \mathbf{w}_\ell - \mathbf{v}_\ell$. Let us now introduce a parameter $\gamma \in \mathbb{R}$. We can continue as

$$\begin{aligned} &\Pr\left(\frac{|z_n^{(1)} \delta_i^{(1)} (\mathbf{w}_\ell - \mathbf{v}_\ell)|}{|\mathbf{h}_i \mathbf{v}_\ell|} \geq \frac{\eta}{4K}\right) \\ &= \Pr\left(|\delta_i^{(1)} (\mathbf{w}_\ell - \mathbf{v}_\ell)| \geq \frac{\eta}{4K} |\delta_i^{(1)} \mathbf{v}_\ell|\right) \\ &\leq \Pr\left(|\delta_i^{(1)} \mathbf{v}_\ell| < \bar{P}^{-\gamma}\right) \\ &\quad + \int_{|\delta_i^{(1)} \mathbf{v}_\ell| \geq \bar{P}^{-\gamma}} \frac{\mathbb{E}\left[|\delta_i^{(1)} (\mathbf{w}_\ell - \mathbf{v}_\ell)|\right]}{\frac{\eta}{4K} y} f_{|\delta_i^{(1)} \mathbf{v}_\ell|}(y) dy \end{aligned} \quad (64)$$

where the first equality comes from the fact that $|\mathbf{h}_i \mathbf{v}_\ell| = z_n^{(1)} |\delta_i^{(1)} \mathbf{v}_\ell|$, and the last inequality from the Law of Total Probability and Markov's Inequality. $f_{|\delta_i^{(1)} \mathbf{v}_\ell|}$ stands for the probability density function of $|\delta_i^{(1)} \mathbf{v}_\ell|$. Let us focus first on the first term of (64), $\Pr\left(|\delta_i^{(1)} \mathbf{v}_\ell| < \bar{P}^{-\gamma}\right)$, which satisfies the following proposition.

Proposition 1. *Let $\gamma > 0$. Then,*

$$\Pr\left(|\delta_i^{(1)} \mathbf{v}_\ell| < \bar{P}^{-\gamma}\right) = o\left(\frac{1}{\log_2(P)}\right). \quad (65)$$

Proof. The proof is relegated to Appendix IV. \square

On the other hand, the integral term of (64) can be rewritten as

$$\begin{aligned} &\int_{|\delta_i^{(1)} \mathbf{v}_\ell| \geq \bar{P}^{-\gamma}} \mathbb{E}\left[|\delta_i^{(1)} (\mathbf{w}_\ell - \mathbf{v}_\ell)|\right] \frac{f_{|\delta_i^{(1)} \mathbf{v}_\ell|}(y)}{\frac{\eta}{4K} y} dy \\ &= \frac{4K}{\eta} \mathbb{E}\left[|\delta_i^{(1)} (\mathbf{w}_\ell - \mathbf{v}_\ell)|\right] \mathbb{E}_{|\delta_i^{(1)} \mathbf{v}_\ell| \geq \bar{P}^{-\gamma}} \left[\frac{1}{|\delta_i^{(1)} \mathbf{v}_\ell|}\right] \\ &\leq \frac{4K}{\eta} \mathbb{E}\left[|\delta_i^{(1)} (\mathbf{w}_\ell - \mathbf{v}_\ell)|\right] \bar{P}^\gamma. \end{aligned} \quad (66)$$

Now, we introduce another useful proposition, whose proof is also relegated to Appendix IV.

Proposition 2. *It holds that*

$$\mathbb{E}\left[|\delta_i^{(1)} (\mathbf{w}_\ell - \mathbf{v}_\ell)|\right] = \mathcal{O}(\bar{P}^{-\alpha_q}). \quad (67)$$

By applying Proposition 1 and Proposition 2 into (64), it is straightforward to see that

$$\begin{aligned} \Pr(\mathcal{F}'_{\mathcal{I}} \geq 1 + \eta) &\leq o\left(\frac{1}{\log_2(P)}\right) \\ &\quad + \frac{8K(K-1)}{\eta} \mathcal{O}(\bar{P}^{-\alpha_q}) \bar{P}^\gamma. \end{aligned} \quad (68)$$

Since $\eta = \bar{P}^{-\varepsilon}$, with $\alpha_q > \varepsilon > 0$,

$$\Pr(\mathcal{F}'_{\mathcal{I}} \geq 1 + \eta) \leq o\left(\frac{1}{\log_2(P)}\right) + \bar{P}^\varepsilon \mathcal{O}(\bar{P}^{-\alpha_q}) \bar{P}^\gamma. \quad (69)$$

Let us select γ such that $\gamma > 0$ and $\varepsilon + \gamma - \alpha_q < 0$. Then, it follows that

$$\Pr(\mathcal{F}'_{\mathcal{I}} \geq 1 + \eta) = o\left(\frac{1}{\log_2(P)}\right), \quad (70)$$

which concludes the proof of the first statement of Lemma 3. We prove in the following the second statement, i.e.,

$$\Pr\left(\frac{1}{\mathcal{F}'_{\mathcal{I}}} \geq 1 + \eta\right) = o\left(\frac{1}{\log_2(P)}\right). \quad (71)$$

This is obtained by switching the vectors \mathbf{v}_ℓ and \mathbf{w}_ℓ and applying the same steps as in the proof of the first statement. To begin with, by following the steps in (63) we can easily obtain that

$$\begin{aligned} \Pr\left(\frac{1}{\mathcal{F}'_{\mathcal{I}}} \geq 1 + \eta\right) &= \Pr\left(\frac{1 + P \sum_{\ell \neq i} |\mathbf{h}_i \mathbf{v}_\ell|^2}{1 + P \sum_{\ell \neq i} |\mathbf{h}_i \mathbf{w}_\ell|^2} \geq 1 + \eta\right) \\ &\leq 2(K-1) \Pr\left(\frac{|z_n^{(1)} \boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell - \mathbf{v}_\ell)|}{|\mathbf{h}_i \mathbf{w}_\ell|} \geq \frac{\eta}{4K}\right). \end{aligned} \quad (72)$$

Furthermore, the final expression in (72) is equal to the one in (63) except for the fact that the denominator is $|\mathbf{h}_i \mathbf{w}_\ell|$ instead of $|\mathbf{h}_i \mathbf{v}_\ell|$. Hence, continuing as in (64)-(69), we obtain that

$$\Pr\left(\frac{1}{\mathcal{F}'_{\mathcal{I}}} \geq 1 + \eta\right) = o\left(\frac{1}{\log_2(P)}\right), \quad (73)$$

which concludes the proof of Lemma 3. \square

B. Proof of Lemma 4

We aim to prove that, for any $\eta = \bar{P}^{-\varepsilon}$, with $\alpha_q > \varepsilon > 0$ and ε arbitrarily small, it holds that

$$\Pr(\mathcal{F}'_{\mathcal{D}} \geq 1 + \eta) = o\left(\frac{1}{\log_2(P)}\right) \quad (74)$$

and

$$\Pr\left(\frac{1}{\mathcal{F}'_{\mathcal{D}'}} \geq 1 + \eta\right) = o\left(\frac{1}{\log_2(P)}\right). \quad (75)$$

Firstly, we focus on (74). Note that, applying similar steps as in (60), it holds that

$$\begin{aligned} |\mathbf{h}_i \mathbf{v}_\ell|^2 &\leq |\mathbf{h}_i \mathbf{w}_\ell|^2 + |z_n^{(1)} \boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell - \mathbf{v}_\ell)|^2 \\ &\quad + 2|\mathbf{h}_i \mathbf{w}_\ell| |z_n^{(1)} \boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell - \mathbf{v}_\ell)|, \end{aligned} \quad (76)$$

$$\begin{aligned} |\mathbf{h}_i \mathbf{v}_i|^2 &\leq |\mathbf{h}_i \mathbf{w}_i|^2 + |\mathbf{h}_i(\mathbf{w}_i - \mathbf{v}_i)|^2 \\ &\quad + 2|\mathbf{h}_i \mathbf{w}_i| |\mathbf{h}_i(\mathbf{w}_i - \mathbf{v}_i)|. \end{aligned} \quad (77)$$

Hence, following the steps applied in (61)-(63), we can write that

$$\begin{aligned} \Pr\left(\frac{1 + P \sum_{\ell \in \mathbb{N}_K} |\mathbf{h}_i \mathbf{v}_\ell|^2}{1 + P \sum_{\ell \in \mathbb{N}_K} |\mathbf{h}_i \mathbf{w}_\ell|^2} \geq 1 + \eta\right) &\leq \Pr(\mathcal{D}_1 + \mathcal{D}_2 + \mathcal{D}_3 + \mathcal{D}_4 \geq \eta) \quad (78) \\ &\leq \sum_{i=1}^4 \Pr\left(\mathcal{D}_i \geq \frac{\eta}{4}\right), \end{aligned}$$

where we have introduced the notations

$$\mathcal{D}_1 \triangleq \sum_{\ell \neq i} \frac{|z_n^{(1)} \boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell - \mathbf{v}_\ell)|^2}{|\mathbf{h}_i \mathbf{w}_i|^2} \quad (79)$$

$$\mathcal{D}_2 \triangleq \sum_{\ell \neq i} \frac{2|\mathbf{h}_i \mathbf{w}_\ell| |z_n^{(1)} \boldsymbol{\delta}_i^{(1)}(\mathbf{w}_i - \mathbf{v}_\ell)|}{|\mathbf{h}_i \mathbf{w}_i|^2} \quad (80)$$

$$\mathcal{D}_3 \triangleq \frac{|\mathbf{h}_i(\mathbf{w}_i - \mathbf{v}_i)|^2}{|\mathbf{h}_i \mathbf{w}_i|^2} \quad (81)$$

$$\mathcal{D}_4 \triangleq \frac{2|\mathbf{h}_i \mathbf{w}_i| |\mathbf{h}_i(\mathbf{w}_i - \mathbf{v}_i)|}{|\mathbf{h}_i \mathbf{w}_i|^2}. \quad (82)$$

The first inequality in (78) is obtained by applying (76)-(77) and eliminating the term $1 + P \sum_{\ell \neq i} |\mathbf{h}_i \mathbf{w}_\ell|^2$ from the denominator. From the analysis of $\mathcal{F}_{\mathcal{I}}$ in the previous section (see (64)), it follows easily that, if $\eta = \bar{P}^{-\varepsilon}$, with $\alpha_q > \varepsilon > 0$, then

$$\begin{aligned} \Pr\left(\mathcal{D}_1 \geq \frac{\eta}{4}\right) &\leq \sum_{\ell \neq i} \Pr\left(z_n^{(1)} \frac{|\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell - \mathbf{v}_\ell)|^2}{|\mathbf{h}_i \mathbf{w}_i|^2} \geq \frac{\eta}{4K}\right) \\ &= o\left(\frac{1}{\log_2(P)}\right). \end{aligned} \quad (83)$$

Similarly,

$$\begin{aligned} \Pr\left(\mathcal{D}_2 \geq \frac{\eta}{4}\right) &= \sum_{\ell \neq i} \Pr\left(z_n^{(1)} \frac{2|\mathbf{h}_i \mathbf{w}_\ell| |\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_i - \mathbf{v}_\ell)|}{|\mathbf{h}_i \mathbf{w}_i|^2} \geq \frac{\eta}{4K}\right) \\ &= o\left(\frac{1}{\log_2(P)}\right). \end{aligned} \quad (84)$$

For the two remaining terms, \mathcal{D}_3 and \mathcal{D}_4 , note that

$$\begin{aligned} \Pr\left(\mathcal{D}_3 \geq \frac{\eta}{4}\right) + \Pr\left(\mathcal{D}_4 \geq \frac{\eta}{4}\right) &= \Pr\left(\frac{|\mathbf{h}_i(\mathbf{w}_i - \mathbf{v}_i)|^2}{|\mathbf{h}_i \mathbf{w}_i|^2} \geq \frac{\eta}{4}\right) \\ &\quad + \Pr\left(\frac{2|\mathbf{h}_i(\mathbf{w}_i - \mathbf{v}_i)|}{|\mathbf{h}_i \mathbf{w}_i|} \geq \frac{\eta}{4}\right) \quad (85) \\ &\leq 2 \Pr\left(\frac{|\mathbf{h}_i(\mathbf{w}_i - \mathbf{v}_i)|}{|\mathbf{h}_i \mathbf{w}_i|} \geq \frac{\eta}{16}\right) \\ &= 2 \Pr\left(\frac{|\tilde{\mathbf{h}}_i(\mathbf{w}_i - \mathbf{v}_i)|}{|\tilde{\mathbf{h}}_i \mathbf{w}_i|} \geq \frac{\eta}{16}\right), \end{aligned}$$

where $\tilde{\mathbf{h}} = \frac{\mathbf{h}}{\|\mathbf{h}\|}$ is unit-norm and it is isotropically distributed on the N_T -dimensional unit-sphere [42]. We can continue as in (64) to write

$$\begin{aligned} \Pr\left(\frac{|\tilde{\mathbf{h}}_i(\mathbf{w}_i - \mathbf{v}_i)|}{|\tilde{\mathbf{h}}_i \mathbf{w}_i|} \geq \frac{\eta}{16}\right) &\leq \Pr\left(|\tilde{\mathbf{h}}_i \mathbf{w}_i| < \bar{P}^{-\gamma}\right) \\ &\quad + \int_{|\tilde{\mathbf{h}}_i \mathbf{w}_i| \geq \bar{P}^{-\gamma}} \frac{\mathbb{E}[|\tilde{\mathbf{h}}_i(\mathbf{w}_i - \mathbf{v}_i)|]}{\frac{\eta}{16} y} f_{|\tilde{\mathbf{h}}_i \mathbf{w}_i|}(y) dy \quad (86) \\ &\leq \mathcal{O}(\bar{P}^{-\gamma}) + 16\bar{P}^\varepsilon \mathbb{E}[|\tilde{\mathbf{h}}_i(\mathbf{w}_i - \mathbf{v}_i)|] \bar{P}^\gamma. \end{aligned}$$

The fact that $\|\tilde{\mathbf{h}}_i\| = 1$ implies that $\mathbb{E}[|\tilde{\mathbf{h}}_i(\mathbf{w}_i - \mathbf{v}_i)|] \leq \mathbb{E}[\|\mathbf{w}_i - \mathbf{v}_i\|]$. Moreover, Corollary 3 states that $\mathbb{E}[\|\mathbf{w}_i - \mathbf{v}_i\|] = \mathcal{O}(\bar{P}^{-\alpha_q})$. Consequently, by selecting

γ such that $\gamma > 0$ and $\varepsilon + \gamma - \alpha_q < 0$, it follows from (86) that

$$\Pr\left(\frac{|\tilde{\mathbf{h}}_i(\mathbf{w}_i - \mathbf{v}_i)|}{|\tilde{\mathbf{h}}_i \mathbf{w}_i|} \geq \frac{\eta}{16}\right) \leq \mathcal{O}(\bar{P}^{-\gamma}) + \bar{P}^\varepsilon \mathcal{O}(\bar{P}^{-\alpha_q}) \bar{P}^\gamma = o\left(\frac{1}{\log_2(P)}\right). \quad (87)$$

We introduce the result of (87) into (85) to obtain from (78) that

$$\Pr(\mathcal{F}'_{\mathcal{D}} \geq 1 + \eta) = o\left(\frac{1}{\log_2(P)}\right). \quad (88)$$

It would remain to prove that $\Pr\left(\frac{1}{\mathcal{F}'_{\mathcal{D}}} \geq 1 + \eta\right) = o\left(\frac{1}{\log_2(P)}\right)$. To do so, we just need to apply the same previous steps, but interchanging \mathbf{w} and \mathbf{v} . Following those steps and following a similar argument as in the proof for $\Pr\left(\frac{1}{\mathcal{F}'_{\mathcal{D}}}\right)$, we directly obtain the result. For this reason, and for the sake of concision, we omit the derivation. \square

APPENDIX III

PROOF OF LEMMA 2 (PROBABILITY OF POWER OUTAGE)

We denote the event of power outage as \mathcal{P}_o . Note that

$$\Pr(\mathcal{P}_o) \leq N_1 \Pr(\|\mathbf{T}_{1,1}\| > 1), \quad (89)$$

and $\mathbf{T}_{1,1} = \mu[\mathbf{w}_{1,1,1}, \mathbf{w}_{2,1,1}, \dots, \mathbf{w}_{K,1,1}]$, where $\mathbf{w}_{i,j,n}$ represents the n -th element of the precoding vector at TX j for the data symbols of RX i . Therefore,

$$\begin{aligned} \Pr(\mathcal{P}_o) &\leq N_1 \Pr(\|\mu[\mathbf{w}_{1,1,1}, \mathbf{w}_{2,1,1}, \dots, \mathbf{w}_{K,1,1}]\| > 1) \\ &\stackrel{(a)}{\leq} N_1 \Pr\left(\bigcup_{i \in \mathbb{N}_K} \|\mu \mathbf{w}_{i,1,1}\| > \|\mathbf{v}_{i,1,1}\|\right) \\ &\stackrel{(b)}{\leq} N_1 K \Pr(\|\mu \mathbf{w}_{1,1,1}\| > \|\mathbf{v}_{1,1,1}\|) \\ &\stackrel{(c)}{\leq} N_1 K \Pr(\mu \|\mathbf{v}_{1,1,1}\| + \mu \|\phi_1\| > \|\mathbf{v}_{1,1,1}\|) \\ &= N_1 K \Pr\left(\|\phi_1\| > \frac{1-\mu}{\mu} \|\mathbf{v}_{1,1,1}\|\right) \end{aligned} \quad (90)$$

where (a) is obtained from the precoder definition as $\|\mathbf{v}_{1,1,1} \dots \mathbf{v}_{K,1,1}\| \leq 1$, (b) follows because $\mathbf{w}_{i,1,1}$ (resp. $\mathbf{v}_{i,1,1}$) is equally distributed for any $i \in \mathbb{N}_K$, and (c) from (29). Now, we obtain the probability by conditioning on $\|\mathbf{v}_{1,1,1}\|$ and then averaging over the distribution of $\|\mathbf{v}_{1,1,1}\|$. Let us denote $\mu' \triangleq \frac{1-\mu}{\mu}$. Hence,

$$\Pr(\mathcal{P}_o) \leq N_1 K \int_{-\infty}^{\infty} \Pr(\|\phi_1\| > \mu' \nu) f_{\|\mathbf{v}_{1,1,1}\|}(\nu) d\nu. \quad (91)$$

Using Markov's inequality we obtain that

$$\begin{aligned} \Pr(\mathcal{P}_o) &\leq N_1 K \int_{-\infty}^{\infty} \frac{\mathbb{E}[\|\phi_1\|]}{\mu' \nu} f_{\|\mathbf{v}_{1,1,1}\|}(\nu) d\nu \\ &= N_1 K \mathbb{E}[\|\phi_1\|] \frac{1}{\mu'} \mathbb{E}[\|\mathbf{v}_{1,1,1}\|^{-1}], \end{aligned} \quad (92)$$

where $\mathbb{E}[\|\mathbf{v}_{1,1,1}\|^{-1}]$ exists from property (ZF2). Let us focus on the first expectation term of (92) ($\mathbb{E}[\|\phi_1\|]$). Recalling (29), ϕ_i is defined as

$$\phi_i = \hat{\mathbf{H}}_{i,1}^\dagger \hat{\mathbf{H}}_{i,\bar{1}} (\mathbf{v}_{i,\bar{1}} - \mathbf{w}_{i,\bar{1}}). \quad (93)$$

Then,

$$\begin{aligned} \mathbb{E}[\|\phi_i\|] &\stackrel{(a)}{\leq} \mathbb{E}\left[\|\hat{\mathbf{H}}_{i,1}^\dagger \hat{\mathbf{H}}_{i,\bar{1}}\| \|\mathbf{v}_{i,\bar{1}} - \mathbf{w}_{i,\bar{1}}\|\right] \\ &\stackrel{(b)}{\leq} \sqrt{\mathbb{E}\left[\|\hat{\mathbf{H}}_{i,1}^\dagger \hat{\mathbf{H}}_{i,\bar{1}}\|^2\right]} \mathbb{E}\left[\|\mathbf{v}_{i,\bar{1}} - \mathbf{w}_{i,\bar{1}}\|^2\right], \end{aligned} \quad (94)$$

where (a) comes from the sub-multiplicative property of the Frobenius norm and (b) from Cauchy-Schwarz inequality. Let us denote $g_m \triangleq \sqrt{\mathbb{E}\left[\|\hat{\mathbf{H}}_{i,1}^\dagger \hat{\mathbf{H}}_{i,\bar{1}}\|^2\right]}$, which is a value that does not depend on P since the channel estimates are equally distributed for any estimation error variance. Then, we have that

$$\mathbb{E}[\|\phi_i\|] \leq g_m \sqrt{\mathbb{E}\left[\|\mathbf{v}_{i,\bar{1}} - \mathbf{w}_{i,\bar{1}}\|^2\right]}. \quad (95)$$

Lemma 5 and the fact that $\mathbb{E}\left[\|\mathbf{v}_{i,\bar{1}} - \mathbf{w}_{i,\bar{1}}\|^2\right] = \sum_{j=2}^M \mathbb{E}\left[\|\mathbf{v}_{i,j} - \mathbf{w}_{i,j}\|^2\right]$ yield

$$\mathbb{E}[\|\phi_1\|] = \mathcal{O}(\bar{P}^{-\alpha_q}). \quad (96)$$

Since $\mu = 1 - \varpi$, with $\varpi = \Theta(\bar{P}^{-\alpha_\mu})$ and $\varpi > 0$, the term $\frac{1}{\mu'} = \frac{\mu}{1-\mu}$ satisfies $\frac{1}{\mu'} = \Theta(\bar{P}^{\alpha_\mu})$. From (ZF2), $\mathbb{E}[\|\mathbf{v}_{1,1,1}\|^{-1}]$ is $\Theta(1)$. Hence, recalling (92),

$$\begin{aligned} \Pr(\mathcal{P}_o) &\leq N_1 K \mathbb{E}[\|\mathbf{v}_{1,1,1}\|^{-1}] \mathbb{E}[\|\phi_1\|] \frac{1}{\mu'} \\ &= \Theta(1) \mathcal{O}(\bar{P}^{-\alpha_q}) \Theta(\bar{P}^{\alpha_\mu}), \end{aligned} \quad (97)$$

which implies that $\Pr(\mathcal{P}_o) = \mathcal{O}(\bar{P}^{\alpha_\mu - \alpha_q})$. By selecting $\alpha_\mu < \alpha_q$, the probability of power outage vanishes and it holds that

$$\Pr(\mathcal{P}_o) = o\left(\frac{1}{\log_2(P)}\right), \quad (98)$$

which concludes the proof. \square

APPENDIX IV

PROOF OF PROPOSITION 1 AND PROPOSITION 2

A. Proof of Proposition 1

We prove in the following that $\Pr(|\delta_i^{(1)} \mathbf{v}_\ell| < \bar{P}^{-\gamma}) = o\left(\frac{1}{\log_2(P)}\right)$ for any $\gamma > 0$ and $i, \ell \in \mathbb{N}_K$ such that $\ell \neq i$. Let us denote the precoder for RX ℓ obtained with perfect knowledge of \mathbf{H} as \mathbf{u}_ℓ . Then,

$$\begin{aligned} \Pr(|\delta_i^{(1)} \mathbf{v}_\ell| < \bar{P}^{-\gamma}) &= \Pr(|\delta_i^{(1)} \mathbf{u}_\ell + \delta_i^{(1)} (\mathbf{v}_\ell - \mathbf{u}_\ell)| < \bar{P}^{-\gamma}) \\ &\leq \Pr(|\delta_i^{(1)} \mathbf{u}_\ell| - |\delta_i^{(1)} (\mathbf{v}_\ell - \mathbf{u}_\ell)| < \bar{P}^{-\gamma}), \end{aligned} \quad (99)$$

where we have applied the inverse triangle inequality. In order to prove Proposition 1, we capitalize the intuition that the term $|\delta_i^{(1)} \mathbf{u}_\ell|$ is independent of the quality of the estimate and P , but the value of $|\delta_i^{(1)} (\mathbf{v}_\ell - \mathbf{u}_\ell)|$ is directly proportional to $P^{-\alpha^{(1)}}$. Before applying this intuition to (99), we first analyze the term $|\delta_i^{(1)} (\mathbf{v}_\ell - \mathbf{u}_\ell)|$ and the probability $\Pr(|\delta_i^{(1)} (\mathbf{v}_\ell - \mathbf{u}_\ell)| > \bar{P}^{-\beta})$ for any $\beta < \alpha^{(1)}$.

Proposition 3. For any $\beta < \alpha^{(1)}$, it holds that

$$\Pr(|\delta_i^{(1)} (\mathbf{v}_\ell - \mathbf{u}_\ell)| > \bar{P}^{-\beta}) = o\left(\frac{1}{\log_2(P)}\right). \quad (100)$$

Proof. It follows by the Cauchy-Schwarz inequality that $|\delta_i^{(1)} (\mathbf{v}_\ell - \mathbf{u}_\ell)| < \|\delta_i^{(1)}\| \|\mathbf{v}_\ell - \mathbf{u}_\ell\|$, which implies that

$\Pr(|\delta_i^{(1)}(\mathbf{v}_\ell - \mathbf{u}_\ell)| > \bar{P}^{-\beta}) \leq \Pr(\|\delta_i^{(1)}\| \|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta})$. Let us define the scalar $\varepsilon > 0$ such that $\beta < \beta + \varepsilon < \alpha^{(1)}$. By applying the law of total probability, we obtain that

$$\begin{aligned} \Pr(|\delta_i^{(1)}(\mathbf{v}_\ell - \mathbf{u}_\ell)| > \bar{P}^{-\beta}) &\leq \Pr(\|\delta_i^{(1)}\| \|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta}) \\ &\leq \Pr(\|\delta_i^{(1)}\| \|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta} \mid \|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta-\varepsilon}) \\ &\quad \times \Pr(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta-\varepsilon}) \\ &\quad + \Pr(\|\delta_i^{(1)}\| \|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta} \mid \|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon}) \\ &\quad \times \Pr(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon}) \\ &\leq \Pr(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta-\varepsilon}) \\ &\quad + \Pr(\|\delta_i^{(1)}\| > \bar{P}^\varepsilon \mid \|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon}). \end{aligned} \quad (101)$$

The term $\Pr(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta-\varepsilon})$ can be upper-bounded by means of the Markov's inequality, such that

$$\begin{aligned} \Pr(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta-\varepsilon}) &\leq \bar{P}^{\beta+\varepsilon} \mathbb{E}[\|\mathbf{v}_\ell - \mathbf{u}_\ell\|] \\ &\stackrel{(a)}{=} \mathcal{O}(\bar{P}^{\beta+\varepsilon-\alpha^{(1)}}) = o\left(\frac{1}{\log_2(P)}\right), \end{aligned} \quad (102)$$

where (a) follows after applying Lemma 5 to vectors whose respective input estimates differ by a $\mathcal{O}(\bar{P}^{-\alpha^{(1)}})$ additive error term. For the last term in (101), $\Pr(\|\delta_i^{(1)}\| > \bar{P}^\varepsilon \mid \|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon})$, it follows that

$$\begin{aligned} \Pr(\|\delta_i^{(1)}\| > \bar{P}^\varepsilon \mid \|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon}) \\ \leq \frac{\Pr(\|\delta_i^{(1)}\| > \bar{P}^\varepsilon)}{\Pr(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon})}. \end{aligned} \quad (103)$$

From (102), it holds that $\Pr(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon}) = 1 - \mathcal{O}(\bar{P}^{\beta+\varepsilon-\alpha^{(1)}})$. Besides this, $\|\delta_i^{(1)}\|^2 = \sum_{n=1}^{N_T} |\delta_{i,n}^{(1)}|^2$, where $\delta_{i,n}^{(1)}$ are i.i.d. as $\mathcal{N}_{\mathbb{C}}(0, 1)$. Consequently, $|\delta_{i,n}^{(1)}|^2$ is distributed following a Rayleigh distribution and

$$\|\delta_i^{(1)}\|^2 \sim \Gamma_d(N_T, 1), \quad (104)$$

where $\Gamma_d(N_T, 1)$ denotes the Gamma distribution. Moreover, $\Gamma_d(N_T, 1)$ is also called the Erlang distribution, and it satisfies that

$$\Pr(X \sim \Gamma_d(N_T, 1) < x) = 1 - \sum_{n=0}^{N_T-1} \frac{1}{n!} e^{-x} x^n. \quad (105)$$

Hence,

$$\begin{aligned} \Pr(\|\delta_i^{(1)}\| > \bar{P}^\varepsilon) &= \Pr(\|\delta_i^{(1)}\|^2 > \bar{P}^{2\varepsilon}) \\ &= \sum_{n=0}^{N_T-1} \frac{1}{n!} e^{-\bar{P}^{2\varepsilon}} \bar{P}^{2n\varepsilon}. \end{aligned} \quad (106)$$

Since $\varepsilon > 0$, $\Pr(\|\delta_i^{(1)}\| > \bar{P}^\varepsilon) = o(\bar{P}^x)$, for any $x \in \mathbb{R}$, and hence it is $o(1/\log_2(P))$. This implies that

$$\frac{\Pr(\|\delta_i^{(1)}\| > \bar{P}^\varepsilon)}{\Pr(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon})} = \frac{o\left(\frac{1}{\log_2(P)}\right)}{1 - \mathcal{O}(\bar{P}^{\beta+\varepsilon-\alpha^{(1)}})}, \quad (107)$$

which, together with (102) and (101), leads to

$$\Pr(|\delta_i^{(1)}(\mathbf{v}_\ell - \mathbf{u}_\ell)| > \bar{P}^{-\beta}) = o\left(\frac{1}{\log_2(P)}\right) \quad (108)$$

for any $\beta < \alpha^{(1)}$. \square

Equipped with this result from Proposition 3, now we focus back on (99). Before, let us introduce the notation $\mathbf{d}_\ell^{\mathbf{v}, \mathbf{u}} \triangleq \mathbf{v}_\ell - \mathbf{u}_\ell$ for the sake of readability. Then, (99) can be expanded by means of the Law of total probability as

$$\begin{aligned} &\Pr(|\delta_i^{(1)} \mathbf{u}_\ell| - |\delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{v}, \mathbf{u}}| < \bar{P}^{-\gamma}) \\ &= \Pr\left(|\delta_i^{(1)} \mathbf{u}_\ell| - |\delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{v}, \mathbf{u}}| < \bar{P}^{-\gamma} \mid |\delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{v}, \mathbf{u}}| \leq \bar{P}^{-\beta}\right) \\ &\quad \times \Pr(|\delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{v}, \mathbf{u}}| \leq \bar{P}^{-\beta}) \\ &\quad + \Pr\left(|\delta_i^{(1)} \mathbf{u}_\ell| - |\delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{v}, \mathbf{u}}| < \bar{P}^{-\gamma} \mid |\delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{v}, \mathbf{u}}| > \bar{P}^{-\beta}\right) \\ &\quad \times \Pr(|\delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{v}, \mathbf{u}}| > \bar{P}^{-\beta}) \\ &= \Pr\left(|\delta_i^{(1)} \mathbf{u}_\ell| - |\delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{v}, \mathbf{u}}| < \bar{P}^{-\gamma} \mid |\delta_i^{(1)} \mathbf{d}_\ell^{\mathbf{v}, \mathbf{u}}| \leq \bar{P}^{-\beta}\right) \\ &\quad + o\left(\frac{1}{\log_2(P)}\right) \\ &\leq \Pr(|\delta_i^{(1)} \mathbf{u}_\ell| < \bar{P}^{-\gamma} + \bar{P}^{-\beta}) + o\left(\frac{1}{\log_2(P)}\right). \end{aligned}$$

Let us assume w.l.o.g. that $\beta < \gamma$, such that $\Pr(|\delta_i^{(1)} \mathbf{u}_\ell| < \bar{P}^{-\gamma} + \bar{P}^{-\beta}) \leq \Pr(|\delta_i^{(1)} \mathbf{u}_\ell| < 2\bar{P}^{-\beta})$. Therefore, it remains to prove that $\Pr(|\delta_i^{(1)} \mathbf{u}_\ell| < 2\bar{P}^{-\beta}) = o\left(\frac{1}{\log_2(P)}\right)$. Let ϵ_β be a scalar such that $0 < \epsilon_\beta < \beta$ and let us define ψ as the angle satisfying

$$\cos(\psi) \triangleq \frac{|\delta_i^{(1)} \mathbf{u}_\ell|}{\|\delta_i^{(1)}\| \|\mathbf{u}_\ell\|}. \quad (109)$$

Then, we use again the Law of Total Probability to obtain

$$\begin{aligned} &\Pr(|\delta_i^{(1)} \mathbf{u}_\ell| < 2\bar{P}^{-\beta}) \\ &= \Pr\left(\|\delta_i^{(1)}\| \|\mathbf{u}_\ell\| \cos(\psi) < 2\bar{P}^{-\beta} \mid \|\mathbf{u}_\ell\| \leq \bar{P}^{-\epsilon_\beta}\right) \\ &\quad \times \Pr(\|\mathbf{u}_\ell\| \leq \bar{P}^{-\epsilon_\beta}) \\ &\quad + \Pr\left(\|\delta_i^{(1)}\| \|\mathbf{u}_\ell\| \cos(\psi) < 2\bar{P}^{-\beta} \mid \|\mathbf{u}_\ell\| > \bar{P}^{-\epsilon_\beta}\right) \\ &\quad \times \Pr(\|\mathbf{u}_\ell\| > \bar{P}^{-\epsilon_\beta}) \\ &\leq \Pr(\|\mathbf{u}_\ell\| \leq \bar{P}^{-\epsilon_\beta}) \\ &\quad + \Pr\left(\|\delta_i^{(1)}\| \|\mathbf{u}_\ell\| \cos(\psi) < 2\bar{P}^{-\beta} \mid \|\mathbf{u}_\ell\| > \bar{P}^{-\epsilon_\beta}\right) \\ &\leq \Pr(\|\mathbf{u}_\ell\| \leq \bar{P}^{-\epsilon_\beta}) \\ &\quad + \Pr\left(\|\delta_i^{(1)}\| \cos(\psi) < 2\bar{P}^{-\beta} \bar{P}^{\epsilon_\beta} \mid \|\mathbf{u}_\ell\| > \bar{P}^{-\epsilon_\beta}\right). \end{aligned} \quad (110)$$

Importantly, $\delta_i^{(1)}$ is isotropically distributed (i.e., the normalized value $\delta_i^{(1)}/\|\delta_i^{(1)}\|$ is uniformly distributed in the sphere surface). Besides this, \mathbf{u}_ℓ is a function of \mathbf{H} . Since \mathbf{H} and $\delta_i^{(1)}$ are mutually independent, so $\delta_i^{(1)}$ and \mathbf{u}_ℓ are. Hence, from isotropy of $\delta_i^{(1)}$, $\cos(\psi)$ is independent of \mathbf{u}_ℓ . On this basis, we can select $\mathbf{u}_\ell = [\mathbf{1}, \mathbf{0}_{1 \times N_T-1}]$ to obtain that

$$\begin{aligned} &\Pr\left(\|\delta_i^{(1)}\| \cos(\psi) < 2\bar{P}^{\epsilon_\beta-\beta} \mid \|\mathbf{u}_\ell\| > \bar{P}^{-\epsilon_\beta}\right) \\ &= \Pr\left(|\delta_{i,1,1}^{(1)}| < 2\bar{P}^{\epsilon_\beta-\beta}\right), \end{aligned} \quad (111)$$

where $\delta_{i,1,1}^{(1)}$ denotes the first element of the vector $\delta_i^{(1)}$, and it is distributed as $\mathcal{N}_{\mathbb{C}}(0, 1)$. Then,

$$\begin{aligned} \Pr\left(|\delta_{i,1,1}^{(1)}| < 2\bar{P}^{\epsilon_\beta - \beta}\right) &= \frac{2}{\sqrt{2\pi}} \int_0^{2\bar{P}^{\epsilon_\beta - \beta}} e^{-x^2/2} dx \\ &\leq \frac{4}{\sqrt{2\pi}} \bar{P}^{\epsilon_\beta - \beta}. \end{aligned} \quad (112)$$

On the other hand, the term $\Pr(\|\mathbf{u}_\ell\| \leq \bar{P}^{-\epsilon_\beta})$ is bounded by

$$\begin{aligned} \Pr(\|\mathbf{u}_\ell\| \leq \bar{P}^{-\epsilon_\beta}) &= \int_0^{\bar{P}^{-\epsilon_\beta}} f_{\|\mathbf{u}_i\|}(x) dx \\ &\leq f_{\|\mathbf{u}_i\|}^{\max} \bar{P}^{-\epsilon_\beta}, \end{aligned} \quad (113)$$

which follows from (ZF3). By introducing (112) and (113) in (110) we obtain that

$$\Pr\left(|\delta_i^{(1)} \mathbf{u}_\ell| < 2\bar{P}^{-\beta}\right) \leq \mathcal{O}(\bar{P}^{\max(-\epsilon_\beta, \epsilon_\beta - \beta)}). \quad (114)$$

Note that ϵ_β satisfies $0 < \epsilon_\beta < \beta$. Hence,

$$\begin{aligned} \Pr(|\delta_i^{(1)} \mathbf{v}_\ell| < \bar{P}^{-\gamma}) &\leq \Pr(|\delta_i^{(1)} \mathbf{u}_\ell| - |\delta_i^{(1)}(\mathbf{v}_\ell - \mathbf{u}_\ell)| < \bar{P}^{-\gamma}) \\ &\leq \Pr(|\delta_i^{(1)} \mathbf{u}_\ell| < 2\bar{P}^{-\beta}) + o\left(\frac{1}{\log_2(P)}\right) \\ &= o\left(\frac{1}{\log_2(P)}\right), \end{aligned} \quad (115)$$

which concludes the proof of Proposition 1. \square

B. Proof of Proposition 2

We prove in the following that $\mathbb{E}[|\delta_i^{(1)}(\mathbf{w}_\ell - \mathbf{v}_\ell)|] = \mathcal{O}(\bar{P}^{-\alpha_q})$ for any $i, \ell \in \mathbb{N}_K : \ell \neq i$. It follows that

$$\begin{aligned} \mathbb{E}[|\delta_i^{(1)}(\mathbf{w}_\ell - \mathbf{v}_\ell)|] &\leq \mathbb{E}[|\delta_i^{(1)}| \|\mathbf{w}_\ell - \mathbf{v}_\ell\|] \\ &= \text{cov}\left(\|\delta_i^{(1)}\|, \|\mathbf{w}_\ell - \mathbf{v}_\ell\|\right) + \mathbb{E}\left[\|\delta_i^{(1)}\|\right] \mathbb{E}\left[\|\mathbf{w}_\ell - \mathbf{v}_\ell\|\right] \\ &\leq \sqrt{\mathbb{E}\left[\|\delta_i^{(1)}\|^2\right]} \sigma_{\|\mathbf{w}_\ell - \mathbf{v}_\ell\|} + \mathbb{E}\left[\|\delta_i^{(1)}\|\right] \mathbb{E}\left[\|\mathbf{w}_\ell - \mathbf{v}_\ell\|\right], \end{aligned}$$

where $\text{cov}(X, Y) \triangleq \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$ is the covariance between X and Y and σ_X^2 represents the variance of the random variable X . The last inequality comes from the fact that $\text{cov}(x, y) \leq \sigma_x \sigma_y$ and $\sigma_x^2 \leq \mathbb{E}[x^2]$. Besides this, it holds from $\|\delta_i^{(1)}\|^2 \sim \Gamma_d(N_T, 1)$ that $\mathbb{E}\left[\|\delta_i^{(1)}\|^2\right] = N_T$. From this point and the fact that $\mathbb{E}[x] \leq \sqrt{\mathbb{E}[x^2]}$, we can write

$$\begin{aligned} \mathbb{E}[|\delta_i^{(1)}(\mathbf{w}_\ell - \mathbf{v}_\ell)|] &\leq \sqrt{N_T} (\sigma_{\|\mathbf{w}_\ell - \mathbf{v}_\ell\|} + \mathbb{E}\left[\|\mathbf{w}_\ell - \mathbf{v}_\ell\|\right]) \\ &\stackrel{(a)}{\leq} \sqrt{N_T} 2\sqrt{\mathbb{E}\left[\|\mathbf{w}_\ell - \mathbf{v}_\ell\|^2\right]} \\ &\stackrel{(b)}{=} \mathcal{O}(\bar{P}^{-\alpha_q}), \end{aligned}$$

where (a) comes from the fact that $\sigma_x + \mathbb{E}[x] \leq 2\sqrt{\mathbb{E}[x^2]}$ and (b) from Corollary 3. \square

APPENDIX V

PROOF OF LEMMA 1 (QUANTIZER CONSISTENCY)

Let $q \triangleq \bar{P}^{-\alpha_q}$ be the quantization step size of the quantizer \mathcal{Q}_u . Then, \mathcal{Q}_u is defined such that, for a scalar value $x \in \mathbb{R}$,

$$\mathcal{Q}_u(x) \triangleq q \left\lfloor \frac{x}{q} + \frac{1}{2} \right\rfloor. \quad (116)$$

We extend the notation for any complex matrix $\mathbf{A} \in \mathbb{C}^{n \times m}$ such that $\mathbf{A}_q = \mathcal{Q}_u(\mathbf{A})$ denotes the element-wise quantization, i.e.,

$$(\mathbf{A}_q)_{i,k} \triangleq \mathcal{Q}_u(\text{Re}(\mathbf{A}_{i,k})) + i\mathcal{Q}_u(\text{Im}(\mathbf{A}_{i,k})), \quad (117)$$

where $\text{Re}(x)$ and $\text{Im}(x)$ stand for the real imaginary part of $x \in \mathbb{C}$, and $i \triangleq \sqrt{-1}$. In this appendix we prove that, for a scalar uniform quantizer \mathcal{Q}_u with $q = \bar{P}^{-\alpha_q}$ and $\alpha^{(j)} > \alpha_q > 0$, $\forall j \in \mathbb{N}_M$, it follows that

$$\Pr\left(\hat{\mathbf{H}}_q^{(j) \leftarrow (1)} \neq \hat{\mathbf{H}}_q^{(j)}\right) = o\left(\frac{1}{\log_2(P)}\right), \quad (118)$$

where $\hat{\mathbf{H}}_q^{(j)} = \mathcal{Q}_u(\hat{\mathbf{H}}^{(j)})$ and $\hat{\mathbf{H}}_q^{(j) \leftarrow (1)}$ is the MAP estimator of $\hat{\mathbf{H}}_q^{(j)}$ given $\hat{\mathbf{H}}^{(1)}$. We start by noting that, by definition of the MAP estimator,

$$\Pr\left(\hat{\mathbf{H}}_q^{(j) \leftarrow (1)} \neq \hat{\mathbf{H}}_q^{(j)}\right) \leq \Pr\left(\mathcal{Q}_u(\hat{\mathbf{H}}^{(1)}) \neq \hat{\mathbf{H}}_q^{(j)}\right). \quad (119)$$

Since $\text{Re}(\hat{\mathbf{H}}_{i,k}^{(1)})$ and $\text{Im}(\hat{\mathbf{H}}_{i,k}^{(1)})$ are i.i.d. for any i, k , it follows that

$$\begin{aligned} \Pr\left(\mathcal{Q}_u(\hat{\mathbf{H}}^{(1)}) \neq \hat{\mathbf{H}}_q^{(j)}\right) &\leq 2KN_T \Pr\left(\mathcal{Q}_u(\text{Re}(\hat{\mathbf{H}}_{1,1}^{(1)})) \neq \mathcal{Q}_u(\text{Re}(\hat{\mathbf{H}}_{1,1}^{(j)}))\right), \end{aligned} \quad (120)$$

where we have selected w.l.o.g. the real part of the (1,1) channel element. Hence, it is sufficient to obtain the probability of disagreement for $\text{Re}(\hat{\mathbf{H}}_{1,1}^{(j)})$. For that purpose, we split each reconstruction level of the quantizer in two parts: The *edge* of the cell and the *center* of the cell. This is done in order to show that, as P increases, the probability of disagreement vanishes if $\hat{\mathbf{H}}_{1,1}^{(1)}$ is in the *center* of the quantization level and, besides this, the probability that $\hat{\mathbf{H}}_{1,1}^{(1)}$ is in the *edge* area also vanishes. We rigorously show it in the following. Before starting, we introduce the simplified notation $\mathbf{h}^{(j)} \triangleq \text{Re}(\hat{\mathbf{H}}_{1,1}^{(j)})$ to ease the readability. Accordingly, we also introduce the notation $\mathbf{h} \triangleq \text{Re}(\hat{\mathbf{H}}_{1,1})$ and $\delta \triangleq \text{Re}(\delta_{1,1}^{(j)})$ such that $\mathbf{h}^{(j)} = \check{z}^{(j)} \mathbf{h} + z^{(j)} \delta^{(j)}$, with $z^{(j)} = \bar{P}^{-\alpha^{(j)}}$ and $\check{z}^{(j)} \triangleq \sqrt{1 - (z^{(j)})^2}$. Furthermore, we recall the notations $\check{z}_{\text{inv}}^{(j)} \triangleq \frac{1}{\check{z}^{(j)}}$ and $z_n^{(j)} \triangleq \frac{z^{(j)}}{\check{z}^{(j)}}$, introduced in Appendix II.

A. Edge and center of the reconstruction level

Let ℓ_n be the n -th quantization level of \mathcal{Q}_u , $n \in \mathbb{Z}$, with $\ell_0 = 0$. Let us define L_n as the input interval that outputs ℓ_n , i.e.,

$$L_n \triangleq \{x \mid \mathcal{Q}_u(x) = \ell_n\}. \quad (121)$$

L_n has a range $[L_n^{\min}, L_n^{\max}]$ such that $|L_n| \triangleq L_n^{\max} - L_n^{\min} = \bar{P}^{-\alpha_q}$. We split L_n in two areas, the *edge* area E_n and the *center* area C_n , depicted in Fig. 8. The *edge* area is defined

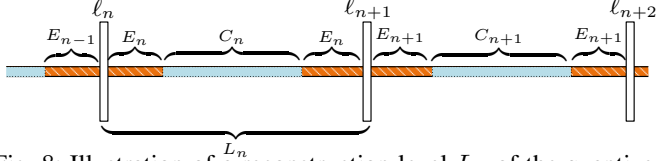


Fig. 8: Illustration of a reconstruction level L_n of the quantizer and the two sub-areas in which we divide it: The central area C_n and the edge area E_n .

as the part of L_n that is at most at distance $\bar{P}^{-c_e\alpha_q}$ of the boundary of the cell, with $c_e > 1$.

$$E_n \triangleq \{x \in L_n \mid x - L_n^{\min} < \bar{P}^{-c_e\alpha_q} \vee L_n^{\max} - x < \bar{P}^{-c_e\alpha_q}\}.$$

The center area is given by

$$C_n \triangleq \{x \in L_n \setminus E_n\}. \quad (122)$$

Intuitively, the probability of disagreement is very high if $h^{(1)}$ lies in the edge area E_n , whereas this probability vanishes in the central area C_n . Mathematically, we have that

$$\begin{aligned} \Pr\left(\mathcal{Q}_u(h^{(1)}) \neq \mathcal{Q}_u(h^{(j)})\right) &\leq \Pr\left(h^{(1)} \in \bigcup_{n \in \mathbb{Z}} E_n\right) \\ &+ \Pr\left(\mathcal{Q}_u(h^{(1)}) \neq \mathcal{Q}_u(h^{(j)}) \mid h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n\right). \end{aligned} \quad (123)$$

Let us analyze separately the two probabilities in the right-hand side of (123).

B. Probability of belonging to the edge area

Consider an arbitrary quantization level ℓ_n . Let $f_{L_n}^{\max}$ be the maximum value of the pdf of $h^{(1)}$ in $L_n = \{x \mid \mathcal{Q}_u(x) = \ell_n\}$. It follows that the probability that $h^{(1)}$ is in E_n is upper-bounded by

$$\begin{aligned} \Pr(h^{(1)} \in E_n) &\leq f_{L_n}^{\max} |E_n| \\ &= 2f_{L_n}^{\max} \bar{P}^{-c_e\alpha_q}, \end{aligned} \quad (124)$$

where $|E_n|$ denotes the length of E_n . The standard normal distribution has a derivative that is, at most, $1/\sqrt{2\pi e}$. Thus, the probability of being in L_n satisfies

$$\begin{aligned} \Pr(h^{(1)} \in L_n) &\geq (f_{L_n}^{\max} - 1/\sqrt{2\pi e} |L_n|) |L_n| \\ &= (f_{L_n}^{\max} - 1/\sqrt{2\pi e} \bar{P}^{-\alpha_q}) \bar{P}^{-\alpha_q}. \end{aligned} \quad (125)$$

Hence, the probability that $h^{(1)}$ is in E_n , given that it is in L_n , satisfies for any n that

$$\begin{aligned} \Pr(h^{(1)} \in E_n \mid L_n) &= \frac{\Pr(h^{(1)} \in E_n)}{\Pr(h^{(1)} \in L_n)} \\ &\leq \frac{2f_{L_n}^{\max}}{(f_{L_n}^{\max} - 1/\sqrt{2\pi e} \bar{P}^{-\alpha_q})} \bar{P}^{-(c_e-1)\alpha_q}. \end{aligned} \quad (126)$$

Let us define g_{\max} as $g_{\max} \triangleq \max_{n \in \mathbb{Z}} \frac{2f_{L_n}^{\max}}{(f_{L_n}^{\max} - 1/\sqrt{2\pi e} \bar{P}^{-\alpha_q})}$. Note that $g_{\max} = \Theta(1)$. Hence, from (126) and the fact that $\sum_{n \in \mathbb{Z}} \Pr(h^{(1)} \in L_n) = 1$, we can write

$$\begin{aligned} \Pr(h^{(1)} \in \bigcup_{n \in \mathbb{Z}} E_n) &= \sum_{n \in \mathbb{Z}} \frac{\Pr(h^{(1)} \in E_n)}{\Pr(h^{(1)} \in L_n)} \Pr(h^{(1)} \in L_n) \\ &\leq g_{\max} \bar{P}^{-(c_e-1)\alpha_q} \sum_{n \in \mathbb{Z}} \Pr(h^{(1)} \in L_n) \\ &= \mathcal{O}(\bar{P}^{-(c_e-1)\alpha_q}). \end{aligned} \quad (127)$$

Consequently, it holds that

$$\Pr(h^{(1)} \in \bigcup_{n \in \mathbb{Z}} E_n) = o\left(\frac{1}{\log_2(P)}\right). \quad (128)$$

C. Probability of disagreement in the center area

From the fact that the minimum distance from any point of C_n to the border of L_n is $\bar{P}^{-c_e\alpha_q}$, it holds that

$$\begin{aligned} \Pr\left(\mathcal{Q}_u(h^{(1)}) \neq \mathcal{Q}_u(h^{(j)}) \mid h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n\right) \\ \leq \Pr\left(|h^{(1)} - h^{(j)}| \geq \bar{P}^{-c_e\alpha_q} \mid h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n\right). \end{aligned} \quad (129)$$

Given that, for two events A, C , $\Pr(A \mid C) \leq \Pr(A)/\Pr(C)$, it follows that

$$\begin{aligned} \Pr\left(|h^{(1)} - h^{(j)}| \geq \bar{P}^{-c_e\alpha_q} \mid h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n\right) \\ \leq \frac{1}{\Pr(h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n)} \Pr\left(|h^{(1)} - h^{(j)}| \geq \bar{P}^{-c_e\alpha_q}\right) \\ \leq \frac{1}{\Pr(h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n)} \frac{\mathbb{E}[|h^{(1)} - h^{(j)}|]}{\bar{P}^{-c_e\alpha_q}}, \end{aligned} \quad (130)$$

which comes from Markov's Inequality. In the following, we obtain the expectation $\mathbb{E}[|h^{(1)} - h^{(j)}|]$. Then,

$$h^{(1)} - h^{(j)} = h(\check{z}^{(1)} - \check{z}^{(j)}) + (z^{(1)}\delta^{(1)} - z^{(j)}\delta^{(j)}). \quad (131)$$

From the assumption of Gaussian variables, it holds that

$$h(\check{z}^{(1)} - \check{z}^{(j)}) \sim \mathcal{N}\left(0, (\check{z}^{(1)} - \check{z}^{(j)})^2\right), \quad (132)$$

$$z^{(1)}\delta^{(1)} - z^{(j)}\delta^{(j)} \sim \mathcal{N}\left(0, (z^{(1)})^2 + (z^{(j)})^2\right). \quad (133)$$

Since $h^{(1)}$ is independent of $\delta^{(1)}$ and $\delta^{(j)}$, it follows that

$$h^{(1)} - h^{(j)} \sim \mathcal{N}\left(0, \sigma_d^2\right), \quad (134)$$

where σ_d^2 is given by

$$\sigma_d^2 = (\check{z}^{(1)} - \check{z}^{(j)})^2 + (z^{(1)})^2 + (z^{(j)})^2. \quad (135)$$

Substituting the variables for their values yields

$$\sigma_d^2 = 2\left(1 - \sqrt{1 - P^{-\alpha^{(1)}} - P^{-\alpha^{(j)}} + P^{-\alpha^{(1)} - \alpha^{(j)}}}\right). \quad (136)$$

Furthermore, if $h^{(1)} - h^{(j)}$ is drawn from a zero-mean Normal distribution of variance σ_d^2 , $|h^{(1)} - h^{(j)}|$ is distributed as a half-normal distribution of mean

$$\mathbb{E}[|h^{(1)} - h^{(j)}|] = \sigma_d \sqrt{\frac{2}{\pi}}. \quad (137)$$

From (136) and the fact that, for any x such that $0 \leq x \leq 1$, it holds that $\sqrt{1-x} \geq 1-x$, it follows that

$$\begin{aligned} \mathbb{E} \left[|h^{(1)} - h^{(j)}| \right] &\leq \sqrt{\frac{4}{\pi} (P^{-\alpha^{(1)}} + P^{-\alpha^{(j)}} - P^{-\alpha^{(1)} - \alpha^{(j)}})} \\ &= \mathcal{O}(\bar{P}^{-\alpha^{(j)}}). \end{aligned} \quad (138)$$

Besides this, it holds from (128) that

$$\begin{aligned} \Pr \left(h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n \right) &= 1 - \Pr \left(h^{(1)} \in \bigcup_{n \in \mathbb{Z}} E_n \right) \\ &= 1 - \mathcal{O}(\bar{P}^{-\alpha_q}). \end{aligned} \quad (139)$$

From (130), both (138) and (139) lead to

$$\begin{aligned} \Pr \left(\mathcal{Q}_u(h^{(1)}) \neq \mathcal{Q}_u(h^{(j)}) \mid h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n \right) &\leq \frac{1}{\Pr \left(h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n \right)} \frac{\mathbb{E} [|h^{(1)} - h^{(j)}|]}{\bar{P}^{-c_e \alpha_q}} \\ &= \mathcal{O} \left(\bar{P}^{c_e \alpha_q - \alpha^{(j)}} \right) \\ &= o \left(\frac{1}{\log_2(P)} \right). \end{aligned} \quad (140)$$

The last inequality is obtained only if $c_e \alpha_q < \alpha^{(j)}$ for any $j \in \mathbb{N}_M$. Thus, it follows from (140) that it is necessary to satisfy that $c_e \alpha_q < \alpha^{(j)}$ for any $j \in \mathbb{N}_M$. Since for any $\alpha_q < \alpha^{(j)}$ we can find a $c_e > 1$ such that $c_e \alpha_q < \alpha^{(j)}$, any $\alpha_q < \alpha^{(j)}$ will satisfy (140) as long as a correct c_e is selected.

D. Assembling probabilities

We make use of (128) and (140) to show that (123) satisfies

$$\begin{aligned} \Pr \left(\mathcal{Q}_u(h^{(1)}) \neq \mathcal{Q}_u(h^{(j)}) \right) &\leq \Pr \left(h^{(1)} \in \bigcup_{n \in \mathbb{Z}} E_n \right) \\ &\quad + \Pr \left(\mathcal{Q}_u(h^{(1)}) \neq \mathcal{Q}_u(h^{(j)}) \mid h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n \right) \\ &= o \left(\frac{1}{\log_2(P)} \right), \end{aligned} \quad (141)$$

which concludes the proof of Lemma 1. \square

APPENDIX VI

PROOF OF LEMMA 5 (ERROR ON NAIVE PRECODER)

In this appendix, we prove both Lemma 5 and Corollary 3. First, we focus on demonstrating that

$$\mathbb{E} [\|\mathbf{v}_{i,2} - \mathbf{w}_{i,2}\|^2] = \mathcal{O}(P^{-\alpha_q}). \quad (142)$$

The generalization of (142) for any $\mathbb{E} [\|\mathbf{v}_{i,k} - \mathbf{w}_{i,k}\|^2]$, $k \in \mathbb{N}_M \setminus 1$ is straightforward. Then, the proof of the other results in the lemma and the corollary will be shown to follow from (142).

In order to prove (142), we make use of the fact that, as presented in Section III-A, we assume that the precoding vectors and matrices can be expressed as a combination of summations, products, and generalized inverses of the channel estimate. Note that, with the previous operations, it is also possible to compute divisions and norms of the channel estimate coefficients.

First, note that both $\mathbf{w}_{i,2}$ and $\mathbf{v}_{i,2}$ are obtained following the same algorithm but based on different information (input). Specifically, $\mathbf{w}_{i,2}$ is computed on the basis of $\hat{\mathbf{H}}_q^{(2)} = \mathcal{Q}_u(\hat{\mathbf{H}}^{(2)})$, where \mathcal{Q}_u is a scalar uniform quantizer with quantization step $q = \bar{P}^{-\alpha_q}$, and $\mathbf{v}_{i,2}$ is computed on the basis of $\hat{\mathbf{H}}^{(1)}$. As in the previous appendix, let $h_q^{(2)}$ (resp. $h^{(j)}$ and h) denote the real or imaginary part of an arbitrary element of the matrix $\hat{\mathbf{H}}_q^{(2)}$ (resp. $\hat{\mathbf{H}}^{(j)}$ and \mathbf{H}). Let us define $h_\zeta^{(2)}$ as

$$h_\zeta^{(2)} \triangleq h^{(2)} + \zeta_q, \quad (143)$$

where ζ_q is distributed as a binary symmetric distribution with points $[-q, q]$, independent of the other variables, such that $\sigma_{\zeta_q}^2 = q^2$. Note that the error $h_q^{(2)} - h^{(1)}$ has smaller or equal variance than $h_\zeta^{(2)} - h^{(1)} = h^{(2)} - h^{(1)} + \zeta_q$. Hence, we can assume that $\mathbf{w}_{i,2}$ is computed on the basis of the estimate $h_\zeta^{(2)}$, since increasing the error variance can only hurt. Consequently, the error $\xi \triangleq h_\zeta^{(2)} - h^{(1)}$ has a variance σ_ξ^2 given by

$$\begin{aligned} \sigma_\xi^2 &\leq \sigma_d^2 + \sigma_{\zeta_q}^2 + 2\sigma_d\sigma_{\zeta_q} \\ &= \mathcal{O}(P^{-\alpha_q}), \end{aligned} \quad (144)$$

where σ_d^2 is given in (136). Therefore, we can write that

$$h_\zeta^{(2)} = h^{(1)} + \bar{P}^{-\alpha_q} \delta_\xi, \quad (145)$$

where δ_ξ is a variable of variance $\Theta(1)$ and bounded density. We continue by showing that the error-variance scaling remains being at most $\Theta(P^{-\alpha_q})$ after applying addition, product, inverse or pseudo-inverse operations. Afterward, based on those results, we prove (142).

A. Error in the addition

Let $a_\xi^{(2)}, b_\xi^{(2)}$, be distributed as (145), i.e., $a_\xi^{(2)} \triangleq a^{(1)} + \bar{P}^{-\alpha_q} \delta_\xi^a$, $b_\xi^{(2)} \triangleq b^{(1)} + \bar{P}^{-\alpha_q} \delta_\xi^b$, where $\delta_\xi^a, \delta_\xi^b$, are variables with variance $\Theta(1)$ and bounded density. It is easy to see that, for any $a_\xi^{(2)}, b_\xi^{(2)}$,

$$\begin{aligned} a_\xi^{(2)} + b_\xi^{(2)} &= a^{(1)} + \bar{P}^{-\alpha_q} \delta_\xi^a + b^{(1)} + \bar{P}^{-\alpha_q} \delta_\xi^b \\ &= a^{(1)} + b^{(1)} + \bar{P}^{-\alpha_q} (\delta_\xi^a + \delta_\xi^b). \end{aligned} \quad (146)$$

This implies that the error variance of the sum is also $\mathcal{O}(P^{-\alpha_q})$ as (144).

B. Error in the product

In a similar way, it follows that

$$\begin{aligned} a_\xi^{(2)} b_\xi^{(2)} &= \left(a^{(1)} + \bar{P}^{-\alpha_q} \delta_\xi^a \right) \left(b^{(1)} + \bar{P}^{-\alpha_q} \delta_\xi^b \right) \\ &= a^{(1)} b^{(1)} + \bar{P}^{-\alpha_q} \left(a^{(1)} \delta_\xi^b + b^{(1)} \delta_\xi^a + \bar{P}^{-\alpha_q} \delta_\xi^a \delta_\xi^b \right), \end{aligned} \quad (147)$$

which implies that the product also maintains the scaling of the variance as $\mathcal{O}(P^{-\alpha_q})$. Moreover, as the sum and product of matrices is a composition of sums and products of its coefficients, the result extends to any two matrices of suitable dimension.

C. Error in the inverse

Let us first assume that $\mathbf{A}_\xi^{(2)}$ and $\mathbf{A}^{(1)}$ are square matrices with full rank with probability one, and with coefficients following (145). We can then write that

$$\begin{aligned} (\mathbf{A}_\xi^{(2)})^{-1} &= \left(\mathbf{A}^{(1)} + \bar{P}^{-\alpha_q} \mathbf{\Delta}_\xi^{\mathbf{A}} \right)^{-1} \\ &= (\mathbf{A}^{(1)})^{-1} \\ &\quad - \bar{P}^{-\alpha_q} (\mathbf{A}^{(1)})^{-1} \mathbf{\Delta}_\xi^{\mathbf{A}} \left(\mathbf{A}^{(1)} + \bar{P}^{-\alpha_q} \mathbf{\Delta}_\xi^{\mathbf{A}} \right)^{-1} \end{aligned} \quad (148)$$

which is obtained from the Woodbury matrix identity [59]. Hence, the error variance of the inverse is again $\mathcal{O}(P^{-\alpha_q})$. Once that it is proved that the inverse operation generates an error with variance $\mathcal{O}(P^{-\alpha_q})$, we extend it for the Moore–Penrose inverse (pseudo-inverse) $(\mathbf{A}_\xi^{(2)})^\dagger$. We assume (as throughout the rest of the document) that each sub-matrix has maximum rank, i.e.,

$$\text{rank} \left(\mathbf{A}_\xi^{(2)} \in \mathbb{C}^{N \times M} \right) = \min(N, M). \quad (149)$$

Let us assume that $\mathbf{A}_\xi^{(2)}$ is full row-rank matrix, i.e., $N \leq M$. Under this assumption, the pseudo-inverse is given by

$$(\mathbf{A}_\xi^{(2)})^\dagger = (\mathbf{A}_\xi^{(2)})^{\text{H}} \left(\mathbf{A}_\xi^{(2)} (\mathbf{A}_\xi^{(2)})^{\text{H}} \right)^{-1}. \quad (150)$$

The case in which $\mathbf{A}_\xi^{(2)}$ is full column-rank matrix ($N \geq M$) will follow the same steps and thus we omit it. It follows from (147) that $\mathbf{A}_\xi^{(2)} (\mathbf{A}_\xi^{(2)})^{\text{H}} = \mathbf{A}^{(1)} (\mathbf{A}^{(1)})^{\text{H}} + \bar{P}^{-\alpha_q} \mathbf{\Delta}_{\text{eq}}$, where $\mathbf{\Delta}_{\text{eq}}$ has variance $\Theta(1)$. This, together with (148), implies that

$$\left(\mathbf{A}_\xi^{(2)} (\mathbf{A}_\xi^{(2)})^{\text{H}} \right)^{-1} = \left(\mathbf{A}^{(1)} (\mathbf{A}^{(1)})^{\text{H}} \right)^{-1} + \bar{P}^{-\alpha_q} \mathbf{\Delta}'_{\text{eq}}, \quad (151)$$

and, by applying again (147), it holds that

$$(\mathbf{A}_\xi^{(2)})^\dagger = (\mathbf{A}^{(1)})^{\text{H}} \left(\mathbf{A}^{(1)} (\mathbf{A}^{(1)})^{\text{H}} \right)^{-1} + \bar{P}^{-\alpha_q} \mathbf{\Delta}''_{\text{eq}}, \quad (152)$$

where $\mathbf{\Delta}'_{\text{eq}}$ and $\mathbf{\Delta}''_{\text{eq}}$ have variance $\Theta(1)$. As explained in [53], under the assumption that $\hat{\mathbf{X}}$ is a full row-rank matrix, any generalized inverse may be expressed as $\hat{\mathbf{X}}^- = \hat{\mathbf{X}}^\dagger + \mathbf{P}_\perp \mathbf{U}$, where $\mathbf{P}_\perp = \mathbf{I} - \hat{\mathbf{X}}^\dagger \hat{\mathbf{X}}$ is the orthogonal projection onto the null space of $\hat{\mathbf{X}}$ and \mathbf{U} is an arbitrary matrix. Hence, a similar result could be obtained for a broad set of generalized inverses.

D. Error variance of the difference of precoders

The centralized precoder $\mathbf{v}_{i,2}$ is based on $\hat{\mathbf{H}}^{(1)}$, i.e., $\mathbf{V} = \mathcal{V}(\hat{\mathbf{H}}^{(1)})$. The distributed precoder at TX 2 is based on its own quantized CSIT $\hat{\mathbf{H}}_q^{(2)}$, and thus $\mathbf{w}_{i,2}$ is obtained from $\mathbf{W} = \mathcal{V}(\hat{\mathbf{H}}_q^{(2)})$. Based on the previous results and the definition of linear precoders, it follows that we can write the distributed precoder based on the CSIT of TX 2 ($\hat{\mathbf{H}}_q^{(2)}$) as

$$\mathbf{w}_{i,2} = \mathbf{v}_{i,2} + \bar{P}^{-\alpha_q} \mathbf{e}_{\mathbf{w}}, \quad (153)$$

where $\mathbb{E} [\|\mathbf{e}_{\mathbf{w}}\|^2] = \mathcal{O}(1)$. Consequently,

$$\begin{aligned} \mathbb{E} [\|\mathbf{v}_{i,2} - \mathbf{w}_{i,2}\|^2] &= \mathbb{E} [\|\mathbf{v}_{i,2} - (\mathbf{v}_{i,2} + \bar{P}^{-\alpha_q} \mathbf{e}_{\mathbf{w}})\|^2] \\ &= P^{-\alpha_q} \mathbb{E} [\|\mathbf{e}_{\mathbf{w}}\|^2] \\ &= \mathcal{O}(P^{-\alpha_q}), \end{aligned} \quad (154)$$

which concludes the proof of (142) and thus of (55). Moreover, since $\mathbb{E} [\|\mathbf{x}\|] \leq \sqrt{\mathbb{E} [\|\mathbf{x}\|^2]}$, it follows that

$$\mathbb{E} [\|\mathbf{v}_{i,j} - \mathbf{w}_{i,j}\|] = \mathcal{O}(\bar{P}^{-\alpha_q}), \quad (155)$$

which concludes the proof of Lemma 5. \square

E. Proof of Corollary 3

In order to prove that $\mathbb{E} [\|\mathbf{v}_i - \mathbf{w}_i\|^2] = \mathcal{O}(P^{-\alpha_q})$ for any $i \in \mathbb{N}_K$, let us recall that the vector $\mathbf{w}_i - \mathbf{v}_i$ can be written as

$$\mathbf{w}_i - \mathbf{v}_i = \begin{bmatrix} \phi_i \\ \mathbf{w}_{i,\bar{1}} - \mathbf{v}_{i,\bar{1}} \end{bmatrix}, \quad (156)$$

where ϕ_i (defined in (29) as $\phi_i = \hat{\mathbf{H}}_{i,1}^\dagger \hat{\mathbf{H}}_{i,\bar{1}} (\mathbf{v}_{i,\bar{1}} - \mathbf{w}_{i,\bar{1}})$) is the difference at TX 1, and $\mathbf{w}_{i,\bar{1}} - \mathbf{v}_{i,\bar{1}}$ denotes the difference for the coefficients of all the TXs but TX 1, i.e.,

$$\mathbf{w}_{i,\bar{1}} - \mathbf{v}_{i,\bar{1}} = [(\mathbf{w}_{i,2} - \mathbf{v}_{i,2})^T, \dots, (\mathbf{w}_{i,K} - \mathbf{v}_{i,K})^T]^T.$$

Let us recall that $N_{\bar{1}}$ has been defined as $N_{\bar{1}} \triangleq N_T - N_1$, and let us further define $\mathbf{G}_{\mathbf{I}}$ as

$$\mathbf{G}_{\mathbf{I}} \triangleq \begin{bmatrix} \hat{\mathbf{H}}_{i,1}^\dagger \hat{\mathbf{H}}_{i,\bar{1}} \\ \mathbf{I}_{N_{\bar{1}}} \end{bmatrix}. \quad (157)$$

From the definition of ϕ_i , we can rewrite (156) as

$$\mathbf{w}_i - \mathbf{v}_i = \mathbf{G}_{\mathbf{I}} (\mathbf{w}_{i,\bar{1}} - \mathbf{v}_{i,\bar{1}}). \quad (158)$$

Generalizing (153) for any $j \in \{\mathbb{N}_K \setminus 1\}$, it holds that

$$\mathbf{w}_{i,j} = \mathbf{v}_{i,j} + \bar{P}^{-\alpha_q} \mathbf{e}_{\mathbf{w},j} \quad \forall j \in \{\mathbb{N}_K \setminus 1\}, \quad (159)$$

where $\mathbb{E} [\|\mathbf{e}_{\mathbf{w},j}\|^2] = \mathcal{O}(1)$. Then, upon defining $\mathbf{e}_{\mathbf{w},\bar{1}} = [\mathbf{e}_{\mathbf{w},2}^T, \dots, \mathbf{e}_{\mathbf{w},K}^T]^T$, it follows that

$$\begin{aligned} \mathbb{E} [\|\mathbf{w}_i - \mathbf{v}_i\|^2] &= \mathbb{E} [\|\mathbf{G}_{\mathbf{I}} (\mathbf{w}_{i,\bar{1}} - \mathbf{v}_{i,\bar{1}})\|^2] \\ &= P^{-\alpha_q} \mathbb{E} [\|\mathbf{G}_{\mathbf{I}} \mathbf{e}_{\mathbf{w},\bar{1}}\|^2] \\ &= \mathcal{O}(P^{-\alpha_q}) \end{aligned} \quad (160)$$

since $\mathbb{E} [\|\mathbf{G}_{\mathbf{I}} \mathbf{e}_{\mathbf{w},\bar{1}}\|^2] = \mathcal{O}(1)$, which proves (57). In order to prove (56), it follows from $\mathbb{E} [\|\mathbf{x}\|] \leq \sqrt{\mathbb{E} [\|\mathbf{x}\|^2]}$ that

$$\mathbb{E} [\|\mathbf{v}_i - \mathbf{w}_i\|] = \mathcal{O}(\bar{P}^{-\alpha_q}), \quad (161)$$

which concludes the proof of Corollary 3. \square

REFERENCES

- [1] M. Costa and A. E. Gamal, "The capacity region of the discrete memoryless interference channel with strong interference (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 33, no. 5, pp. 710–711, Sep. 1987.
- [2] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian Broadcast Channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.
- [3] A. Lozano, A. M. Tulino, and S. Verdú, "High-SNR power offset in multiantenna communication," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4134–4151, Dec 2005.
- [4] B. Hassibi and M. Sharif, "Fundamental Limits in MIMO Broadcast Channels," *IEEE J. Selected Areas in Commun. (JSAC)*, vol. 25, no. 7, pp. 1333–1344, Sep. 2007.
- [5] N. Jindal and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO Broadcast Channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1783–1794, May 2005.
- [6] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

- [7] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian Broadcast Channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.
- [8] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.
- [9] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.
- [10] M. Kountouris, R. de Francisco, D. Gesbert, D. T. M. Slock, and T. Salzer, "Multiuser diversity - multiplexing tradeoff in MIMO broadcast channels with limited feedback," in *Proc. IEEE Asilomar Conf. on Signals, Systems and Computers*, Oct 2006, pp. 364–368.
- [11] A. S. Motahari and A. K. Khandani, "Capacity Bounds for the Gaussian Interference Channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 620–643, Feb 2009.
- [12] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.
- [13] E. Piovano and B. Clerckx, "Optimal DoF Region of the K-User MISO BC With Partial CSIT," *IEEE Commun. Lett.*, vol. 21, no. 11, pp. 2368–2371, Nov 2017.
- [14] A. G. Davoodi and S. A. Jafar, "Aligned Image Sets Under Channel Uncertainty: Settling Conjectures on the Collapse of Degrees of Freedom Under Finite Precision CSIT," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5603–5618, Oct 2016.
- [15] R. Tandon, S. A. Jafar, S. Shamai (Shitz), and H. V. Poor, "On the synergistic benefits of alternating CSIT for the MISO BC," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4106–4128, 2013.
- [16] S. Lashgari, R. Tandon, and S. Avestimehr, "MISO Broadcast Channel With Hybrid CSIT: Beyond Two Users," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7056–7077, Dec 2016.
- [17] 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.
- [18] J. Xu, J. G. Andrews, and S. A. Jafar, "MISO Broadcast Channels with delayed finite-rate feedback: Predict or observe?" *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1456–1467, Apr. 2012.
- [19] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, "Degrees of freedom of time correlated MISO Broadcast Channel with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 315–328, Jan. 2013.
- [20] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr, "Approximate Capacity Region of the MISO Broadcast Channels With Delayed CSIT," *IEEE Trans. on Commun.*, vol. 64, no. 7, pp. 2913–2924, July 2016.
- [21] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai (Shitz), "Downlink multicell processing with limited-backhaul capacity," *EURASIP J. on Advances in Signal Process.*, May 2009.
- [22] M. Jaber, M. A. Imran, R. Tafazolli, and A. Tukmanov, "5G backhaul challenges and emerging research directions: A survey," *IEEE Access*, vol. 4, pp. 1743–1766, 2016.
- [23] P. Popovski, J. J. Nielsen, C. Stefanovic, E. d. Carvalho, E. Strom, K. F. Trillingsgaard, A. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sorensen, "Wireless access for ultra-reliable low-latency communication: Principles and building blocks," *IEEE Network*, vol. 32, no. 2, pp. 16–23, March 2018.
- [24] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, Oct 2018.
- [25] R. Radner, "Team decision problems," *The Annals of Mathematical Statistics*, 1962.
- [26] S. Grandhi, R. Vijayan, and D. Goodman, "Distributed power control in cellular radio systems," *IEEE Trans. on Commun.*, vol. 42, no. 234, pp. 226–228, Feb 1994.
- [27] B. L. Ng, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5491–5499, 2008.
- [28] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov 2010.
- [29] V. Aggarwal, Y. Liu, and A. Sabharwal, "Sum capacity of interference channels with a local view: Impact of distributed decisions," *IEEE Trans. on Inf. Theory*, vol. 58, no. 3, pp. 1630–1659, 2012.
- [30] A. Vahid and R. Calderbank, "Two-User Erasure Interference Channels With Local Delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 4910–4923, Sept 2016.
- [31] E. Lampiris, A. Bazco-Nogueras, and P. Elia, "Resolving the feedback bottleneck of multi-antenna coded caching," 2020. [Online]. Available: arxiv.org/abs/1811.03935
- [32] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of Caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.
- [33] J. Zhang and P. Elia, "Fundamental Limits of Cache-Aided Wireless BC: Interplay of Coded-Caching and CSIT Feedback," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 3142–3160, May 2017.
- [34] M. A. Hail, M. Amadeo, A. Molinaro, and S. Fischer, "Caching in named data networking for the wireless internet of things," in *Int. Conf. on Recent Advances in Internet of Things (RIoT)*, April 2015, pp. 1–6.
- [35] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, "A survey on mobile edge networks: Convergence of computing, caching and communications," *IEEE Access*, vol. 5, pp. 6757–6779, 2017.
- [36] 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.
- [37] A. Bazco-Nogueras, P. de Kerret, D. Gesbert, and N. Gresset, "Distributed CSIT does not reduce the Generalized DoF of the 2-user MISO Broadcast Channel," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 685–688, June 2019.
- [38] —, "On the Degrees-of-Freedom of the K-user distributed broadcast channel," *IEEE Trans. Inf. Theory*, vol. 66, no. 9, pp. 5642–5659, 2020.
- [39] P. de Kerret, A. Bazco, and D. Gesbert, "Enforcing coordination in network MIMO with unequal CSIT," in *Proc. IEEE Asilomar Conf. on Signals, Systems and Computers*, Nov 2016, pp. 39–43.
- [40] S. Shamai and S. Verdú, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1302–1327, May 2001.
- [41] J. Lee and N. Jindal, "High SNR analysis for MIMO Broadcast Channels: Dirty paper coding versus linear precoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4787 – 4792, Dec. 2007.
- [42] N. Jindal, "MIMO Broadcast Channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.
- [43] A. Bazco-Nogueras, L. Miretti, P. de Kerret, D. Gesbert, and N. Gresset, "Achieving vanishing rate loss in decentralized Network MIMO," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, July 2019, pp. 1457–1461.
- [44] K. T. Truong and R. W. Heath, "The viability of distributed antennas for massive MIMO systems," in *Proc. IEEE Asilomar Conf. on Signals, Systems and Computers*. IEEE, 2013, pp. 1318–1323.
- [45] D. E. Knuth, "Big Omicron and Big Omega and Big Theta," *SIGACT News*, vol. 8, no. 2, pp. 18–24, Apr. 1976.
- [46] G. Hardy, E. Wright, D. Heath-Brown, and J. Silverman, *An Introduction to the Theory of Numbers*, ser. Oxford mathematics. Oxford University Press, 2008.
- [47] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP Journal on Wireless Commun. and Networking*, vol. 2019, no. 1, pp. 1–13, 2019.
- [48] Z. H. Shaik, E. Björnson, and E. G. Larsson, "MMSE-optimal sequential processing for cell-free massive MIMO with radio stripes," *IEEE Trans. Commun.*, 2021.
- [49] L. Miretti, E. Björnson, and D. Gesbert, "Team MMSE precoding with applications to cell-free massive MIMO," *Submitted to IEEE Trans. Wireless Commun.*, May 2021, 2021.
- [50] A. G. Davoodi, B. Yuan, and S. A. Jafar, "GDoF Region of the MISO BC: Bridging the Gap Between Finite Precision and Perfect CSIT," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7208–7217, Nov 2018.
- [51] A. Bazco Nogueras, "Fundamental limits and algorithms in decentralized and cooperative wireless networks," Ph.D. dissertation, Sorbonne Université, 11 2019. [Online]. Available: <http://www.eurecom.fr/publication/6121>
- [52] G. Caire and K. R. Kumar, "Information theoretic foundations of adaptive coded modulation," *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2274–2298, 2007.
- [53] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, 2008.
- [54] C. Rao, "Generalized inverse of a matrix and its applications," in *Vol. 1 Theory of Statistics*. University of California Press, 1972, pp. 601–620.
- [55] C. T. K. Ng, D. Gunduz, A. J. Goldsmith, and E. Erkip, "Distortion minimization in Gaussian layered broadcast coding with successive Refinement," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5074–5086, 2009.
- [56] O. E. Ayach and R. W. Heath, "Interference alignment with analog channel state feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 626–636, Feb. 2012.

- [57] A. Bazco-Nogueras, L. Miretti, P. de Kerret, D. Gesbert, and N. Gresset, "Transmission robuste de zéro-forçage asymptotiquement optimale pour coopération imparfaite de transmetteurs," in *2019 GRETSI Colloque*, August 2019.
- [58] I. Atzeni and D. Gesbert, "Cooperative MIMO Precoding with Distributed CSI: A Hierarchical Approach," in *IEEE Int. Workshop on Signal Processing Advances in Wireless Commun. (SPAWC)*, June 2018, pp. 1–5.
- [59] H. V. Henderson and S. R. Searle, "On deriving the inverse of a sum of matrices," *Siam Review*, vol. 23, no. 1, pp. 53–60, Jan. 1981.

Antonio Bazco-Nogueras (M'20) received the B.S. and M.S. degrees in Telecommunications Engineering, both from University of Zaragoza, Spain, in 2014 and 2016, respectively. He obtained the Ph.D. degree from Sorbonne Université, Paris, France, in collaboration with the Mitsubishi Electric R&D Centre Europe, Rennes, France, in 2019. He was a post-doctoral researcher at EURECOM, Sophia-Antipolis, France, in 2020 and 2021. He is currently a post-doctoral researcher at IMDEA Networks Institute, Madrid, Spain. His research interests include multi-user information theory, intelligent networks, decentralized systems, content delivery networks, and cooperative wireless networks.

Paul de Kerret has an Engineering degree from both the French Graduate School, IMT Atlantique, and from the Munich University of Technology through a double-degree program. In 2013, he obtained his doctorate from the French Graduate School, Telecom Paris. From 2015 to 2019, he was a Senior Researcher at EURECOM. In September 2019, he joined the Mantu Artificial Intelligence Laboratory to lead the Research Group and he is currently working as Lead Data Scientist within the Startup Company Greenly where he puts his skills to measure and reduce our carbon footprints. He has been involved in several European collaborative projects on mobile communications and co-presented several tutorials at major IEEE international conferences. He has authored more than 30 articles in the IEEE flagship conferences.

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 Telecommunications, 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 340 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. In 2015, he was awarded 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. In 2020, he was awarded funding by the French Interdisciplinary Institute on Artificial Intelligence for a Chair in the area of AI for the future IoT.

Nicolas Gresset was born in France, in 1978. He received the Ph.D. degree from Institut Polytechnique de Paris, France, in collaboration with the Mitsubishi Electric R&D Centre Europe, Rennes, France, in 2004. He was a Post-Doctoral Researcher with Orange Labs, Issy-les-Moulineaux, France, in 2005. He was a Senior Engineer with Wipro-Newlogic and delegate to the IEEE 802.11n standardization, in 2006. In 2007, he joined the Mitsubishi Electric R&D Centre Europe, where he is currently the Research Manager of the Wireless Communication Systems Group focusing on research and standardization for industrial, automotive, railway and satellite networks.