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 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 $2 \times 2$ single-antenna scenario in which one TX has better knowledge of the full channel matrix that 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 DoF 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 $2 \times 2$ 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.

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.

C. Linear Approximation of Rate at High-SNR

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

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

(1)

where $P$ denotes the SNR, $\text{DoF}$ is the pre-logarithmic factor (or Degrees-of-Freedom), and $\mathcal{R}_\infty$ is the rate offset (or vertical offset). The approximation in (1) can also be written in terms of the power offset (horizontal offset) $\mathcal{L}_\infty$, where $\mathcal{R}_\infty = \text{DoF} \mathcal{L}_\infty$. An illustrative representation is shown in Fig. 1. The term $\mathcal{L}_\infty$ 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

$$\text{DoF} \triangleq \lim_{P \to \infty} \frac{R(P)}{\log_2(P)},$$

(2)

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

(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 $SNR^{-\alpha}$ in order to attain a DoF per user of $DoF_{RX_1} = \alpha$. Furthermore, having a CSIT error variance scaling with $SNR^{-\alpha}$ was shown to be equivalent to obtaining a quantized feedback of $\alpha \log_2(SNR)$ bits from the RX, which could be attained if the feedback resources scale proportionally to $\log_2(SNR)$.

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 $2 \times 2$ 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 $2 \times 2$ 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 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 ($N_i$) 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:** $N_N$ stands for $N_N \triangleq \{1, 2, \ldots, N\}$. We follow the asymptotic notation presented in [45], based on the prevalent Bachmann–Landau notation [46]. In consequence, if $f(x) = o(g(x))$ implies that $\lim_{x \to \infty} \frac{f(x)}{g(x)} = 0$, $f(x) = \Omega(g(x))$ implies that $\limsup_{x \to \infty} \frac{|f(x)|}{|g(x)|} < \infty$, and $f(x) = \Theta(g(x))$ implies that $\lim_{x \to \infty} \frac{|f(x)|}{|g(x)|} = c$, $0 < c < \infty$. 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 $X$. $A_{i,k}$ or $(A)_{i,k}$ represent the element of the matrix $A$ located in the $i$-th row and the $k$-th column. $I_n$ writes for the identity matrix of size $n \times n$. $0_{M \times N}$ (resp. $0_{M \times N}$) represents the all-zeroes (resp. all-ones) matrix of size $M \times N$. $\|A\|$ denotes the Frobenius norm of the matrix $A$.

### 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

$$y \triangleq \sqrt{P}HWs + n,$$  \tag{4}

where $P$ is the transmit power, $y \triangleq [y_1, \ldots, y_K]^T$ is the received signal vector, and $y_i$ is the received signal at RX $i$. $n$ stands for the Additive White Gaussian Noise (AWGN) distributed as $N_c(0, 1)$, where $N_c(0, \Gamma)$ stands for the circularly-symmetric complex normal distribution with covariance matrix (or variance) $\Gamma$. The vector $s \triangleq [s_1, \ldots, s_K]^T$ 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 $E[ss^H] = I$. The channel matrix is given by

$$H \triangleq \begin{bmatrix} h_1 \\ \vdots \\ h_K \end{bmatrix} \triangleq \begin{bmatrix} h_{1,1} & \ldots & h_{1,M} \\ \vdots & \ddots & \vdots \\ h_{K,1} & \ldots & h_{K,M} \end{bmatrix} \in \mathbb{C}^{K \times N_T}. \tag{5}$$

Hence, $h_i \in \mathbb{C}^{1 \times N_T}$ denotes the global channel vector towards RX $i$ and $h_{i,j} \in \mathbb{C}^{1 \times N_j}$ is the channel vector from TX $j$ to RX $i$. Note that we have defined the row vectors as $h_i$ and $h_{i,j}$ in place of the usual Hermitian notation $h_i^H$ and $h_{i,j}^H$. This is done to ease the notation for the remainder of the document. The channel coefficients are assumed to be i.i.d. as
where each TX must independently compute and agree on \(\mathcal{N}_C(0,1)\) such that all the channel sub-matrices are full rank with probability one. The precoding matrix is given by

\[
W \triangleq \begin{bmatrix}
T_1 \\
\vdots \\
T_M
\end{bmatrix} = \mu \begin{bmatrix}
w_1 & \ldots & w_K
\end{bmatrix}
\tag{6}
\]

\[
\triangleq \mu \begin{bmatrix}
w_{1,1} & \ldots & w_{K,1} \\
\vdots & \ddots & \vdots \\
w_{1,M} & \ldots & w_{K,M}
\end{bmatrix} \in \mathbb{C}^{N_T \times K}. \tag{7}
\]

Hence, \(T_j \in \mathbb{C}^{N_j \times K}\) 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 \(w_{i,j,n}\) as the coefficient at the \(n\)-th antenna of TX \(j\). The parameter \(\mu \leq 1\) is a power correction value that will be detailed afterwards. We define \(T_{j,n} \in \mathbb{C}^{1 \times K}\) as the precoding vector applied at the \(n\)-th antenna of TX \(j\) for the vector \(s\), with \(n \in \mathbb{N}_N\). We assume that the precoder has a per-antenna instantaneous unit-norm constraint, such that

\[
\|T_{j,n}\| \leq 1. \tag{8}
\]

The results presented here also hold under the assumption of per-TX instantaneous constraint \((\|T_j\| \leq 1)\). Note that, even if we set \(\|T_j\| = 1\), 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 \((\ref{eq:power_constraints})\) 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 \((\mathbb{E}[\|T_j\|^2] \leq 1)\) which has been assumed in other works \([27], [38]\). 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

\[
\hat{\mathbf{H}}^{(j)} \triangleq \sqrt{1 - Z^{(j)}} \mathbf{H} + \sqrt{Z^{(j)}} \Delta^{(j)}, \tag{9}
\]

where \(\Delta^{(j)} \triangleq [\delta_1^{(j)}, \ldots, \delta_K^{(j)}]^T\) is a random matrix that contains the additive estimation noise and whose elements are i.i.d. as \(\mathcal{N}_C(0,1)\). The matrix \(\Delta^{(j)}\) 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 \(\ref{eq:power_constraints}\), 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. \(\ref{fig:network_miso}\).

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.\footnote{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.} 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\cite{47,48,49}. 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\cite{43} for the $2 \times 2$ setting, in which a distributed adaptation of the Random Vector Quantization (RVQ) feedback model of\cite{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 $\alpha$\footnote{Moreover, it was shown in\cite{43} that this normalization parameter can be calculated with a less restricting precision without affecting the asymptotic performance.}.\footnote{Moreover, it was shown in\cite{43} that this normalization parameter can be calculated with a less restricting precision without affecting the asymptotic performance.} As our focus is on the high-SNR regime, we consider a high-SNR modeling in which the estimation error scales such that

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

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

$$1 \geq \alpha^{(1)} \geq \alpha^{(2)} \geq \cdots \geq \alpha^{(M)} \geq 0,$$

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

The reason for which the inequality $\alpha^{(1)} > \alpha^{(2)}$ is strict will be explained in Section V. 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

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

where $z^{(j)}_i \triangleq \bar{P}^{-\alpha^{(j)}}, z^{(j)}_i \triangleq \sqrt{1 - (z^{(j)}_i)^2}$, and $\hat{h}^{(j)}_i, \delta^{(j)}_i$, are the $i$-th rows of the matrices $\hat{H}^{(j)}, \Delta^{(j)}$, respectively. As stated before, the precision parameters $\alpha_M$ are assumed to be long-term coefficients that vary slowly. Based on that, it is assumed that every TX knows the full set $\alpha_M$, 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\cite{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 $\hat{H}$ and the estimate for the channel vector of RX $i$ as $\hat{h}_i$. We further denote the CSIT precision for the ideal centralized case as $\alpha^*$. Note that in the sorted setting, where $\alpha^{(1)} > \cdots > \alpha^{(M)}$, the ideal centralized setting is equivalent to a MISO Broadcast Channel with $N_T$ transmit antennas, CSIT $\hat{H}$ equal to $\hat{H}^{(1)}$, 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\cite{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 latter, although it was shown in\cite{38} that does not attain a greater DoF would benefit from the fact that the knowledge of the matched filters onto the null spaces of the interfered vectors has a bounded probability density function (pdf), i.e., that is neither predetermined nor constant. This last property follows from the assumption that \( \mathbf{h}_i \) and \( \mathbf{h}_j \) are drawn from Gaussian distributions and that the ZF precoding is a continuous map from estimates \( \mathbf{h}_i \) to vectors \( \mathbf{v}_i \). 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 inverses 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.,

\[
\mathbf{v}_i = \frac{\mathbf{P}_{\mathbf{h}_i^+} \mathbf{h}_i^H}{\|\mathbf{P}_{\mathbf{h}_i^+} \mathbf{h}_i^H\|}
\]

where \( \mathbf{P}_{\mathbf{h}_i^+} \) is defined as

\[
\mathbf{P}_{\mathbf{h}_i^+} \triangleq \left( \mathbf{I}_{N_T} - \mathbf{H}_i^H (\mathbf{H}_i^H \mathbf{H}_i^H)^{-1} \mathbf{H}_i \right),
\]

and where the matrix \( \mathbf{H}_i \) stands for the global channel matrix with the \( i \)-th row removed. Note that the inversion in\cite{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:

\[
\mathbf{v}_i \triangleq \mathcal{V}(\hat{\mathbf{H}}).
\]

**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 \( \{\mathbf{H}, \hat{\mathbf{H}}^{(1)}, \ldots, \hat{\mathbf{H}}^{(M)}\} \)\cite{52}. For the sake of readability, we omit hereinafter the reference to the joint process \( \{\mathbf{H}, \hat{\mathbf{H}}^{(1)}, \ldots, \hat{\mathbf{H}}^{(M)}\} \) 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

\[
r_i \triangleq \log_2 \left( 1 + \frac{P |\mathbf{h}_i^H t_i|^2}{1 + \sum_{j \neq i} P |\mathbf{h}_j^H t_j|^2} \right),
\]

where \( t_i \) 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\cite{1} for the rate of the Distributed CSIT setting. Let us denote the rate achieved in a Distributed CSIT setting characterized by \( \alpha_M \) as \( R(\alpha_M) \). Hence, we aim to find the values DoF\(_d\), \( \mathcal{R}_d^\infty \), such that

\[
R(\alpha_M) = \text{DoF}_d \log_2(P) - \mathcal{R}_d^\infty + o(1).
\]

**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\cite{1} 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 algorithm satisfying

1. \( \hat{\mathbf{h}}_i \mathbf{v}_i = 0, \quad \forall \ell \neq i \) (Zero-Forcing condition) (ZF1)
2. \( \mathbb{E} \left[ |\mathbf{v}_{i,j,n}|^{-1} \right] = \Theta(1) \) (ZF2)
3. \( f_{\|\mathbf{v}_i\|} \leq f_{\|\mathbf{v}_i\|}^{\max} < \infty \) (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 is neither predetermined nor constant. This last property follows from the assumption that \( \mathbf{h}_i \) and \( \mathbf{h}_j \) are drawn from Gaussian distributions and that the ZF precoding is a continuous map from estimates \( \mathbf{h}_i \) to vectors \( \mathbf{v}_i \). 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 inverses 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.,

\[
\mathbf{v}_i = \frac{\mathbf{P}_{\mathbf{h}_i^+} \mathbf{h}_i^H}{\|\mathbf{P}_{\mathbf{h}_i^+} \mathbf{h}_i^H\|}
\]

\( \mathbf{P}_{\mathbf{h}_i^+} \triangleq \left( \mathbf{I}_{N_T} - \mathbf{H}_i^H (\mathbf{H}_i^H \mathbf{H}_i^H)^{-1} \mathbf{H}_i \right),
\]

and where the matrix \( \mathbf{H}_i \) stands for the global channel matrix with the \( i \)-th row removed. Note that the inversion in\cite{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:

\[
\mathbf{v}_i \triangleq \mathcal{V}(\hat{\mathbf{H}}).
\]
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 (ZF) 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 \(\mathbb{E} \left[ \| T_j \|_2^2 \right] \leq 1\) for the 2 \(\times\) 2 setting \[37\] and for some regimes of the \(K \times K\) setting \[38\]. Under the instantaneous power constraint \(\| T_j,\alpha \|_2^2 \leq 1\) 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 an 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 \(T_1 = f(\hat{H}^{(1)}, \hat{g}(\hat{H}^{(2)})\), where \(f, g\) are generic functions. In a D-CSIT setting, TX 1 has no access to \(H^{(2)}\). Let us assume that TX 1 estimates \(\hat{g}(\hat{H}^{(2)})\) from \(\hat{H}^{(1)}\), as \(\hat{g}(\hat{H}^{(1)})\). Then, the computation is said to be Consistent if and only if \(\hat{g}(\hat{H}^{(1)}) = g(\hat{H}^{(2)})\). Otherwise, it is said to be Inconsistent.

**Definition 3** (Power Outage). Let \(T_j\) denote the linear precoding matrix computed at TX \(j\). For any precoder \(T_j\), 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 precode 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 V-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

\[
\Delta R \triangleq R^*(\alpha^{(1)}) - R(\alpha_M). \tag{19}
\]

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.,

\[
\lim_{P \to \infty} R^*(\alpha^{(1)}) - R(\alpha_M) = 0. \tag{20}
\]

**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.

**Corollary 1** (Rate-Offset under Distributed CSIT). It holds from Theorem 1 that the rate offset \(R_{\Delta,\alpha}^*(\alpha)\) 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 implies that it is possible to achieve not only the multiplexing gain but also the beamforming gain.

\footnote{We 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 \(H^{(2)}\) as a noisy version of \(H^{(1)}\), 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 \( \lim_{P \to \infty} R^*(\alpha^{(1)}) - R(\alpha_M) = \infty \). 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 \( \alpha^{(j)} \) 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.

\[ \Delta R = \Delta R_A \Pr(A) + \Delta R_{A^c} \Pr(A^c). \]

The transmission scheme has to be both feasible consistent \( (\Delta R_A = 0) \) and, for the consistent cases, it has to be precise \( (\Delta R_{A^c} = 0) \). 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 \( (\Pr(A) \to 0) \) but also that it attains consistency fast enough with respect to the CSI scaling, i.e., that \( \Delta R_{A^c} \Pr(A^c) \to 0 \).

As a matter of example, consider that the rate gap for the inconsistent cases \( (\Delta R_A \to 0) \) scales proportionally to \( \log(P) \) (as it will proven later on) such that \( \Delta R_{A^c} = \Theta(\log(P)) \). Then, the probability of inconsistency \( \Pr(A^c) \) approached to 0 at a convergence rate proportional to \( \frac{1}{\log(P)} \), 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 \( \Delta R_A \) in the consistent cases. This last condition is critical: As \( P \) increases, 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.
The characteristics of the quantizer \( Q \) 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 \( \hat{H}_{q}^{(j)} \)). Since the quantization transforms the continuous variable \( \hat{H}^{(j)} \) 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 \( \hat{H}_{q}^{(j)} \) based on its own information \( \hat{H}^{(j)} \), e.g. by computing the Maximum A Posteriori estimator (MAP) of \( \hat{H}_{q}^{(j)} \): 

\[
\hat{H}_{q}^{(j)\rightarrow(1)} = \arg \max_{\hat{H}_{q}^{(j)} \in Q(\hat{H}^{(j)})} \Pr(\hat{H}_{q}^{(j)} | \hat{H}^{(1)}).
\]

It is important to notice that the quantized value \( \hat{H}_{q}^{(j)} \) 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 

\[
\hat{H}_{q}^{(j)\rightarrow(1)} = \hat{H}_{q}^{(j)}.
\]
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 \( \mathbf{H}^{(1)} \). In order to provide some insights, we first describe the single-antenna case with 2 TXs and 2 RXs.

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

\[
|\mathbf{h}_{i,1}^{(1)} \mathbf{w}_{i,1} + \mathbf{h}_{i,2}^{(1)} \mathbf{w}_{i,2}| = |\mathbf{h}_{i,1}^{(1)} \mathbf{v}_{i,1} + \mathbf{h}_{i,2}^{(1)} \mathbf{v}_{i,2}|.
\]

(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 \( \mathbf{H}^{(2)} \), it knows \( \mathbf{w}_{i,2} \). Then, TX 1 computes its precoder such that

\[
\mathbf{w}_{i,1} = \mathbf{v}_{i,1} + (\mathbf{h}_{i,1}^{(1)})^\dagger (\mathbf{h}_{i,2}^{(1)} \mathbf{v}_{i,2} - \mathbf{w}_{i,2}),
\]

(26)

where \( (x)^\dagger \) denotes the pseudo-inverse of \( x \), which is known to have minimal Frobenius norm among all the generalized inverses [53]. The term \( \phi_i \) 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 \( i \in \mathbb{N}_K \),

\[
\sum_{\ell \in \mathbb{N}_N \setminus i} |\mathbf{h}_{i,1}^{(1)} \mathbf{w}_{\ell}|^2 = \sum_{\ell \in \mathbb{N}_N \setminus i} |\mathbf{h}_{i}^{(1)} \mathbf{v}_{\ell}|^2.
\]

(27)

The equality in (27) is attained if (but not only if) \( \mathbf{h}_{i,1}^{(1)} \mathbf{w}_{\ell} = \mathbf{h}_{i}^{(1)} \mathbf{v}_{\ell} \), 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 \( \mathbf{h}_{i,1} \) denote the sub-vector corresponding to the antennas of TX 1, whereas \( \mathbf{v}_{\ell,1}, \mathbf{w}_{\ell,1} \) and \( \mathbf{h}_{\ell,1}^{(1)} \) represent the sub-vector corresponding to the antennas from TX 2 to TX \( M \). The sub-matrices \( \mathbf{H}_{\ell,1} \) and \( \mathbf{H}_{\ell,1}^{(1)} \) 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 \( N_1 \equiv N_2 - N_1 \). We can expand the condition \( \mathbf{h}_{i,1}^{(1)} \mathbf{w}_{\ell} = \mathbf{h}_{i}^{(1)} \mathbf{v}_{\ell} \) as a matrix equation in which \( \mathbf{w}_{\ell,1} \) has to satisfy

\[
\mathbf{H}_{\ell,1} \mathbf{w}_{\ell,1} = \mathbf{H}_{\ell,1}^{(1)} \mathbf{v}_{\ell,1} + \mathbf{H}_{\ell,1}^{(1)} (\mathbf{v}_{\ell,1} - \mathbf{w}_{\ell,1}).
\]

(28)

We 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.

\( \phi_i \) could also represent the regularized pseudo-inverse.

**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 \( \mathbf{h}_i^{(1)} \mathbf{w}_i \) 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 \( n \in \mathbb{N}_N \), 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 \( Q \) is said to be Asymptotically Consistent if the probability of correct estimation of the MAP estimator at TX 1 satisfies

\[
\Pr \left( \mathbf{H}_j^{(j)}(\mathbf{w}_j) = \mathbf{H}_j' \right) = o \left( \frac{1}{\log_2(P)} \right), \quad \forall j \in \mathbb{N}_M.
\]

(P1)
Property \([\text{P1}]\) 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 \([\text{IV-A}]\). Clearly, it remains to prove that there exists some quantizer \(\mathcal{Q}_u\) that satisfies \([\text{P1}]\) 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 \(\mathcal{Q}_u(X)\) be a scalar uniform quantizer with quantization step \(q = \bar{P}/\alpha_q\), where \(\alpha_q\) is such that \(\alpha_q > 0\), for all \(j \in \mathbb{N}_M\). Then, \(\mathcal{Q}_u\) is an Asymptotically Consistent Quantizer and

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

**Proof.** The proof is relegated to Appendix \([\text{V}]\) \(\blacksquare\)

Note that \(\mathcal{Q}_u\) is a scalar quantizer. Thus, for a matrix \(A\), the notation \(A_q = \mathcal{Q}_u(A)\) denotes (with an abuse of notation) that \(A_q\) is composed of the independent scalar quantization of the real and imaginary part of each element in \(A\). Obviously, using vector quantization would improve the performance. However, as the proof of Theorem \([\text{I}]\) 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 \(P_o\) and recall that the precoder at the \(n\)-th antenna of TX \(j\) is expressed by \(\mathbf{T}_{j,n}\). In that case,

\[
P_o = \bigcup_{n \in \mathbb{N}_{N_j}, j \in \mathbb{N}_M} \{|\mathbf{T}_{j,n}| > 1\}. \tag{31}
\]

The power outage probability is handled by the parameter \(\mu\) of the precoding vector \(\mathbf{P} = \mathbf{C}^{N_T \times 1}\). The value of \(\mu\) 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 \(\varpi > 0\), \(\varpi = \Theta(\bar{P}/\alpha)\), and \(\alpha \mu \leq \alpha_q\). Then,

\[
\Pr (P_o) = o \left( \frac{1}{\log_2(P)} \right). \tag{P2}
\]

**Proof.** See Appendix \([\text{III}]\) \(\blacksquare\)

Similarly to property \([\text{P1}]\), property \([\text{P2}]\) in Lemma \([\text{II}]\) 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 \([\text{V}]\). Hereinafter, we assume that the uniform quantizer of Lemma \([\text{I}]\) is applied in the CD-ZF.

### C. Hierarchical CSIT Setting

Theorem \([\text{I}]\) 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 \([\text{IV-B}]\), 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 \(\mathbf{H}^{(j)}\), 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 \(\alpha(1) > \cdots > \alpha(M)\), TX \(j\) has access to \(\{\mathbf{H}^{(j)}, \mathbf{H}^{(j+1)}, \ldots, \mathbf{H}^{(M)}\}\).

This scenario, although it may seem less practical, may arise in many heterogeneous networks. Fig.\(\text{C}\) 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 \([33]\) or analog feedback \([36]\) is used.

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

**Proof.** The proof follows directly from the proof of Theorem \([\text{I}]\) in Section \([\text{V}]\) \(\blacksquare\)

In this setting, TX 1 already knows \(\mathbf{H}^{(j)}\) 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 \([\text{VI}]\).
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 \( \Delta R = \sum_{i \in N_K} \Delta R_i \) 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 \( o(1) \)). Let \( \mathcal{H}_\|^\) denote the set of inconsistent events, i.e., \( \mathcal{H}_\|^\triangleq \bigcup_{2 \leq j \leq M} \{ \mathcal{H}_j^{(j)} \neq \hat{\mathcal{H}}_j^{(j)} \} \). Hence, the probability of having feasible consistent precoding is \( \Pr(\mathcal{P}_o^c \cap \mathcal{H}_\|^\) By means of the law of total expectation, we can split the expected rate gap for TX \( i \) \( \Delta R_i \) as

\[
\Delta R_i = \Pr(\mathcal{P}_o^c \cup \mathcal{H}_\|^\Delta R_{i|\mathcal{P}_o^c \cup \mathcal{H}_\|^\),
\]

Let us focus on \( \Delta R_{i|\mathcal{P}_o^c \cup \mathcal{H}_\|^\). 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^c \cup \mathcal{H}_\|^\leq R_{i|\mathcal{P}_o^c \cup \mathcal{H}_\|^\(\alpha^{(1)}\)\). By neglecting the received interference, we can write that

\[
R_{i|\mathcal{P}_o^c \cup \mathcal{H}_\|^*\(\alpha^{(1)}\)} \leq \mathbb{E}_{\mathcal{P}_o^c \cup \mathcal{H}_\|^}[\log_2 \left( 1 + P |h_i v_i|^2 \right)]
\]

\[
\leq \log_2(P) + \mathbb{E}_{\mathcal{P}_o^c \cup \mathcal{H}_\|^} \left[ \log_2 \left( 1 + |h_i v_i|^2 \right) \right].
\]

(33)

Moreover, the set \( \mathcal{P}_o^c \cup \mathcal{H}_\|^\ 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^c \cup \mathcal{H}_\|^*\(\alpha^{(1)}\)} = \Theta(\log_2(P)) \). Hence, it follows from Lemma 1 and Lemma 2 that

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

and consequently

\[
\Delta R_i = \Delta R_{i|\mathcal{P}_o^c \cup \mathcal{H}_\^} + o(1).
\]

(34)

Thus, in the remainder of the proof we assume w.l.o.g. that TX 1 knows \( \hat{\mathcal{H}}_j^{(j)} \) for any \( j \in N_M \), 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 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 \( \Delta R_{i|\mathcal{P}_o^c \cap \mathcal{H}_\^} \) converges to the centralized setting rate.

B. Reformulating the Rate Gap

We can rewrite the rate gap for RX \( i \) as

\[
\Delta R_i = \mathbb{E} \left[ \log_2 \left( 1 + \frac{P |h_i v_i|^2}{1 + P |\sum_{\ell \neq i} |h_i v_i|^2} \right) \right] - \mathbb{E} \left[ \log_2 \left( 1 + \frac{P |h_i w_i|^2}{1 + P |\sum_{\ell \neq i} |h_i w_i|^2} \right) \right]
\]

\[
= \mathbb{E} \left[ \log_2 \left( 1 + \frac{P |h_i v_i|^2}{1 + P |\sum_{\ell \neq i} |h_i v_i|^2} \right) \right]_{\mathcal{F}_D} - \mathbb{E} \left[ \log_2 \left( 1 + \frac{P |h_i w_i|^2}{1 + P |\sum_{\ell \neq i} |h_i w_i|^2} \right) \right]_{\mathcal{F}_I}
\]

(35)

This rewriting of \( \Delta R_i \) allows us to separate the ratio of received interference power \( (\mathcal{F}_D) \) and the ratio of total received power \( (\mathcal{F}_D) \). In the following, we will prove that \( \lim_{P \to \infty} \Delta R_i = 0 \) by showing that \( \lim_{P \to \infty} \mathbb{E}[\log_2(\mathcal{F}_I)] = 0 \) 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 \( \mu \leq 1 \),

\[
\mathbb{E} \left[ \log_2 \left( 1 + \frac{P |\sum_{\ell \neq i} |h_i w_i|^2}{1 + P |\sum_{\ell \neq i} |h_i v_i|^2} \right) \right] \leq \mathbb{E} \left[ \log_2 \left( 1 + \frac{P |h_i w_i|^2}{1 + P |\sum_{\ell \neq i} |h_i w_i|^2} \right) \right],
\]

(36)

where we have introduced the notation \( \mathcal{F}_I^\) for the sake of readability. Let \( \eta \) 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,

\[
\mathbb{E}[\log_2(\mathcal{F}_I^\)] = \Pr(\mathcal{F}_I^\ < 1 + \eta) \mathbb{E}[\mathcal{F}_I^\ < 1 + \eta] + \Pr(\mathcal{F}_I^\ \geq 1 + \eta) \mathbb{E}[\mathcal{F}_I^\ \geq 1 + \eta].
\]

(37)

Now we present a useful lemma.

Lemma 3. Let \( \eta = \rho^{-\varepsilon} \), with \( \alpha_\eta > \varepsilon > 0 \) and \( \varepsilon \) arbitrarily small. Then,

\[
\Pr(\mathcal{F}_I^\ \geq 1 + \eta) = o \left( \frac{1}{\log_2(P)} \right)
\]

(38)
and
\[ \Pr \left( \frac{1}{\mathcal{F}_D} \geq 1 + \eta \right) = o \left( \frac{1}{\log_2(P)} \right). \] (39)

**Proof.** The proof is relegated to Appendix [II-A]. □

Let \( \eta = P^{-\varepsilon} \), with \( \alpha_q > \varepsilon > 0 \) and \( \varepsilon \) arbitrarily small. Then, (37) becomes
\[
\mathbb{E} \left[ \log_2 (\mathcal{F}_T) \right] \leq \mathbb{E}_{\mathcal{F}_T<1+\eta} \left[ \log_2 (\mathcal{F}_T) \right] + o \left( \frac{1}{\log_2(P)} \right) \mathbb{E} \left[ \log_2 (\mathcal{F}_T) \right] \leq \log_2(1 + \eta) + o(1)
\]
since \( \mathbb{E}_{\mathcal{F}_T\geq1+\eta} \left[ \log_2 (\mathcal{F}_T) \right] = O(\log_2(P)) \). In order to prove that \( \mathbb{E}_{\mathcal{F}_T\geq1+\eta} \left[ \log_2 (\mathcal{F}_T) \right] = O(\log_2(P)) \), note that
\[
\mathbb{E}_{\mathcal{F}_T\geq1+\eta} \left[ \log_2 (\mathcal{F}_T) \right] = \frac{1}{\Pr (\mathcal{F}_T \geq 1 + \eta)} \mathbb{E} \left[ \log_2 (\mathcal{F}_T) \right] - \Pr (\mathcal{F}_T < 1 + \eta) \mathbb{E}_{\mathcal{F}_T<1+\eta} \left[ \log_2 (\mathcal{F}_T) \right] \leq \frac{1}{\Pr (\mathcal{F}_T \geq 1 + \eta)} \mathbb{E} \left[ \log_2 (\mathcal{F}_T) \right].
\]

Further, we have
\[
\mathbb{E} \left[ \log_2 (\mathcal{F}_T) \right] \leq \mathbb{E} \left[ \log_2 \left( 1 + P \sum_{i \neq i} \left| \mathbf{h}_i \mathbf{w}_i \right|^2 \right) \right] \leq \log_2(P) + \mathbb{E} \left[ \log_2 \left( 1 + \sum_{i \neq i} \left| \mathbf{h}_i \mathbf{w}_i \right|^2 \right) \right].
\]
From (41), (42), and the fact that \( \Pr (\mathcal{F}_T \geq 1 + \eta) = \Theta(1) \), we obtain that \( \mathbb{E} \left[ \log_2 (\mathcal{F}_T) \right] = O(\log_2(P)) \).

2) **Lower-bounding** \( \mathbb{E} \left[ \log_2 (\mathcal{F}_T) \right] \): Let us now lower-bound the expectation. Note that
\[
\mathbb{E} \left[ \log_2 (\mathcal{F}_T) \right] \geq \log_2 \left( \frac{1}{\mathcal{F}_D} \right) + \log_2(1 + \mu) + o(1) \text{ and hence}
\]
\[
\mathbb{E} \left[ \log_2 (\mathcal{F}_T) \right] \geq \log_2(1 + \mu) + o(1).
\]
Consequently, the term \( \mathbb{E} \left[ \log_2 (\mathcal{F}_T) \right] \) can be bounded as
\[
\log_2(1 + \mu^2) - \log_2(1 + \mu) + o(1) \leq \mathbb{E} \left[ \log_2 (\mathcal{F}_T) \right] \leq \log_2(1 + \mu) + o(1).
\]
Since \( \lim_{P \to \infty} \mu = 1 \) and \( \lim_{P \to \infty} \eta = 0 \), it follows that
\[
\lim_{P \to \infty} \mathbb{E} \left[ \log_2 (\mathcal{F}_T) \right] = 0.
\]

**D. Analysis of the Received Signal Ratio (F_D)**

It remains to prove that the first expectation in (35) also converges to zero. As for \( \mathcal{F}_T \), we can write
\[
\mathbb{E} \left[ \log_2 (\mathcal{F}_D) \right] \leq \log_2 \left( \frac{1}{\mu^2} \right) + \mathbb{E} \left[ \log_2 \left( \frac{1 + \sum_{i \in H_K} \left| \mathbf{h}_i \mathbf{v}_i \right|^2}{1 + \sum_{i \in H_K} \left| \mathbf{h}_i \mathbf{w}_i \right|^2} \right) \right].
\]

Moreover, the equivalent to Lemma 3 also holds for \( \mathcal{F}_D \).

**Lemma 4.** Let \( \eta = P^{-\varepsilon} \), with \( \alpha_q > \varepsilon > 0 \) and \( \varepsilon \) arbitrarily small. Then,
\[
\Pr \left( \frac{1}{\mathcal{F}_D} \geq 1 + \eta \right) = o \left( \frac{1}{\log_2(P)} \right).
\]

**Proof.** The proof is relegated to Appendix [II-B]. □

Thus, applying the same step as in (40) yields
\[
\mathbb{E} \left[ \log_2 (\mathcal{F}_D) \right] \leq \log_2(1 + \eta) + o(1).
\]
We can similarly lower-bound \( \mathbb{E} \left[ \log_2 (\mathcal{F}_D) \right] \) to obtain that
\[
- \log_2(1 + \eta) + o(1) \leq \mathbb{E} \left[ \log_2 (\mathcal{F}_D) \right] \leq \log_2(1 + \eta) + o(1).
\]
The fact that \( \lim_{P \to \infty} \mu = 1 \) and \( \lim_{P \to \infty} \eta = 0 \) leads to
\[
\lim_{P \to \infty} \mathbb{E} \left[ \log_2 (\mathcal{F}_D) \right] = 0.
\]

**E. Merging Previous Sections**

Given that \( \lim_{P \to \infty} \Delta R = \lim_{P \to \infty} \sum_{i=1}^K \Delta R_i \), we obtain that
\[
\lim_{P \to \infty} \Delta R = \lim_{P \to \infty} K (\mathbb{E} \left[ \log_2 (\mathcal{F}_D) \right] + \mathbb{E} \left[ \log_2 (\mathcal{F}_T) \right]) = 0,
\]
which concludes the proof of Theorem 1. □

**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.
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 [59], [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

\[
\mathbb{E} \left[ \| v_{i,j} - \bar{w}_{i,j} \|^2 \right] = \mathcal{O} \left( \bar{P}^{-\alpha_q} \right) \tag{54}
\]

\[
\mathbb{E} \left[ \| v_{i,j} - w_{i,j} \|^2 \right] = \mathcal{O} \left( P^{-\alpha_q} \right). \tag{55}
\]

Proof. The proof is provided in Appendix VI. \hfill \Box

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 $P^{-\alpha}$ for any $\alpha > 0$. 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

\[
\mathbb{E} \left[ \| v_i - w_i \|^2 \right] = \mathcal{O} \left( \bar{P}^{-\alpha_q} \right) \tag{56}
\]

\[
\mathbb{E} \left[ \| v_i - w_i \|^2 \right] = \mathcal{O} \left( P^{-\alpha_q} \right). \tag{57}
\]

Proof. The proof is relegated to Appendix VI. \hfill \Box

![Fig. 7: Performance of different precoding schemes for a setting with 2 single-antenna TXs and 2 RXs with instantaneous power constraint and $\alpha^{(1)} = 1, \alpha^{(2)} = 0.6.$](image)
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 \( \varepsilon \) arbitrarily small, it holds that

\[
\Pr \left( \mathcal{F}_2' \geq 1 + \eta \right) = o \left( \frac{1}{\log_2(P)} \right) \tag{58}
\]

and

\[
\Pr \left( \frac{1}{\mathcal{F}_2'} \geq 1 + \eta \right) = o \left( \frac{1}{\log_2(P)} \right) \tag{59}
\]

We start by noting that \( h_i \) can be written as \( h_i = \frac{1}{2\pi} (\hat{h}_i^{(i)} - z^{(i)} \delta_i^{(i)}) \) from the definition of the estimate in (53). Let us introduce the notations \( \hat{z}^{(i)}_{\text{inv}} = \frac{1}{2\pi} z^{(i)}_{\text{inv}} \) and \( z_n^{(i)} = \frac{1}{2\pi} z^{(i)}_n \). Hence, for any \( i \neq \ell \), it follows that

\[
|h_i w_\ell| = \left| z^{(i)}_{\text{inv}} \delta_i^{(i)} w_\ell - z_n^{(i)} \delta_i^{(i)} w_\ell \right|
\]

where (a) and (c) come from the D-CIT model of (13) and \( b \) from the precoder definition in (25) since \( h_i^{(i)} w_\ell = h_i^{(i)} w_\ell \). Let us define \( d_i^{(i)} w_\ell = w_\ell - v_\ell \) for the sake of readability and space. Hence, from the triangular inequality it follows that

\[
1 + P \sum_{\ell 
eq i} |h_i w_\ell|^2 \leq 1 + \frac{P}{1 + P \sum_{\ell 
eq i} |h_i v_\ell|^2} \sum_{\ell 
eq i} \left( |z_n^{(i)} \delta_i^{(i)} d_i^{(i)} w_\ell|^2 + 2 |h_i v_\ell| |z_n^{(i)} \delta_i^{(i)} d_i^{(i)} w_\ell| \right) \tag{60}
\]

Let us recall that

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

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

\[
\Pr \left( \mathcal{F}_2' \geq 1 + \eta \right)
\leq \sum_{\ell 
eq i} \Pr \left( \frac{\left( |z_n^{(i)} \delta_i^{(i)} d_i^{(i)} w_\ell|^2 + 2 |h_i v_\ell| |z_n^{(i)} \delta_i^{(i)} d_i^{(i)} w_\ell| \right)}{1 + P \sum_{\ell 
eq i} |h_i v_\ell|^2} \geq \eta \frac{K}{K} \right)
\leq \sum_{\ell 
eq i} \Pr \left( \frac{|z_n^{(i)} \delta_i^{(i)} d_i^{(i)} w_\ell|^2}{|h_i v_\ell|^2} \geq \eta \frac{K}{K} \right)
\leq (K - 1) \Pr \left( \frac{|z_n^{(i)} \delta_i^{(i)} d_i^{(i)} w_\ell|^2}{|h_i v_\ell|^2} \geq \eta \frac{K}{K} \right)
\leq \frac{c}{K} \frac{1}{\log_2(P)}
\]

where (a) comes from symmetry, (b) from (62), and (c) because \( \eta < 1 \) and \( d_i^{(i)} w_\ell = w_\ell - v_\ell \). Let us now introduce a parameter \( \gamma \in \mathbb{R} \). We can continue as

\[
\Pr \left( \frac{|z_n^{(i)} \delta_i^{(i)} (w_\ell - v_\ell)|}{|h_i v_\ell|} \geq \eta \frac{K}{K} \right)
= \Pr \left( |\delta_i^{(i)} (w_\ell - v_\ell)| \geq \eta \frac{K}{K} |\delta_i^{(i)} v_\ell| \right)
\leq \Pr \left( |\delta_i^{(i)} v_\ell| < P^{-\gamma} \right)
\]

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

**Proposition 1.** Let \( \gamma > 0 \). Then,

\[
\Pr \left( |\delta_i^{(i)} v_\ell| < P^{-\gamma} \right) = o \left( \frac{1}{\log_2(P)} \right). \tag{65}
\]

**Proof.** The proof is relegated to Appendix [V]. \( \square \)

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

\[
\int_{|\delta_i^{(i)} v_\ell| \geq P^{-\gamma}} E \left[ |\delta_i^{(i)} (w_\ell - v_\ell)| \right] f_{|\delta_i^{(i)} v_\ell|}(y) \frac{1}{P^{-\gamma}} dy \leq 4K \frac{1}{\eta} E \left[ \delta_i^{(i)} (w_\ell - v_\ell) \right] E \left[ |\delta_i^{(i)} v_\ell| \geq P^{-\gamma} \right] \frac{1}{|\delta_i^{(i)} v_\ell|} \tag{66}
\]

Now, we introduce another useful proposition, whose proof is also relegated to Appendix [V].

**Proposition 2.** It holds that

\[
E \left[ |\delta_i^{(i)} (w_\ell - v_\ell)| \right] = O(\bar{P}^{-\alpha_q}). \tag{67}
\]

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

\[
\Pr \left( \mathcal{F}_2' \geq 1 + \eta \right) \leq \alpha \left( \frac{1}{\log_2(P)} \right) + 8K(K-1)O(P^{-\alpha_q})P^{\gamma}. \tag{68}
\]

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

\[
\Pr \left( \mathcal{F}_2' \geq 1 + \eta \right) \leq \alpha \left( \frac{1}{\log_2(P)} \right) + \bar{P}^{\varepsilon}O(P^{-\alpha_q})P^{\gamma}. \tag{69}
\]

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

\[
\Pr \left( \mathcal{F}_2' \geq 1 + \eta \right) \leq o \left( \frac{1}{\log_2(P)} \right). \tag{70}
\]
which concludes the proof of the first statement of Lemma 3.

We prove in the following the second statement, i.e.,

$$\Pr \left( \frac{1}{F_i} \geq 1 + \eta \right) = o \left( \frac{1}{\log_2(P)} \right). \quad (71)$$

This is obtained by switching the vectors $v_i$ and $w_i$ 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

$$\Pr \left( \frac{1}{F_i} \geq 1 + \eta \right) = \Pr \left( \frac{1 + P \sum_{t \neq i} |h_t v_t|^2}{1 + P \sum_{t \neq i} |h_t w_t|^2} \geq 1 + \eta \right)$$

$$\leq 2(K-1) \Pr \left( \frac{|z_n^{(1)} \delta_n^{(1)} (w_i - v_i)|}{|h_i w_i|^2} \geq \eta \frac{4K}{4K} \right). \quad (72)$$

Furthermore, the final expression in (72) is equal to the one in (63) except for the fact that the denominator is $|h_i v_i|$ instead of $|h_i v_i|$. Hence, continuing as in (64)-(69), we obtain that

$$\Pr \left( \frac{1}{F_i} \geq 1 + \eta \right) = o \left( \frac{1}{\log_2(P)} \right), \quad (73)$$

which concludes the proof of Lemma 3. \qed



B. Proof of Lemma 4

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

$$\Pr \left( F_D \geq 1 + \eta \right) = o \left( \frac{1}{\log_2(P)} \right) \quad (74)$$

and

$$\Pr \left( \frac{1}{F_D} \geq 1 + \eta \right) = o \left( \frac{1}{\log_2(P)} \right). \quad (75)$$

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

$$|h_i v_i|^2 \leq |h_i w_i|^2 + |z_n^{(1)} \delta_n^{(1)} (w_i - v_i)|^2$$

$$+ 2 |h_i w_i| |z_n^{(1)} \delta_n^{(1)} (w_i - v_i)|,$$  

$$\quad (76)$$

$$|h_i v_i|^2 \leq |h_i w_i|^2 + |h_i (w_i - v_i)|^2$$

$$+ 2 |h_i w_i| |h_i (w_i - v_i)|.$$  

\quad (77)

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

$$\Pr \left( \frac{1 + P \sum_{t \neq i} |h_t v_t|^2}{1 + P \sum_{t \neq i} |h_t w_t|^2} \geq 1 + \eta \right)$$

$$\leq \Pr \left( D_1 + D_2 + D_3 + D_4 \geq \eta \right) \quad (78)$$

where we have introduced the notations

$$D_1 \triangleq \sum_{t \neq i} \frac{|z_n^{(1)} \delta_n^{(1)} (w_t - v_t)|^2}{|h_t w_t|^2} \quad (79)$$

$$D_2 \triangleq \sum_{t \neq i} 2 \frac{|h_t w_t| |z_n^{(1)} \delta_n^{(1)} (w_t - v_t)|}{|h_t w_t|^2} \quad (80)$$

$$D_3 \triangleq \frac{|h_i (w_i - v_i)|^2}{|h_i w_i|^2} \quad (81)$$

$$D_4 \triangleq \frac{2 |h_i w_i| |h_i (w_i - v_i)|}{|h_i w_i|^2}.$$  

\quad (82)

The first inequality in (78) is obtained by applying (76)-(77) and eliminating the term $1 + P \sum_{t \neq i} |h_t w_t|^2$ from the denominator. From the analysis of $F_D$ in the previous section (see (64)), it follows easily that, if $\eta = \tilde{P}^{-\varepsilon}$, with $\alpha_q > \varepsilon > 0$, then

$$\Pr \left( D_1 \geq \eta \frac{4}{4} \right) \leq \sum_{t \neq i} \Pr \left( \frac{|z_n^{(1)} \delta_n^{(1)} (w_t - v_t)|^2}{|h_t w_t|^2} \geq \eta \frac{4K}{4K} \right)$$

$$= o \left( \frac{1}{\log_2(P)} \right). \quad (83)$$

Similarly,

$$\Pr \left( D_2 \geq \eta \frac{4}{4} \right) = \sum_{t \neq i} \Pr \left( \eta \frac{2 |h_t w_t| |z_n^{(1)} \delta_n^{(1)} (w_t - v_t)|}{|h_t w_t|^2} \geq \eta \frac{4K}{4K} \right)$$

$$= o \left( \frac{1}{\log_2(P)} \right). \quad (84)$$

For the two remaining terms, $D_3$ and $D_4$, note that

$$\Pr \left( D_3 \geq \eta \frac{4}{4} \right) + \Pr \left( D_4 \geq \eta \frac{4}{4} \right)$$

$$= \Pr \left( \frac{|h_i (w_i - v_i)|^2}{|h_i w_i|^2} \geq \eta \frac{4}{16} \right)$$

$$+ \Pr \left( \frac{2 |h_i w_i| |h_i (w_i - v_i)|}{|h_i w_i|^2} \geq \eta \frac{4}{16} \right) \quad (85)$$

$$\leq 2 \Pr \left( \frac{|h_i (w_i - v_i)|}{|h_i w_i|} \geq \eta \frac{4}{16} \right)$$

$$= 2 \Pr \left( \frac{|h_i (w_i - v_i)|}{|h_i w_i|} \geq \eta \frac{4}{16} \right).$$

\quad (86)

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

$$\Pr \left( \frac{|h_i (w_i - v_i)|}{|h_i w_i|} \geq \eta \frac{4}{16} \right) \leq \Pr \left( \tilde{h}_i w_i < \tilde{P}^{-\gamma} \right)$$

$$+ \int_{|h_i w_i| \geq \tilde{P}^{-\gamma}} \frac{E[|\tilde{h}_i (w_i - v_i)|]}{2 \pi y} \frac{y}{\pi y} y^{N_T - 2} dy \quad (86)$$

$$\leq O(\tilde{P}^{-\gamma}) + 16 \tilde{P}^{-\gamma} E[|\tilde{h}_i (w_i - v_i)|] \tilde{P}^{-\gamma}. \quad (87)$$

The fact that $|\tilde{h}_i| = 1$ implies that $E[|\tilde{h}_i (w_i - v_i)|] \leq E[|w_i - v_i|]$. Moreover, Corollary 3 states that $E[|w_i - v_i|] = O(\tilde{P}^{-\alpha_4})$. Consequently, by selecting...
We introduce the result of (87) into (85) to obtain from (78) that
\[ \Pr \left( \frac{\|h_i(w_i - v_i)\|}{\|h_i w_i\|} \geq \frac{\eta}{16} \right) \leq O(\bar{P}^{-\gamma}) + \bar{P}^c O(\bar{P}^{-\alpha_q} \bar{P}^\gamma) = o \left( \frac{1}{\log_2(\bar{P})} \right). \] (87)

We introduce the result of (87) into (85) to obtain from (78) that
\[ \Pr (\mathcal{F}_D \geq 1 + \eta) = o \left( \frac{1}{\log_2(\bar{P})} \right). \] (88)

It would remain to prove that \( \Pr \left( \frac{1}{\bar{P}} \geq 1 + \eta \right) = o \left( \frac{1}{\log_2(\bar{P})} \right). \)

To do so, we just need to apply the same previous steps, but interchanging \( w \) and \( v \). Following those steps and following a similar argument as in the proof for \( \Pr \left( \frac{1}{\bar{P}} \right) \), 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 \( \mathcal{R}_o \). Note that
\[ \Pr (\mathcal{R}_o) \leq N_1 \Pr (\|T_{1,1}\| > 1), \] (89)
and \( T_{1,1} = \mu \|w_{1,1}w_{2,1}, \ldots, w_{K,1}\| \), where \( w_{i,j,n} \) represents the \( n \)-th element of the precoding vector at TX \( j \) for the data symbols of RX \( i \). Therefore,
\[ \Pr (\mathcal{R}_o) \leq N_1 \Pr (\|\mu \|w_{1,1}, w_{2,1}, \ldots, w_{K,1}\| > 1) \]
\begin{align*}
& \leq N_1 \Pr \left( \bigcup_{i \in \mathbb{N}_K} \|w_{i,1,1}\| > \|v_{i,1,1}\| \right) \quad \text{(a)} \\
& \leq N_1 K \Pr (\|\|w_{1,1,1}\| > \|v_{1,1,1}\|) \quad \text{(b)} \\
& \leq N_1 K \Pr (\|v_{1,1,1}\| + \|\phi_1\| > \|v_{1,1,1}\|) \quad \text{(c)} \\
& = N_1 K \Pr (\|\phi_1\| > \frac{1 - \mu}{\mu} \|v_{1,1,1}\|) \\
& \leq N_1 K \int_{-\infty}^{\infty} \mathbb{E} [\|\phi_1\|] \frac{1}{\mu} \mathbb{E} [\|v_{1,1,1}\|^{-1}] \, d\nu. \quad \text{(91)}
\end{align*}

Using Markov’s inequality we obtain that
\[ \Pr (\mathcal{R}_o) \leq N_1 K \int_{-\infty}^{\infty} \mathbb{E} [\|\phi_1\|] \frac{1}{\mu} \mathbb{E} [\|v_{1,1,1}\|^{-1}] \, d\nu = N_1 K \mathbb{E} [\|\phi_1\|] \frac{1}{\mu} \mathbb{E} [\|v_{1,1,1}\|^{-1}], \] (92)
where \( \mathbb{E} [\|v_{1,1,1}\|^{-1}] \) exists from property (ZF). Let us focus on the first expectation term of (92). \( \mathbb{E} [\|\phi_1\|] \). Recalling (29), \( \phi_i \) is defined as
\[ \phi_i = \hat{H}_{i,1} \hat{H}_{i,1} \|v_{i,1} - w_{i,1}\|. \] (93)

Then,
\[ \mathbb{E} [\|\phi_1\|] \leq \mathbb{E} [\|\hat{H}_{i,1} \hat{H}_{i,1} \|v_{i,1} - w_{i,1}\|] \]
\[ \leq \sqrt{\mathbb{E} [\|\hat{H}_{i,1} \hat{H}_{i,1}\|^2] \mathbb{E} [\|v_{i,1} - w_{i,1}\|^2],} \] (94)
where (a) comes from the sub-multiplicativity property of the Frobenius norm and (b) from Cauchy-Schwarz inequality. Let us denote \( g_m \triangleq \sqrt{\mathbb{E} [\|\hat{H}_{i,1} \hat{H}_{i,1}\|^2] \mathbb{E} [\|v_{i,1} - w_{i,1}\|^2]} \), 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
\[ \mathbb{E} [\|\phi_1\|] \leq g_m \sqrt{\mathbb{E} [\|v_{i,1} - w_{i,1}\|^2].} \] (95)

Lemma 5 and the fact that \( \mathbb{E} [\|v_{i,1} - w_{i,1}\|^2] = \sum_{j=2}^{M} \mathbb{E} [\|v_{i,j} - w_{i,j}\|^2] \) yield
\[ \mathbb{E} [\|\phi_1\|] = O (\bar{P}^{-\alpha_q}). \] (96)

Since \( \mu = 1 - \omega \), with \( \omega = \Theta (\bar{P}^{-\alpha_q}) \) and \( \omega > 0 \), the term \( \frac{1}{\mu} \) satisfies \( \frac{1}{\mu} = \Theta (\bar{P}^{\alpha_q}). \) From (ZF2), \( \mathbb{E} [\|v_{1,1,1}\|^{-1}] \) is \( \Theta (1) \). Hence, recalling (29),
\[ \Pr (\mathcal{R}_o) \leq N_1 K \mathbb{E} [\|v_{1,1,1}\|^{-1}] \mathbb{E} [\|\phi_1\|] \frac{1}{\mu} \]
\[ = \Theta (1) O (\bar{P}^{-\alpha_q}) \Theta (\bar{P}^{\alpha_q}), \] (97)
which implies that \( \Pr (\mathcal{R}_o) = O (\bar{P}^{-\alpha_q}) \). By selecting \( \alpha_m < \alpha_q \), the probability of power outage vanishes and it holds that
\[ \Pr (\mathcal{R}_o) = o \left( \frac{1}{\log_2(\bar{P})} \right), \] (98)
which concludes the proof.

\[ \square \]

**Appendix IV**

**Proof of Proposition 1 and Proposition 2**

**A. Proof of Proposition 7**

We prove in the following that \( \Pr (|\delta_i^{(1)} v_{\ell}| < \bar{P}^{-\gamma}) = o \left( \frac{1}{\log_2(P)} \right) \) for any \( \gamma > 0 \) and \( i, \ell \in \mathbb{N}_K \) such that \( \ell \neq i \).

Let us denote the precoder for RX \( \ell \) obtained with perfect knowledge of \( H \) as \( v_{\ell} \). Then,
\[ \Pr (|\delta_i^{(1)} v_{\ell}| < \bar{P}^{-\gamma}) = \Pr (|\delta_i^{(1)} u_{\ell} + \delta_j^{(1)} v_{\ell} - u_{\ell}| < \bar{P}^{-\gamma}) \]
\[ \leq \Pr (|\delta_i^{(1)} u_{\ell} - \delta_j^{(1)} v_{\ell} - u_{\ell}| < \bar{P}^{-\gamma}), \] (99)
where we have applied the inverse triangle inequality. In order to prove Proposition 1 we capitalize the intuition that the term \( |\delta_i^{(1)} u_{\ell}| \) is independent of the quality of the estimate and \( P \), but the value of \( |\delta_j^{(1)} v_{\ell} - u_{\ell}| \) is directly proportional to \( P^{-\alpha}\).

Before applying this intuition to (99), we first analyze the term \( |\delta_i^{(1)} v_{\ell} - u_{\ell}| \) and the probability \( \Pr (|\delta_i^{(1)} v_{\ell} - u_{\ell}| > \bar{P}^{-\beta}) \) for any \( \beta < \alpha^{(1)} \).

**Proposition 3.** For any \( \beta < \alpha^{(1)} \), it holds that
\[ \Pr (|\delta_i^{(1)} v_{\ell} - u_{\ell}| > \bar{P}^{-\beta}) = o \left( \frac{1}{\log_2(P)} \right). \] (100)

**Proof.** It follows by the Cauchy-Schwarz inequality that \( |\delta_i^{(1)} v_{\ell} - u_{\ell}| < |\delta_i^{(1)} v_{\ell} - u_{\ell}| \), which implies that
where \( \beta < \alpha \)
\[
\Pr \left( \| \delta_1 \| (\mathbf{v}_t - \mathbf{u}_t) > \bar{P}^{-\beta} \right) \leq \Pr \left( \left\| \delta_1 \right\| \| \mathbf{v}_t - \mathbf{u}_t \| > \bar{P}^{-\beta} \right).
\]
Let us define the scalar \( \varepsilon > 0 \) such that \( \beta < \beta + \varepsilon < \alpha \). By applying the law of total probability, we obtain that
\[
\Pr \left( \| \delta_1 \| (\mathbf{v}_t - \mathbf{u}_t) > \bar{P}^{-\beta} \right) \leq \Pr \left( \| \delta_1 \| \| \mathbf{v}_t - \mathbf{u}_t \| > \bar{P}^{-\beta} - \varepsilon \right) \times \Pr \left( \| \mathbf{v}_t - \mathbf{u}_t \| > \bar{P}^{-\beta} \varepsilon \right)
+ \Pr \left( \| \delta_1 \| \| \mathbf{v}_t - \mathbf{u}_t \| > \bar{P}^{-\beta} \varepsilon \right) \times \Pr \left( \| \mathbf{v}_t - \mathbf{u}_t \| > \bar{P}^{-\beta} - \varepsilon \right)
+ \Pr \left( \| \delta_1 \| > \bar{P}^\varepsilon \varepsilon \| \mathbf{v}_t - \mathbf{u}_t \| \right). (101)
\]
The term \( \Pr \left( \| \mathbf{v}_t - \mathbf{u}_t \| > \bar{P}^{-\beta} - \varepsilon \right) \) can be upper-bounded by means of the Markov’s inequality, such that
\[
\Pr \left( \| \mathbf{v}_t - \mathbf{u}_t \| > \bar{P}^{-\beta} - \varepsilon \right) \leq \bar{P}^{\beta + \varepsilon} \mathbb{E} \left[ \| \mathbf{v}_t - \mathbf{u}_t \| \right] (102)
\]
where \( a \) follows after applying Lemma [3] to vectors whose respective input estimates differ by an \( O(\bar{P}^{-\alpha(1)}) \) additive error term. For the last term in (101), \( \Pr \left( \| \delta_1 \| > \bar{P}^\varepsilon \varepsilon \| \mathbf{v}_t - \mathbf{u}_t \| \right) \), it follows that
\[
\Pr \left( \| \delta_1 \| > \bar{P}^\varepsilon \varepsilon \| \mathbf{v}_t - \mathbf{u}_t \| \leq \bar{P}^{-\beta} - \varepsilon \right)
\]
\[
\leq \frac{\Pr \left( \| \mathbf{v}_t - \mathbf{u}_t \| > \bar{P}^{-\beta} - \varepsilon \right) \Pr \left( \| \delta_1 \| > \bar{P}^\varepsilon \varepsilon \right)}{\Pr \left( \| \mathbf{v}_t - \mathbf{u}_t \| > \bar{P}^{-\beta} - \varepsilon \right)} (103)
\]
From (102), it holds that \( \Pr \left( \| \mathbf{v}_t - \mathbf{u}_t \| > \bar{P}^{-\beta} - \varepsilon \right) = 1 - O(\bar{P}^{-\alpha(1)} \delta_{1,n}^2), \) where \( \delta_{1,n} \) are i.i.d. as \( N(0, 1) \), consequently, \( \delta_{1,n}^2 \) is distributed following a Rayleigh distribution and
\[
\| \delta_1 \|^2 \sim \Gamma_d(N_t, 1), (104)
\]
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
\[
\Pr \left( X \sim \Gamma_d(N_t, 1) < x \right) = 1 - \sum_{n=0}^{N_t-1} \frac{1}{n!} e^{-\bar{P}^\varepsilon} \bar{P}^2 n x. (105)
\]
Hence,
\[
\Pr \left( \| \delta_1 \| > \bar{P}^\varepsilon \varepsilon \right) = \frac{\Pr \left( \| \delta_1 \|^2 > \bar{P}^2 \varepsilon \varepsilon \right)}{\sum_{n=0}^{N_t-1} \frac{1}{n!} e^{-\bar{P}^\varepsilon} \bar{P}^2 n x}. (106)
\]
Since \( \varepsilon > 0 \), \( \Pr \left( \| \delta_1 \| > \bar{P}^\varepsilon \varepsilon \right) = o(1/\log_2(P)) \), for any \( x \in \mathbb{R} \), and hence it is \( o(1/\log_2(P)) \). This implies that
\[
\frac{\Pr \left( \| \delta_1 \| > \bar{P}^\varepsilon \varepsilon \right)}{\Pr \left( \| \mathbf{v}_t - \mathbf{u}_t \| > \bar{P}^{-\beta} - \varepsilon \right)} = \frac{1}{\log_2(P)} (107)
\]
which, together with (102) and (101), leads to
\[
\Pr \left( \| \delta_1 \| (\mathbf{v}_t - \mathbf{u}_t) > \bar{P}^{-\beta} \right) = o \left( \frac{1}{\log_2(P)} \right) (108)
\]
for any \( \beta < \alpha(1) \).
where \( \delta^{(1)}_{i,1} \) denotes the first element of the vector \( \delta^{(1)}_i \), and it is distributed as \( N_C(0,1) \). Then,

\[
\Pr \left( \delta^{(1)}_{i,1} < 2 \bar{P}^{-\beta} \right) = \frac{2}{\sqrt{2\pi}} \int_0^{2 \bar{P}^{-\beta}} e^{-x^2/2} \, dx \leq \frac{4}{\sqrt{2\pi}} \bar{P}^{-\beta}. \tag{112}
\]

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

\[
\Pr (\| u \| \leq \bar{P}^{-\beta}) = \int_0^{\bar{P}^{-\beta}} f_u(x) \, dx \leq f_{\| u \|}^{\max}(\bar{P}^{-\beta}), \tag{113}
\]

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

\[
\Pr (\| \delta^{(1)} u \| < 2 \bar{P}^{-\beta}) \leq O(\bar{P}^{-\beta}-\epsilon_{\beta}, \epsilon_{\beta}-\beta), \tag{114}
\]

Note that \( \epsilon_{\beta} \) satisfies \( 0 < \epsilon_{\beta} < \beta \). Hence,

\[
\Pr (\| \delta^{(1)} v \| < \bar{P}^{-\gamma}) \leq \Pr (\| \delta^{(1)} u \| - \| \delta^{(1)} (v-u) \| < \bar{P}^{-\gamma}) \leq \Pr (\| \delta^{(1)} u \| < 2 \bar{P}^{-\beta}) + o\left( \frac{1}{\log_2(P)} \right) \tag{115}
\]

which concludes the proof of Proposition.

\[\square\]

### B. Proof of Proposition

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

\[
\mathbb{E}[\| \delta^{(1)} (w_{\ell} - v_{\ell}) \|] \leq \mathbb{E}[\| \delta^{(1)}_i \||w_{\ell} - v_{\ell}||, \] 

\[
= \text{cov} \left( \| \delta^{(1)}_i \||, ||w_{\ell} - v_{\ell}|| \right) + \mathbb{E} \left[ \| \delta^{(1)}_i \|| \mathbb{E}[||w_{\ell} - v_{\ell}||] \right] 
\]

\[
\leq \sqrt{\mathbb{E} \left[ \| \delta^{(1)}_i \|^2 \right]} \sigma_{|w_{\ell} - v_{\ell}|} + \mathbb{E} \left[ \| \delta^{(1)}_i \|| \mathbb{E}[||w_{\ell} - v_{\ell}||] \right],
\]

where \( \text{cov}(X,Y) \triangleq \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] \) is the covariance between \( X \) and \( Y \) and \( \sigma^2_X \) 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 \( \| \delta^{(1)}_i \|^2 \sim \Gamma_{\left(\frac{1}{2},N_T,1\right)} \) that \( \mathbb{E}[\| \delta^{(1)}_i \|^2] = N_T \).

From this point and the fact that \( \mathbb{E}[x] \leq \sqrt{\mathbb{E}[x^2]} \), we can write

\[
\mathbb{E}[\| \delta^{(1)} (w_{\ell} - v_{\ell}) \|] \leq \sqrt{N_T} \left( \sigma_{|w_{\ell} - v_{\ell}|} + \mathbb{E}[||w_{\ell} - v_{\ell}||] \right) \leq \sqrt{N_T} \left( \sigma_{|w_{\ell} - v_{\ell}|} + \mathbb{E}[||w_{\ell} - v_{\ell}||] \right) \leq \mathcal{O}(\bar{P}^{-\alpha_q}),
\]

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

### Appendix V

#### Proof of Lemma I (Quantizer Consistency)

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

\[
Q_u(x) \triangleq q \left[ \frac{x}{q} + \frac{1}{2} \right]. \tag{116}
\]

We extend the notation for any complex matrix \( A \in \mathbb{C}^{n \times m} \) such that \( A_q \triangleq Q_u(A) \) denotes the element-wise quantization, i.e.,

\[
(A_q)_{i,j} \triangleq Q_u(\text{Re}(A_{i,j}))+iQ_u(\text{Im}(A_{i,j})), \tag{117}
\]

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 \( Q_u \) with \( q = \bar{P}^{-\alpha_q} \) and \( \alpha_q > \alpha_q > 0 \), \( \forall j \in \mathbb{N}_M \), it follows that

\[
\Pr (H^{(0)}_q \neq H^{(1)}_q) = o\left( \frac{1}{\log_2(P)} \right), \tag{118}
\]

where \( H^{(0)}_q = Q_u(H^{(0)}_0) \) and \( H^{(0)}_q \neq H^{(1)}_q \) is the MAP estimator of \( H^{(0)}_q \) given \( H^{(1)}_q \). We start by noting that, by definition of the MAP estimator,

\[
\Pr (H^{(0)}_q \neq H^{(1)}_q) \leq \Pr (Q_u(H^{(0)}_q) \neq Q_u(H^{(1)}_q)). \tag{119}
\]

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

\[
\Pr (Q_u(H^{(0)}_q) \neq H^{(0)}_q) \leq 2KN_T \Pr (Q_u(\text{Re}(H^{(1)}_{1,k})) \neq Q_u(\text{Re}(H^{(1)}_{1,k}))), \tag{120}
\]

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 \( \text{Re}(H^{(1)}_{1,1}) \). 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 \( H^{(1)}_{1,i} \) is in the center of the quantization level and, besides this, the probability that \( H^{(1)}_{1,i} \) is in the edge area also vanishes. We rigorously show it in the following. Before starting, we introduce the simplified notation \( h^{(j)} \triangleq \text{Re}(H^{(1)}_{1,1}) \) to ease the readability. Accordingly, we also introduce the notation \( h \triangleq \text{Re}(H^{(1)}_{1,1}) \) and \( \delta \triangleq \text{Re}(H^{(1)}_{1,1}) \) such that \( h^{(j)} = \delta^{(j)}h + z^{(j)}\delta^{(j)} \), with \( z^{(j)} = \bar{P}^{-\alpha_q} \) and \( z^{(j)} \triangleq \sqrt{1 - (\bar{P}^{2})/2} \).

Furthermore, we recall the notations \( z^{(j)} \triangleq \frac{1}{\sqrt{2}} \) and \( z^{(j)} \triangleq \frac{1}{\sqrt{2}} \), introduced in Appendix I.

### A. Edge and center of the reconstruction level

Let \( \ell_n \) be the \( n \)-th quantization level of \( Q_u, n \in \mathbb{Z} \), with \( \ell_0 = 0 \). Let us define \( L_n \) as the input interval that outputs \( \ell_n \), i.e.,

\[
L_n \triangleq \{ x \mid Q_u(x) = \ell_n \}. \tag{121}
\]

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

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

The center area is given by

$$C_n \triangleq \{ x \in L_n \setminus E_n \}.$$  

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

$$\Pr \left( Q_u(h^{(1)}) \neq Q_u(h^{(j)}) \right) \leq \Pr \left( h^{(1)} \in \bigcup_{n \in \mathbb{Z}} E_n \right)$$

$$+ \Pr \left( Q_u(h^{(1)}) \neq Q_u(h^{(j)}) \mid h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n \right).$$

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 $\ell_n$. Let $f_{L_n}^{\max}$ be the maximum value of the pdf of $h^{(1)}$ in $L_n = \{ x \mid Q_u(x) = \ell_n \}$. It follows that the probability that $h^{(1)}$ is in $E_n$ is upper-bounded by

$$\Pr \left( h^{(1)} \in E_n \right) \leq f_{L_n}^{\max} \left| E_n \right|$$

$$= \frac{1}{2} f_{L_n}^{\max} \bar{P}^{-\alpha_q},$$

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

$$\Pr \left( h^{(1)} \in L_n \right) \geq \left( f_{L_n}^{\max} - 1/\sqrt{2\pi} \epsilon \right) L_n$$

$$= \left( f_{L_n}^{\max} - 1/\sqrt{2\pi} \epsilon \bar{P}^{-\alpha_q} \right) \bar{P}^{-\alpha_q}. $$

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

$$\Pr \left( h^{(1)} \in E_n \mid L_n \right) = \frac{\Pr \left( h^{(1)} \in E_n \right)}{\Pr \left( h^{(1)} \in L_n \right)}$$

$$\leq \frac{f_{L_n}^{\max}}{2 f_{L_n}^{\max} - 1/\sqrt{2\pi} \epsilon \bar{P}^{-\alpha_q}} \bar{P}^{-(\alpha_q - 1)/\alpha_q}.$$  

Let us define $g_{\text{max}}$ as

$$g_{\text{max}} \triangleq \max_{n \in \mathbb{Z}} \left\{ \frac{2 f_{L_n}^{\max}}{f_{L_n}^{\max} - 1/\sqrt{2\pi} \epsilon \bar{P}^{-\alpha_q}} \right\}.$$  

Note that $g_{\text{max}} = \Theta(1)$. Hence, from (126) and the fact that

$$\sum_{n \in \mathbb{Z}} \Pr \left( h^{(1)} \in L_n \right) = 1,$$  

and the fact that

$$\Pr \left( h^{(1)} \in \bigcup_{n \in \mathbb{Z}} E_n \right) = \sum_{n \in \mathbb{Z}} \Pr \left( h^{(1)} \in E_n \right) \Pr \left( h^{(1)} \in L_n \right)$$

$$\leq g_{\text{max}} \bar{P}^{-\alpha_q} \sum_{n \in \mathbb{Z}} \Pr \left( h^{(1)} \in L_n \right) = O \left( \bar{P}^{-\alpha_q} \right).$$  

Consequently, it holds that

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

**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 $\bar{P}^{-\alpha_q}$, it holds that

$$\Pr \left( Q_u(h^{(1)}) \neq Q_u(h^{(j)}) \mid h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n \right)$$

$$\leq \Pr \left( \left| h^{(1)} - h^{(j)} \right| \geq \bar{P}^{-\alpha_q} \mid h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n \right).$$

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

$$\Pr \left( \left| h^{(1)} - h^{(j)} \right| \geq \bar{P}^{-\alpha_q} \mid h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n \right)$$

$$\leq \frac{1}{\Pr \left( h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n \right)} \Pr \left( \left| h^{(1)} - h^{(j)} \right| \geq \bar{P}^{-\alpha_q} \right)$$

$$\leq \frac{1}{\Pr \left( h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n \right)} \mathbb{E} \left[ \left| h^{(1)} - h^{(j)} \right| \right],$$

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

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

From the assumption of Gaussian variables, it holds that

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

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

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

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

where $\sigma_d^2$ is given by

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

Substituting the variables for their values yields

$$\sigma_d^2 = 2 \left( 1 - \sqrt{1 - P^{-\alpha_q(1)} - P^{-\alpha_q} + P^{-\alpha_q(1)} - P^{-\alpha_q}} \right).$$

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

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

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 $E_n$. 


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
\[
\mathbb{E}[|h^{(1)} - h^{(j)}|] \leq \sqrt{\frac{4}{\pi} (P^{-\alpha(j)} + P^{-\alpha(i)} - P^{-\alpha(i)} - P^{-\alpha(j)})} = O(\tilde{P}^{-\alpha(j)}). \tag{138}
\]
Besides this, it holds from (128) that
\[
\Pr(h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n) = 1 - \Pr(h^{(1)} \in \bigcup_{n \in \mathbb{Z}} E_n) = 1 - O(\tilde{P}^{-\alpha(q)}). \tag{139}
\]
From (130), both (138) and (139) lead to
\[
\Pr(Q_u(h^{(1)}) \neq Q_u(h^{(j)}) \mid h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n) \leq \frac{1}{\Pr(h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n)} \mathbb{E}[|h^{(1)} - h^{(j)}|] = O(P^{-c_0\alpha_q}) = o\left(\frac{1}{\log_2(P)}\right). \tag{140}
\]
The last inequality is obtained only if \( c_0\alpha_q < \alpha(j) \) for any \( j \in \mathbb{N}_M \). Thus, it follows from (140) that it is necessary to satisfy that \( c_0\alpha_q < \alpha(j) \) for any \( j \in \mathbb{N}_M \). Since for any \( \alpha_q < \alpha(j) \) we can find a \( c_0 > 1 \) such that \( c_0\alpha_q < \alpha(j) \), any \( \alpha_q < \alpha(j) \) will satisfy (140) as long as a correct \( c_0 \) is selected.

**D. Assembling probabilities**

We make use of (128) and (140) to show that (123) satisfies
\[
\Pr(Q_u(h^{(1)}) \neq Q_u(h^{(j)}) \mid h^{(1)} \in \bigcup_{n \in \mathbb{Z}} E_n) + \Pr(Q_u(h^{(1)}) \neq Q_u(h^{(j)}) \mid h^{(1)} \in \bigcup_{n \in \mathbb{Z}} C_n) = o\left(\frac{1}{\log_2(P)}\right),
\]
which concludes the proof of Lemma 1.

**APPENDIX VI**

**PROOF OF LEMMA 3 (ERROR ON NAIVE PRECODER)**

In this appendix, we prove both Lemma 3 and Corollary 3. First, we focus on demonstrating that
\[
\mathbb{E}[|v_{i,2} - w_{i,2}|^2] = O(P^{-\alpha_q}). \tag{142}
\]
The generalization of (142) for any \( \mathbb{E}[|v_{i,k} - w_{i,k}|^2] \), \( k \in \mathbb{N}_M \setminus 1 \) 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 \( w_{i,2} \) and \( v_{i,2} \) are obtained following the same algorithm but based on different information (input). Specifically, \( w_{i,2} \) is computed on the basis of \( \hat{H}_q(2) = \mathbb{Q}_u(H(2)) \), where \( \mathbb{Q}_u \) is a scalar uniform quantizer with quantization step \( q = \tilde{P}^{-\alpha_q} \), and \( v_{i,2} \) is computed on the basis of \( H(1) \). As in the previous appendix, let \( h_q(2) \) (resp. \( h(1) \) and \( h \)) denote the real or imaginary part of an arbitrary element of the matrix \( \hat{H}_q(2) \) (resp. \( H(1) \) and \( H \)). Let us define \( h_c(2) \) as
\[
h_c(2) = h(2) + \varsigma_q, \tag{143}
\]
where \( \varsigma_q \) is distributed as a binary symmetric distribution with points \([-q, q]\), independent of the other variables, such that \( \sigma^2_q = q^2 \). Note that the error \( h_c(2) - h(1) \) has smaller or equal variance than \( h_c(2) - h(1) = h(2) - h(1) + \varsigma_q \). Hence, we can assume that \( w_{i,2} \) is computed on the basis of the estimate \( h_c(2) \), since increasing the error variance can only hurt. Consequently, the error \( \xi = h_c(2) - h(1) \) has a variance \( \sigma^2_{\xi} \) given by
\[
\sigma^2_{\xi} \leq \sigma^2_d + \sigma^2_q + 2\sigma_d\sigma_q \tag{144}
\]
where \( \sigma^2_d \) is given in (136). Therefore, we can write that
\[
h_c(2) = h(1) + \tilde{P}^{-\alpha_q} \delta_{\xi}, \tag{145}
\]
where \( \delta_{\xi} \) 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 (135), i.e., \( a_{\xi}(2) \triangleq a(1) + \tilde{P}^{-\alpha_q}\delta_{a,\xi} \), \( b_{\xi}(2) \triangleq b(1) + \tilde{P}^{-\alpha_q}\delta_{b,\xi} \), where \( \delta_{a,\xi}, \delta_{b,\xi} \), are variables with variance \( \Theta(1) \) and bounded density. It is easy to see that, for any \( a_{\xi}(2), b_{\xi}(2), \)
\[
a_{\xi}(2) + b_{\xi}(2) = a(1) + \tilde{P}^{-\alpha_q}\delta_{a,\xi} + b(1) + P^{-\alpha_q}\delta_{b,\xi}
\]
\[
= a(1) + b(1) + P^{-\alpha_q}(\delta_{a} + \delta_{b}). \tag{146}
\]
This implies that the error variance of the sum is also \( \Theta(P^{-\alpha_q}) \) as (144).

**B. Error in the product**

In a similar way, it follows that
\[
a_{\xi}(2) b_{\xi}(2) = \left(a(1) + \tilde{P}^{-\alpha_q}\delta_{a,\xi}\right)\left(b(1) + \tilde{P}^{-\alpha_q}\delta_{b,\xi}\right)
\[
= a(1)b(1) + \tilde{P}^{-\alpha_q}\left(a(1)\delta_{b} + b(1)\delta_{a} + \tilde{P}^{-\alpha_q}\delta_{a}\delta_{b}\right), \tag{147}
\]
which implies that the product also maintains the scaling of the variance as \( \Theta(P^{-\alpha_q}) \). 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 $A^{(2)}$ and $A^{(1)}$ are square matrices with full rank with probability one, and with coefficients following \(145\). We can then write that

\[
(A^{(2)})^{-1} = (A^{(1)} + \bar{P}^{-\alpha} \Delta^A_{\xi})^{-1} = (A^{(1)})^{-1} - \bar{P}^{-\alpha} (A^{(1)})^{-1} \Delta^A_{\xi} (A^{(1)} + \bar{P}^{-\alpha} \Delta^A_{\xi})^{-1},
\]

which is obtained from the Woodbury matrix identity \(59\). Hence, the error variance of the inverse is again $O(P^{-\alpha_q})$.

Once it is proved that the inverse operation generates an error with variance $O(P^{-\alpha_q})$, we extend it for the Moore–Penrose inverse (pseudo-inverse) \((A^{(2)})^\dagger\). We assume (as throughout the rest of the document) that each sub-matrix has maximum rank, i.e.,

\[
\text{rank} \left( A^{(2)} \right) \in \mathbb{C}^{N \times M} = \min(N, M).
\]

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

\[
(A^{(2)})^\dagger = (A^{(2)})^H \left( A^{(2)} (A^{(2)})^H \right)^{-1}.
\]

The case in which $A^{(2)}$ is full column-rank matrix ($N \geq M$) will follow the same steps and thus we omit it. It follows from \(147\) that

\[
(A^{(2)})^H (A^{(2)}) = (A^{(1)})^H (A^{(1)})^H + \bar{P}^{-\alpha} \Delta^t_{eq},
\]

where $\Delta^t_{eq}$ has variance $\Theta(1)$. This, together with \(148\), implies that

\[
(A^{(2)})^H (A^{(2)})^{-1} = (A^{(1)})^H (A^{(1)})^{-1} + \bar{P}^{-\alpha} \Delta^t_{eq},
\]

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

\[
(A^{(2)})^\dagger = (A^{(1)})^H \left( A^{(1)} (A^{(1)})^H \right)^{-1} + \bar{P}^{-\alpha} \Delta^t_{eq},
\]

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


d

D. Error variance of the difference of precoders

The centralized precoder $v_{i,2}$ is based on $H^{(1)}$, i.e., $V = \sqrt{\lambda} (H^{(1)})$. The distributed precoder at TX 2 is based on its own quantized CSI $\hat{H}^{(2)}$. And thus $w_{i,2}$ is obtained from $W = \sqrt{\lambda} (\hat{H}^{(2)})$. Based on the previous results and the definition of linear precoders, it follows that we can write the distributed precoder based on the CSI of TX 2 $(\hat{H}^{(2)})$ as

\[
w_{i,2} = v_{i,2} + \bar{P}^{-\alpha_q} e_w,
\]

where $E[\|e_w\|^2] = O(1)$. Consequently,

\[
E[\|v_{i,2} - w_{i,2}\|^2] = E[\|v_{i,2} - (v_{i,2} + \bar{P}^{-\alpha_q} e_w)\|^2] = P^{-\alpha_q} E[\|e_w\|^2] = O(P^{-\alpha_q}),
\]

which concludes the proof of \(142\) and thus of \(53\). Moreover, since $E[\|x\|^2] \leq \sqrt{E[\|k\|^2]}$, it follows that

\[
E[\|v_{i,j} - w_{i,j}\|^2] = O(P^{-\alpha_q}),
\]

which concludes the proof of Lemma \[5\]  \(\square\)

E. Proof of Corollary \[2\]

In order to prove that $E[\|w_i - v_i\|^2] = O(P^{-\alpha_q})$ for any $i \in \mathbb{N}_K$, we will recall that the vector $w_i - v_i$ can be written as

\[
w_i - v_i = \left[ \begin{array}{c} \phi_i \\ w_{i,1} - v_{i,1} \end{array} \right],
\]

where $\phi_i$ (defined in \[29\]) as $\phi_i = \hat{H}^{i,1}_I \hat{H}^{i,1}_I (v_{i,1} - w_{i,1})$ is the difference at TX 1, and $w_{i,1} - v_{i,1}$ denotes the difference for the coefficients of all the TXs but TX 1, i.e.,

\[
w_{i,1} - v_{i,1} = [(w_{i,2} - w_{1,2})^T, \ldots, (w_{i,K} - w_{1,K})^T]^T.
\]

Let us recall that $N_1$ has been defined as $N_1 = N_T - N_1$, and let us further define $G_1$ as

\[
G_1 = \left[ \begin{array}{c} \hat{H}^{i,1}_I \\ \hat{H}^{i,1}_I \\ I_{N_1} \end{array} \right].
\]

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

\[
w_{i,j} = v_{i,j} + \bar{P}^{-\alpha_q} e_{w,j}
\]

where $E[\|e_{w,j}\|^2] = O(1)$. Then, upon defining $e_{w,j} = [e_{w,2}, \ldots, e_{w,K}]^T$, it follows that

\[
E[\|w_i - v_i\|^2] = E[\|G_1 (w_{i,1} - v_{i,1})\|^2] = P^{-\alpha_q} E[\|G_1 e_{w,1}\|^2] = O(P^{-\alpha_q})
\]

since $E[\|G_1 e_{w,1}\|^2] = O(1)$, which proves \(57\). In order to prove \(56\), it follows from from $E[\|x\|^2] \leq \sqrt{E[\|k\|^2]}$ that

\[
E[\|v_i - w_i\|^2] = O(P^{-\alpha_q}),
\]

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 multiuser 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.
[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.