On the Degrees-of-Freedom
of the K-user Distributed Broadcast Channel

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

Abstract—We study the Degrees-of-Freedom (DoF) in a wireless setting in which K Transmitters (TXs) aim at jointly serving K users. The performance is studied when the TXs are faced with a distributed Channel State Information (CSI) configuration in which each TX has access to its own multi-user imperfect channel estimate based on which it designs its transmit coefficients. The channel estimates are not only imperfectly acquired but they are also imperfectly shared between the TXs. Our first contribution consists of computing a genie-aided upper bound for the DoF of that setting. Our main contribution is then to develop a new robust transmission scheme that leverages the different qualities of CSI available at the TXs to improve the achieved DoF. We show the surprising result that there is a CSI regime, coined the Weak-CSIT regime, in which the genie-aided upper bound is achieved by the proposed transmission scheme. Interestingly, the optimal DoF in the Weak-CSIT regime only depends on the CSI quality at the best informed TX and not on the CSI quality at all other TXs.

Index Terms—Degrees-of-Freedom (DoF), Network MIMO, Broadcast Channel, decentralized wireless networks.

I. INTRODUCTION

A. Limited Channel State Information on the Transmitter Side

CHANNEL capacity characterization of multi-user wireless networks is known to be an elusive problem for many practical scenarios, in particular for the cases in which the Channel State Information at the Transmitter (CSIT) is not perfect. In order to tackle this problem, capacity approximations at high Signal-to-Noise Ratio (SNR), such as Degrees-of-Freedom (DoF) analysis [3], have been used as a first step towards the complete characterization of the system capacity, and they have successfully led to many important insights. For example, the DoF of the MISO Broadcast Channel (BC) with imperfect noisy CSIT was characterized by Davoodi and Jafar in [4] by showing that a CSIT error variance scaling in $\mathrm{SNR}^{-\alpha}$, for $\alpha \in [0,1]$, leads to a DoF of $1 + (K-1)\alpha$.

A different line of work in the area of BC with limited feedback has been focused on the exploitation of delayed CSIT. This research area was triggered by the seminal work from Maddah-Ali and Tse [5] where it was shown that completely

outdated CSIT could still be exploited via a multi-phase protocol involving the retransmission of the interference generated. While the original model assumed completely outdated CSIT, a large number of works have developed generalized schemes for the case of partially outdated [6]–[8], alternating [9], or evolving CSIT [10], to name just a few.

In all the above literature, however, *centralized* CSIT is typically assumed, i.e., the transmission is optimized at the TX side on the basis of a *single* imperfect/outdated channel estimate being common at every transmit antenna. Recently, the increasing importance of cooperation of non-collocated TXs—as, for example, in Unmanned Aerial Device (UAV) aided networks [11]—has led to an increasing number of works challenging this assumption of centralized CSIT. In [12], [13], methods have been developed to reduce the CSIT required to achieve MIMO Interference Alignment (IA), and the DoF achieved with delayed and local CSIT in the Interference Channel (IC) has been also studied in several works [14]–[16]. Furthermore, the assumption of centralized CSIT has been challenged in capacity analysis for the Multiple Access Channel [17] and the Relay Channel [18], among others.

B. Network MIMO with Distributed CSIT Model

In order to account for TX-dependent limited feedback in the network MIMO channel, we focus in this work on a wireless configuration first studied in [19], in which the user's data symbols are available and jointly transmitted from all TXs, whereas the channel estimates could only be imperfectly obtained at the TXs. This network MIMO setting can be seen as a Distributed BC setting. We will show below why such assumptions, although seemingly contradictory at first sight, are actually very relevant in current wireless networks, and even more in future 5G-and-beyond networks. Yet, let us first try to convey the main intuition before diving in a more detailed and precise discussion: In many scenarios of interest, the latency constraint for data delivery is significantly looser than the CSI outdating constraint (which is related to the coherence time, and hence very short in many relevant mobility scenarios). This has for consequence that the data caching or sharing between TXs can be achieved in practice while *timely* CSI acquisition and sharing becomes the main bottleneck.

Imperfect CSI Acquisition and Sharing

We consider in this work a TX-dependent limited feedback where each TX receives its own multi-user imperfect estimate. This CSIT configuration is coined as the *Distributed CSI*

configuration (D-CSIT), in opposition to the aforementioned centralized CSIT model. The imperfect multi-user channel estimate is obtained at the TXs from a CSI acquisition and sharing mechanism not discussed in this work. Yet, due to unavoidable delay and imperfections in the CSI sharing mechanism, this CSI sharing step leads to a setting where the TXs have received different imperfect estimates of the true channel. We provide below several practical examples that are illustrated in Fig. 1.

Example 1. *In a network with several Base Stations (TXs) cooperating to jointly serve their users, each TX obtains feedback from its attached user such that each TX knows accurately only a part of the channel state information. For example, in Time Division Duplexing (TDD) transmission with reciprocity, each TX acquires a good estimate of its local channel; instead, in Frequency Division Duplexing (FDD) with user feedback, each TX obtains a good estimate of the whole channel vector towards its attached user.*

Considering a CSI exchange step with heterogeneous links in the sense that one direction is of better quality, requires less quantization, or introduces less delay (which is a very reasonable assumption when considering heterogeneous networks where some of the TXs are UAVs [11], [20] or vehicles [21]), we obtain a setting where one of the TXs is uniformly more informed than the other. A particular example is depicted in Fig. 1.a, in which the link between TX 2 and TX 3 is a limited Device-to-Device (D2D) link. □

Example 2. *Let us consider the previous setting in the case where the CSI exchange is limited by a very restrictive delay. Then, the sharing can be done by a transmission of the accurate CSI to a specific (main) TX, which then forwards a coarser version of the whole channel matrix to all TXs due to delay constraints. This retransmission from the main TX could also be broadcast, such that the resources spent on the sharing are reduced. Using layered encoding [22], every TX would obtain an estimate with a different accuracy. This setting is shown in Fig. 1.b.* □

Example 3. *In a wireless network with one principal TX receiving feedback from all served users and several remote radio heads helping in the joint transmission, a distributed CSIT configuration is obtained when the CSI sharing from the main TX to the remote radio heads is done using limited and imperfect communication links, as illustrated in Fig. 1.c. It could also be envisioned that the remote radio heads directly acquire low precision channel state information from direct feedback from the users by means of layered encoding [22] or analog feedback [23]. Due to the weaker capabilities of the remote radio heads, a Distributed CSIT (D-CSIT) configuration with homogeneous quality at each TX would then be obtained.* □

The analysis and the design of this CSI acquisition mechanism is an important research topic per-se and is consequently out of the scope of this paper and not further discussed. We assume in the following that this CSI acquisition step has already occurred through limited and imperfect communication links and has led to each TX having access to its own imperfect estimate of the multi-user channel state.

User's Data Symbols Caching

Our second main assumption comes from the fact that the user's data symbols are available at all TXs. This is made possible without putting in question the scenario described before because of two recent major techniques envisioned for future 5G-and-beyond wireless networks: Caching [24]–[26] and Cloud-Ran/Fog-Ran [27], [28].

Through caching, the user's data symbols are pre-fetched at the TX nodes before the transmission occurs [25]. Caching is an increasingly important feature that already exists in many scenarios [29], [30] and is envisioned in many more [31]. With the user's data symbols available at the TXs, even at mobile and cost-efficient ones, the accurate and timely acquisition of the multi-user channel becomes the main bottleneck for efficient interference reduction. This leads to a D-CSIT configuration wherever the cooperation links are not of sufficient quality.

In the Cloud-RAN paradigm, the centralization of the processing of all nodes is envisioned so as to gain full benefits of cooperation. This centralization is however limited by its cost and its delay, such that partial centralization is considered a promising solution [32]. Considering decentralized precoding at the TXs allows to reduce the *delay* in CSI acquisition. In that case, the backhaul links are solely used to convey the user's data since, for many data-oriented applications, the application's latency requirements are orders of magnitude slower than the rate at which the fading channel evolves. The CSI estimates are directly exchanged between the TXs through direct links, thus reducing the delay of the complete CSI acquisition at the TXs. This CSI exchange between TXs through limited resources also leads to a D-CSIT configuration.

Previous Works

In [33], the finite-SNR performance of regularized Zero-Forcing (ZF) under distributed CSIT has been computed in the large system limit, and heuristic robust precoding schemes have been provided in [34], [35] for practical cellular networks. In terms of DoF, it was shown in a previous work [19] that using a conventional ZF precoder (regularized or not) leads to a severe DoF degradation caused by the lack of a consistent CSI shared by the cooperating TXs. Exploiting the fact that there is a single interfered user in the two-user case, it was shown that exploiting the CSIT at a single TX was sufficient to completely suppress the interference, thus recovering from the DoF loss due to the D-CSIT setting. This result was extended in [36] to the Generalized DoF (GDoF) [37], where it was shown that the 2-user case centralized GDoF is recovered for any path-loss topology *and CSIT allocation*. In [38], a Deep Learning based robust precoder was proposed. This approach has been then extended in [39] to settings with a cooperation link between TXs and further generalized in [40] to arbitrary topologies with local CSI at the TXs. An overview of the use of Machine Learning for the physical layer of wireless communications with a particular focus on such decentralized applications can be found in [41].

The extension of the high SNR results for the two-user setting to an arbitrary number of users is not straightforward and has remained an open problem for several years. Tackling this challenging problem is the focus of this work.
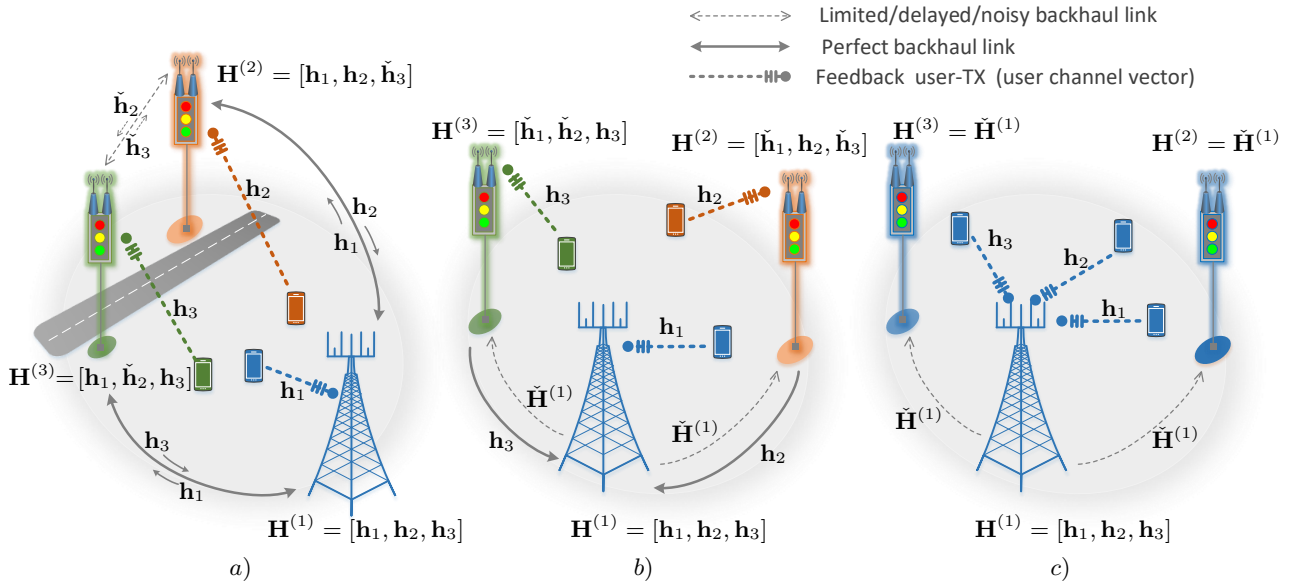
Figure 1: Schematic illustration of three different example scenarios with distributed CSIT. FDD transmission is assumed. $\mathbf{h}_i$ denotes the (highly accurate) estimate of user i's channel that is fed back to the attached TX. $\mathbf{H}^{(j)}$ denotes the CSI matrix obtained at TX j after cooperation. $\check{\mathbf{H}}^{(j)}$ and $\check{\mathbf{h}}_i$ represent a quantized/coarse version of the respective estimates, transmitted through limited backhaul communications.

C. Main Contributions

Our main contributions read as follows:

- We compute the DoF of a genie-aided setting where all TXs are given the knowledge of all the channel estimates at all TXs, the so-called *centralized upper bound*.
- We show the surprising result that this bound is actually tight for a range of D-CSIT configurations, coined the *Weak-CSIT regime* and defined rigorously further down. Interestingly, the optimal DoF for such D-CSIT regime only depends on the CSIT quality at the most informed TXs. Sharing the instantaneous CSIT among the TXs is hence not necessary for achieving the genie-aided centralized DoF and does not improve the optimal DoF.
- Building on the fundamental principles of the previous scheme, we derive a robust transmission scheme adapted to any CSIT configuration and any number of users, and which significantly improves the achieved DoF with respect to state-of-the-art methods.

A byproduct of our work which completes our main contributions is the development of new methods used as building blocks to our main algorithm, and which are of interest by themselves for other applications. The first one is the non-trivial extension of the Active-Passive precoding scheme introduced in [19]. Specifically, we extend the scheme to multiple passive TXs which turns out to be essential to transmit multi-stream transmissions to a single-antenna user and hence create an overloaded transmission. The second method is the translation to the distributed CSIT setting of the idea introduced by Maddah-Ali and Tse in [5] consisting in estimating and retransmitting the interference generated. In contrast to [5], the interference terms are estimated *before* they even take place and are retransmitted in the same time slot. This principle could be applied in other wireless configurations

where some nodes are more informed than others.

Notations: We denote the probability density function of a variable $\mathcal{X}$ as $f_{\mathcal{X}}(x)$. For a conditional probability density function, the simplified notation $f_{\mathcal{X}|\mathcal{Y}}(x|y)$ stands for $f_{\mathcal{X}|\mathcal{Y}}(\mathcal{X} = x|\mathcal{Y} = y)$. The circularly symmetric complex Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$. We use $\doteq$ to denote exponential equality, i.e., we write $f(P) \doteq P^x$ to denote $\lim_{P \to \infty} \frac{\log f(P)}{\log P} = x$. The exponential inequalities $\dot{\leq}$ and $\dot{\geq}$ are defined in the same way. We also consider the conventional Landau notation with a slight variation, such that $f(x) = \mathcal{O}(g(x))$ stands for $\lim_{x \to \infty} \frac{|f(x)|}{|g(x)|} = a$, with $0 < a < \infty$, and $f(x) = o(g(x))$ stands for $\lim_{x \to \infty} \frac{|f(x)|}{|g(x)|} = 0$. We use the shorthand notation $\mathcal{K} \triangleq \{1, \ldots, K\}$. Given a matrix $\mathbf{A}$, the sub-matrix obtained by taking from $\mathbf{A}$ the rows $\{r_i, \ldots, r_j\}$ and the columns $\{c_i, \ldots, c_j\}$ is denoted by $\mathbf{A}_{r_i:r_j, c_i:c_j}$.

II. SYSTEM MODEL

A. Transmission Model

We study a communication system where K TXs jointly serve K users (RXs) over a Network (Broadcast) MISO channel. We assume a D-CSIT configuration, where each TX has access to its own CSIT. Hereinafter, we denote this setting as the K-user MISO BC with distributed CSIT. We consider that both TXs and RXs are equipped with a single antenna. The assumption of single-antenna TX is done for ease of exposition, and the extension to multiple-antenna TX is straightforward. The signal received at RX i is written as

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

where $\boldsymbol{h}_i^{\mathrm{H}} \in \mathbb{C}^{1 \times K}$ is the channel vector towards RX i, $\mathbf{x} \in \mathbb{C}^{K \times 1}$ is the transmitted multi-user signal, which satisfies an average power constraint $\mathbb{E}[\|\mathbf{x}\|^2] \leq P$, and P is the

maximum transmit power. $z_i \in \mathbb{C}$ is the additive noise at RX i, independent of the channel and the transmitted signal, and distributed as $\mathcal{N}_{\mathbb{C}}(0,1)$. We further define the channel matrix $\mathbf{H} \triangleq [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]^{\mathrm{H}} \in \mathbb{C}^{K \times K}$ and the channel coefficient from TX k to RX i as $\mathbf{H}_{i,k}$. The channel is assumed to be drawn from a continuous distribution with density such that all the channel matrices and all their sub-matrices are full rank with probability one. Furthermore, the channel coefficients are assumed to change after one channel use and to be independent from one channel use to another.

The transmitted multi-user signal $\mathbf{x}$ is obtained from the symbol vector $\boldsymbol{s} \in \mathbb{C}^{b \times 1}$ having its elements independently and identically distributed (i.i.d.) according to $\mathcal{N}_{\mathbb{C}}(0,1)$, where b is the number of independent data symbols delivered. We will differentiate in this work between the *private* data symbols, destined to be decoded at a particular user, and the *common* data symbols, broadcast and destined to be decoded at all users. Note that the term private is used only in contrast to common and does not refer to any privacy/secrecy constraint, but to the fact that only one user will decode the symbol.

B. Distributed CSIT Model

The Distributed CSIT (D-CSIT) setting differs from the conventional centralized one in that each TX receives a possibly different multi-user CSI, based on which it designs its own transmission parameters without any additional communication with the other TXs. Specifically, TX j receives the imperfect multi-user channel estimate $\hat{\mathbf{H}}^{(j)} = [\hat{\boldsymbol{h}}_1^{(j)}, \ldots, \hat{\boldsymbol{h}}_K^{(j)}]^{\mathrm{H}} \in \mathbb{C}^{K \times K}$, where $(\hat{\boldsymbol{h}}_i^{(j)})^{\mathrm{H}}$ refers to the estimate at TX j of the channel from all TXs towards RX i. TX j then designs its transmit coefficients solely as a function of $\hat{\mathbf{H}}^{(j)}$ and the statistics of the channel. This scenario can be seen as a multi-agent cooperative decision with common goal, where each node knows the structure of the system but not the information that the others own [42].

Remark 1. *It is critical to this work to understand well how the* Distributed *CSIT setting differs from the many different heterogeneous CSIT configurations studied in the literature. Indeed, a heterogeneous CSIT configuration typically refers to a centralized CSIT setting (i.e., with a channel estimate common to all TXs), where each element of the channel is known with a different quality owing to specific feedback mechanisms [43]–[46]. In contrast, the distributed setting considered here has as many different channel estimates as TXs, and each TX does not have access to the CSIT knowledge at the other TXs.* □

We follow the conventional model in the literature [6]–[8], [47] to model the dependency of the CSIT accuracy as a function of the SNR, such that the channel estimate at TX j is given by

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

where $\bar{P} \triangleq \sqrt{P}$ and P is the nominal SNR parameter. $\boldsymbol{\Delta}^{(j)} \in \mathbb{C}^{K \times K}$ is a random variable with zero mean and bounded covariance matrix. The scalar $\alpha^{(j)}$ is called the *CSIT scaling coefficient* at TX j and represents the average accuracy of the estimate at TX j. Following insights obtained from
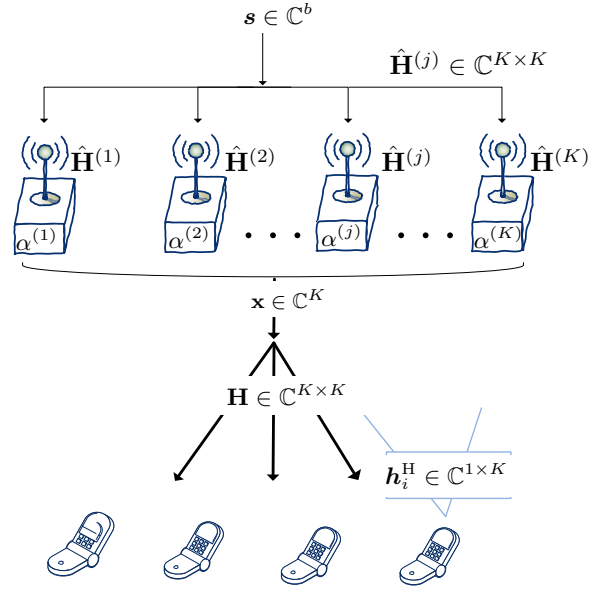

Figure 2: $K \times K$ Distributed Broadcast Channel. The accuracy of the channel estimate at TX j is modeled through the CSIT scaling coefficient $\alpha^{(j)}$.

the analysis of the centralized CSI configuration [4], [48], we focus in this work on the configurations where $\alpha^{(j)} \in [0,1]$.

Remark 2. *For the centralized case, it is shown in [4] that, if the CSI accuracy does not improve polynomially with the SNR (i.e., a CSIT scaling coefficient equal to 0), the channel estimate leads to no DoF improvement. Similarly, a CSIT scaling coefficient equal to 1 is shown to be sufficient to attain the DoF obtained with perfect CSIT. These results are in fact very intuitive when considering the scaling of the interference.* □

For later use, we also denote the i-th row of $\boldsymbol{\Delta}^{(j)}$ as $(\boldsymbol{\delta}_i^{(j)})^{\mathrm{H}}$, such that

$$\hat{\boldsymbol{h}}_i^{(j)} = \boldsymbol{h}_i + \bar{P}^{-\alpha^{(j)}} \boldsymbol{\delta}_i^{(j)}. \tag{3}$$

In this work, we restrict the D-CSIT model to the case of *homogeneous CSIT quality* at a given TX, such that every channel coefficient is known at TX j with the same average CSIT quality $\alpha^{(j)}$, as it can be seen in (2). This limitation is not inherent to the D-CSIT model and it is solely done here for simplicity. How to deal with different CSIT qualities for the different channel coefficients is a topic of undergoing research, and it is out of the scope of this work.

Given that one TX has the same CSIT quality ($\alpha^{(j)}$) for all channel coefficients, we can order the TXs without loss of generality such that

$$1 \geq \alpha^{(1)} \geq \alpha^{(2)} \geq \cdots \geq \alpha^{(K)} \geq 0. \tag{4}$$

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

$$\boldsymbol{\alpha} \triangleq \begin{bmatrix} \alpha^{(1)}, & \ldots, & \alpha^{(K)} \end{bmatrix}^{\mathrm{T}}. \tag{5}$$

The parameters $\boldsymbol{\alpha}$ represent the average accuracy of the estimates. They are long-term coefficients that vary slowly in time and can be easily shared to all TXs. Consequently, the parameters $\boldsymbol{\alpha}$ are assumed in the following to be fixed and known at all TXs.

Our upper bound builds on the centralized upper bound of [4]. On that account, and for the sake of completeness, we recall the Bounded Density assumption introduced in [4], which is necessary for the proof of the upper bound.

Definition 1 (Bounded Density Coefficients [4]). *A set of random variables, $\mathcal{A}$, is said to satisfy the bounded density assumption if there exists a finite positive constant $f_{\max}$,*

$$0 < f_{\max} < \infty, \tag{6}$$

such that for all finite cardinality disjoint subsets $\mathcal{A}_1$, $\mathcal{A}_2$ of $\mathcal{A}$, $\mathcal{A}_1 \subset \mathcal{A}$, $\mathcal{A}_2 \subset \mathcal{A}$, $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$, $|\mathcal{A}_1| < \infty$, $|\mathcal{A}_2| < \infty$, the conditional probability density functions exist and are bounded as follows

$$\forall A_1, A_2, \quad f_{\mathcal{A}_1|\mathcal{A}_2}(A_1|A_2) \le f_{\max}^{|\mathcal{A}_1|}. \tag{7}$$

We assume hereinafter that all channel realizations $\mathbf{H}_{i,k}$ and estimation noise variables $\boldsymbol{\Delta}_{i,k}^{(j)}$ satisfy the bounded density property. Furthermore, the channel realizations and the estimation noise are mutually independent.

C. CSIR Model

In this work, we focus on the impact of the imperfect CSI on the TX side as the CSI acquisition is widely acknowledged to be more challenging on the TX side than on the RX side in FDD, because the CSI that has been estimated at the RX needs to be fed back towards the TX. Therefore, we consider that every RX has perfect knowledge of its own channel. As in the important literature on delayed CSIT [5]–[10], to name just a few, we assume that the RX has been able to obtain perfect knowledge of the multi-user channel, i.e., also the channel to the other users. This assumption is key to the approach used in this work.

However, it is also important to note that this assumption can be weakened in our work as it is sufficient for the RXs to obtain the multi-user CSIT up to the best CSIT quality across the TXs (not necessarily the same estimate, but of the same quality). Furthermore, the estimate should be made available at the RX for the decoding, such that its latency constraint stems from the user's data, not from the time coherency of the channel.

D. Degrees-of-Freedom Analysis

We assume that every user $i \in \mathcal{K}$ wishes to receive message $W_i \in \mathbb{W}_i$. After n channel uses, the rate $R_i(P)$ is achievable for RX i if $R_i(P) = \frac{\log |\mathbb{W}_i|}{n}$ and the probability of wrong decoding goes to zero as n goes to infinity. The sum capacity $\mathcal{C}(P)$ is defined as the supremum of the sum of all achievable rates [49]. The optimal sum DoF in the D-CSIT setting with CSIT scaling coefficients $\boldsymbol{\alpha} \in \mathbb{R}^K$ is defined by

$$\mathrm{DoF}^{\mathrm{DCSI}}(\boldsymbol{\alpha}) \triangleq \lim_{P \to \infty} \frac{\mathcal{C}(P)}{\log_2(P)}. \tag{8}$$

III. MAIN RESULTS

As a preliminary, let us first state the optimal DoF of the centralized K-user BC setting where a single estimate with CSIT scaling coefficient α is perfectly shared by all TXs. It was shown in [4] that the sum DoF in that configuration, denoted by $\mathrm{DoF}^{\mathrm{CCSI}}(\alpha)$, is equal to

$$\mathrm{DoF}^{\mathrm{CCSI}}(\alpha) = 1 + (K-1)\alpha. \tag{9}$$

We can now present our main results.

A. Upper Bound

Theorem 1. *The optimal sum DoF of the K-user MISO BC with Distributed CSIT satisfies*

$$\mathrm{DoF}^{\mathrm{DCSI}}(\boldsymbol{\alpha}) \le \mathrm{DoF}^{\mathrm{CCSI}}(\alpha^{(1)}). \tag{10}$$

Proof: The proof relies on the following lemma whose technical proof is relegated to Appendix II for clarity.

Lemma 1. *Let $\hat{\mathcal{H}}^{(j)} \triangleq \mathcal{H} + \bar{P}^{-\alpha^{(j)}}\Delta^{(j)}$, where $\mathcal{H}$, $\Delta^{(j)}$, $\forall j \in \mathcal{K}$, are independent continuous random variables satisfying the Bounded Density assumption. Then, the conditional probability density function $f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)},\dots,\hat{\mathcal{H}}^{(K)}}$ satisfies that*

$$\max_{\mathcal{H}} f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)},\dots,\hat{\mathcal{H}}^{(K)}} = \mathcal{O}\big(\bar{P}^{\max_{j \in \mathcal{K}} \alpha^{(j)}}\big). \tag{11}$$

Let us now consider a genie-aided scenario where all channel estimates are exchanged between all the TXs. Such setting corresponds to a (logically) centralized scenario with a shared CSI composed by $\{\hat{\mathbf{H}}^{(1)}, \dots, \hat{\mathbf{H}}^{(K)}\}$. Using Lemma 1, we obtain that the peak of the probability density function of this genie-aided scenario with multiple estimates has the same scaling as the centralized setting endowed only with $\hat{\mathbf{H}}^{(1)}$. Then, it follows directly from the proof in [4, Section V.8] that the DoF of the genie-aided scenario, denoted by $\mathrm{DoF}^{\mathrm{CCSI}}_{genie}(\boldsymbol{\alpha})$, is given by

$$\mathrm{DoF}^{\mathrm{CCSI}}_{genie}(\boldsymbol{\alpha}) = \mathrm{DoF}^{\mathrm{CCSI}}(\alpha^{(1)}). \tag{12}$$

From this equivalence, and the fact that providing with more information does not hurt, the proof is concluded. ∎

Remark 3. *Lemma 1 is expected to hold for a more general group of distributions, i.e., including cases where the different noise variables are partially correlated. Indeed, for the Gaussian case where the noise variables $\{\Delta_{i,k}^{(j)}\}_{\forall j \in \mathcal{K}}$ are drawn from partially correlated jointly Gaussian distributions, it is easy to show analytically that (11) is also satisfied.* □

B.1) Achievability: Weak-CSIT Regime

Theorem 2. *Let us assume that the m first TXs have the same accuracy, i.e., $\alpha^{(1)} = \cdots = \alpha^{(m)}$, $m < K$. Then, the sum DoF of the K-user MISO BC with Distributed CSIT satisfies*

$$\mathrm{DoF}^{\mathrm{DCSI}}(\boldsymbol{\alpha}) \ge \mathrm{DoF}^{\mathrm{CCSI}}(\alpha^{(1)}) \tag{13}$$

if $\alpha^{(1)} \le \alpha_m^{\mathrm{Weak}}$, where α_m^{Weak} is defined as

$$\alpha_m^{\mathrm{Weak}} \triangleq \frac{1}{1 + K(K-m-1)}, \tag{14}$$

and is called the "m-TX Weak-CSIT" regime. For $m = 1$, we simplify the notation and call it the "Weak-CSIT" regime.

Proof: The result follows directly from the analysis of the proposed scheme presented in detail in Section V. ∎

In the so-called Weak-CSIT regime, the upper bound of Theorem 1 is tight. Surprisingly, for $m = 1$, the most heterogeneous case, the DoF depends only on the CSI quality of TX 1, although with the downside of reducing the range of possible CSIT configurations. For $m = K - 1$, it holds that $\alpha_{K-1}^{\text{Weak}} = 1$ and hence the Weak-CSIT regime encloses every possible value of $\alpha^{(1)}$, which is consistent with the simple use of single-stream Active-Passive Zero-Forcing (AP-ZF) in [19].

Remark 4. *The fact that it is possible to achieve the DoF of the centralized upper bound with uniformed or badly informed TXs is a surprising result which is not expected to extend to many other CSIT configurations. Indeed, for linear schemes, it can be intuitively seen using basic linear algebra that at least $K - 1$ TXs are necessary to cancel $K - 1$ ZF constraints.* □

B.2) Achievability: Extension to Arbitrary CSIT Configurations

Theorem 2 presents the CSIT regime for which the upper bound of Theorem 1 is tight. In the following, we present a robust transmission scheme that builds on the scheme attaining Theorem 2 but which is extended to adapt to any CSIT configuration. The main challenge comes from the large number of CSIT scaling parameters, which leads to an even larger (combinatorially large) number of possible CSIT configurations depending on their relative values. Before presenting the achievability result, we define three terms that play an important role in the proposed scheme.

Definition 2 (Transmitting TX). *A TX is said to be a "Transmitting TX" if it is connected and sends information to the RXs. It may or may not use its instantaneous CSI for precoding.*

This definition is made necessary by the distributed nature of the CSIT. Indeed, in contrast to the centralized setting where adding antennas cannot reduce the performance [45], [50], using additional antennas in the distributed setting can decrease the achievable DoF by creating additional interference. Hence, although we are considering a setting with K TXs, it may be beneficial in some CSI configurations to "turn off" some TXs and use a smaller number of "Transmitting TXs".

Definition 3 (Active TX). *A TX is said to be an "Active TX" if it is a Transmitting TX and it makes use of its instantaneous CSI.*

Definition 4 (Passive TX). *A TX is said to be a "Passive TX" if it is a Transmitting TX but it does not make use of its instantaneous CSI.*

A more thorough explanation of these definitions and their relevance is provided later on, along with the description of the proposed scheme. In brief, it will become clear that the two critical parameters to optimize are both the number of "Transmitting TXs" and the number of "Active TXs". In relation to these two notions, we introduce the following definition.

Definition 5 (Transmission Mode (n, k)). *We define the Transmission Mode (n, k) as a transmission where only k Transmitting TXs and $n \leq k$ Active TXs are used.*

Building on these definitions, the following lower bound is exactly obtained by optimizing the performance of the proposed scheme over the different Transmission Modes.

Theorem 3. *The sum DoF of the K-user D-CSIT BC with CSIT scaling coefficients $\boldsymbol{\alpha}$ is lower-bounded by the solution of the next linear program, which we denote as $\mathrm{DoF}^{\mathrm{APZF}}(\boldsymbol{\alpha})$:*

$$\underset{\gamma_{n,k}}{\text{maximize}} \quad \sum_{k=2}^{K} \sum_{n=1}^{k-1} \gamma_{n,k} \big(1 + (k-1)\alpha^{(n)} \big) \quad (15)$$

$$\text{subject to} \quad \sum_{k=2}^{K} \sum_{n=1}^{k-1} \gamma_{n,k} = 1, \quad \gamma_{n,k} \geq 0 \quad (16)$$

$$\sum_{k=2}^{K} \sum_{n=1}^{k-1} d_{n,k} \gamma_{n,k} \geq 0, \quad (17)$$

where $\gamma_{n,k}$ is a time-sharing variable representing the percentage of time allocated to the Transmission Mode (n, k), and $d_{n,k} \triangleq 1 - \alpha^{(n)} - k(k - n - 1)\alpha^{(n)}$.

Proof: The transmission scheme for a particular Transmission Mode is described in Section V and, building on this result, the explanation and proof of the theorem is given in Section VI. ∎

The transmission scheme and the achieved DoF are obtained by solving a simple linear programming problem with low complexity. Interestingly, the expression $1 + (k-1)\alpha^{(n)}$ in (15) corresponds to the DoF of the k-user centralized setting with k TXs having access to a CSIT of quality $\alpha^{(n)}$ (See (9)).

Remark 5. *The linear program of Theorem 3 depends only on the $K - 1$ best CSIT coefficients and not on $\alpha^{(K)}$. This property was already highlighted in [19] and follows from the fact that it is possible to solve $K - 1$ linear equations with $K - 1$ Active TXs, which means removing interference at $K - 1$ users, and thus serving K users at the same time. As consequence, it can always be assumed that $\alpha^{(K)} = 0$ without reducing the DoF.* □

Furthermore, we can show that time sharing between only two Transmission Modes is sufficient.

Corollary 1. *The solution of the linear program in Theorem 3 is composed only of two Transmission Modes, (n_1, k_1) and (n_2, k_2), such that $\gamma_{n_1,k_1} > 0$, $\gamma_{n_2,k_2} \geq 0$, and $\gamma_{n,k} = 0$ for any pair $(n, k) \notin \{(n_1, k_1), (n_2, k_2)\}$.*

Proof: The proof is relegated to Appendix III. ∎

Remark 6. *The Distributed CSIT setting here assumed and the Delayed CSIT setting share an important feature: It is not possible to cancel out all the interference. However, the means for solving this challenge at each setting differ considerably. For the Distributed CSIT setting, the best strategy consists on overloading the transmission so as to use the interference created as side information that is useful at different RXs. In addition, the more RXs receive interference, the more overloaded the transmission has to be. Corollary 1 reflects the main insight: If there are two modes of transmission, the first one is a generator of interference, i.e., it creates side-information at the RXs through the overloaded transmission.*
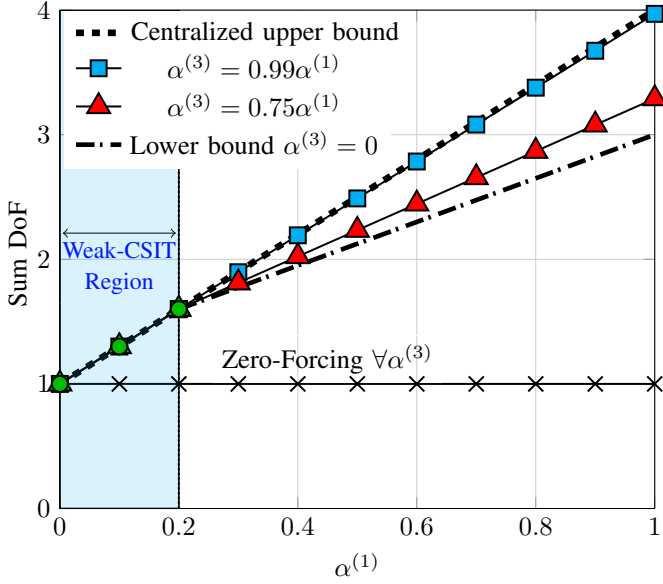
Figure 3: Sum DoF for the case with $K = 4$ TXs, $\alpha^{(2)} = \alpha^{(1)}$, and $\alpha^{(4)} = 0$, for different values of $\alpha^{(3)}$ as a function of $\alpha^{(1)}$.

Figure 4: Sum DoF of the $K = 4$ MISO BC with distributed CSIT as a function of $\alpha^{(2)}$, with $\alpha^{(1)} = 1$, $\alpha^{(3)} = \alpha^{(4)} = 0$.

Then, it relies on a successive second Transmission Mode to retransmit some of this interference (side information) in order to decode the overloaded transmission. When only one mode is used, the interference is directly retransmitted through rate splitting using the common data symbol. □

We show in Fig. 3 the DoF as a function of $\alpha^{(1)}$ for a setting with $K = 4$ TXs and $\alpha^{(2)} = \alpha^{(1)}$, and we compare the achievable DoF with the centralized upper bound for different values of $\alpha^{(3)}$. The centralized upper bound is achieved for any $\alpha^{(1)} \le 0.2 = \alpha_2^{\text{Weak}}$, no matter the value of $\alpha^{(3)}$ (as stated in Theorem 2). The upper bound is also achieved when $\alpha^{(3)}$ becomes equal to $\alpha^{(1)}$, which is consistent with the results in [19].

In Fig. 4, we show the DoF achieved by AP-ZF for a network with $K = 4$ TXs when fixing the number of Transmitting TXs (i.e., the value of k in Theorem 3) for the specific case where $\alpha^{(1)} = 1$, $\alpha^{(3)} = \alpha^{(4)} = 0$, and $\alpha^{(2)}$ varies from 0 to 1. Depending on the value of $\alpha^{(2)}$, it is optimal to use either 2 Transmitting TXs or 3 Transmitting TXs, but never $K = 4$ Transmitting TXs.

IV. A SIMPLE EXAMPLE

We present in the following a simple transmission scheme in a toy example so as to exemplify the key features of our approach and convey the main intuition in a clear manner. The full scheme achieving the DoF of Section III will be described in Section V.

We consider a 3-user setting with $\alpha^{(1)} = 0.1$, $\alpha^{(2)} = 0$, and $\alpha^{(3)} = 0$. The conventional regularized Zero-Forcing would achieve a DoF of 1 [1], which is the same as for the no CSIT scenario. We will show how it is possible to achieve a DoF of $1 + 2\alpha^{(1)} = 1.2$, which is the DoF that would be achieved in a centralized setting if TX 2 and TX 3 had received the same estimate as TX 1 [4], such that there is no DoF loss from *not* sharing the CSIT between the TXs.
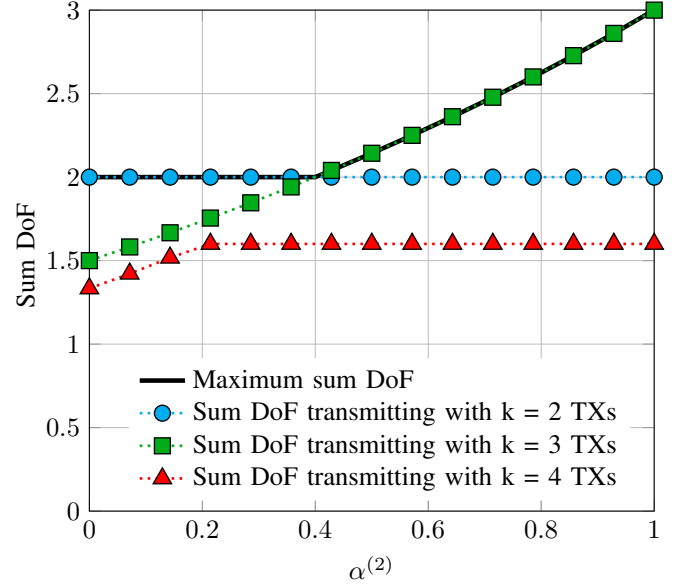
A. Encoding

The transmission scheme consists in a single channel use during which 3 private data symbols of $\alpha^{(1)} \log_2(P)$ bits are sent to each user, thus leading to 9 data symbols being sent in one channel use. Additionally, a common data symbol of rate $(1 - \alpha^{(1)}) \log_2(P)$ bits is broadcast from TX 1 to all users using superposition coding [49]. Note that the information contained in this common data symbol is not only composed of "fresh" information bits destined to one user, but is also composed of side information necessary for the decoding of the private data symbols, as will be detailed below. The transmitted signal $\mathbf{x} \in \mathbb{C}^3$ is then equal to

$$\mathbf{x} = \boldsymbol{s}_1 + \boldsymbol{s}_2 + \boldsymbol{s}_3 + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} s_0 \tag{18}$$

where

- $\boldsymbol{s}_i \in \mathbb{C}^3$ is a vector containing the 3 private data symbols destined to user i, each one with power $P^{\alpha^{(1)}}/9$ and rate $\alpha^{(1)} \log_2(P)$ bits.
- s_0 is the common data symbol transmitted *only from* TX 1 and destined to be decoded at all users, transmitted with power $P - P^{\alpha^{(1)}}$ and rate $(1 - \alpha^{(1)}) \log_2(P)$ bits.

The signal received at RX i, illustrated in Fig. 5, is given by

$$y_i = \underbrace{\mathbf{H}_{i,1} s_0}_{\doteq P} + \underbrace{\boldsymbol{h}_i^{\text{H}} \boldsymbol{s}_1}_{\doteq P^{\alpha^{(1)}}} + \underbrace{\boldsymbol{h}_i^{\text{H}} \boldsymbol{s}_2}_{\doteq P^{\alpha^{(1)}}} + \underbrace{\boldsymbol{h}_i^{\text{H}} \boldsymbol{s}_3}_{\doteq P^{\alpha^{(1)}}} \tag{19}$$

where the power scaling is written under the bracket, and where the noise term has been neglected for clarity.

B. Interference Estimation and Quantization at TX 1

The key element of the scheme is that the common data symbol s_0 is used to convey side information, enabling
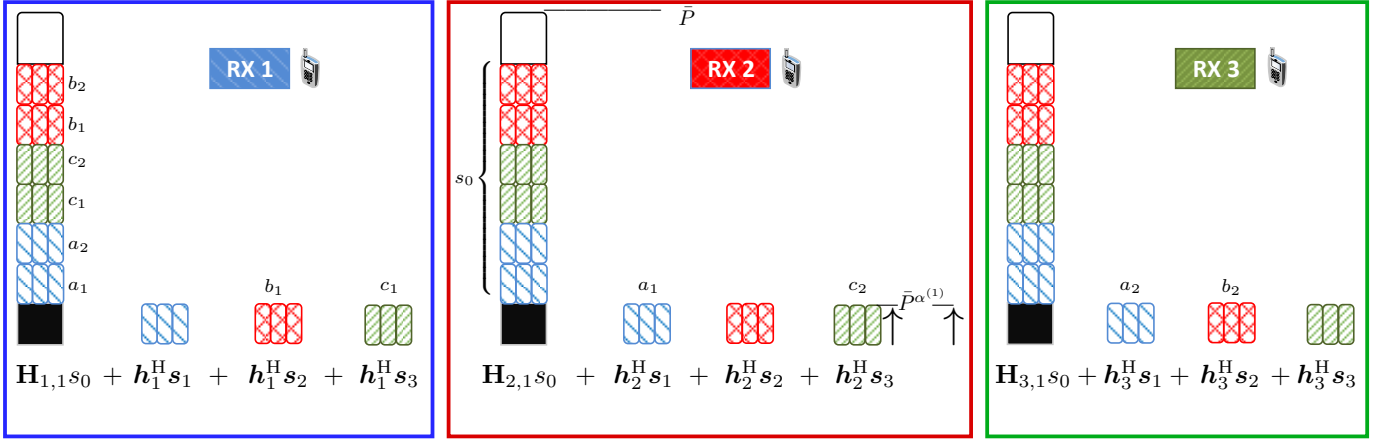
Figure 5: Illustration of the received signals. Each RX receives its desired private data symbols and interference scaling both in $\bar{P}^{\alpha^{(1)}}$. Through superposition coding, it also receives the common data symbol s_0 containing a mix of fresh desired data symbols (illustrated in white), and side information to remove interference (illustrated with the color of the relevant RX).

each user to decode its desired private data symbols. More specifically, TX 1 uses its local CSIT $\hat{\mathbf{H}}^{(1)}$ to *estimate* the interference terms $(\hat{\boldsymbol{h}}_i^{(1)})^{\mathrm{H}}\boldsymbol{s}_k$ generated by the first layer of transmission, for any i, k such that $k \neq i$. Then, TX 1 quantizes and transmits these interference terms using the common data symbol s_0. Each interference term has a variance scaling in $\bar{P}^{\alpha^{(1)}}$ and is quantized using $\alpha^{(1)} \log_2(P)$ bits, such that the quantization noise can be made to remain at the noise floor using an appropriate uniform or Lloyd quantizer [49]. In total, the transmission of all the quantized estimated interference requires a transmission of $6\alpha^{(1)} \log_2(P)$ bits.

These $6\alpha^{(1)} \log_2(P)$ bits can be transmitted via the data symbol s_0 if $6\alpha^{(1)} \log_2(P) \leq (1-\alpha^{(1)}) \log_2(P)$, which is the case for the example considered here since $6 \times 0.1 < 1 - 0.1$. If the inequality is strict (as it is in this case), s_0 carries some additional $(1 - 7\alpha^{(1)}) \log_2(P)$ fresh-information bits to any particular user.

C. Decoding and DoF Analysis

It remains to verify that this transmission scheme achieves the claimed DoF. Let us consider w.l.o.g. the decoding at RX 1, because the decoding at the other users is the same up to a circular permutation of the RX's indices. Note that signals at the noise floor are systematically omitted.

Using successive decoding [49], the common data symbol s_0 is decoded first, followed by the private data symbols $\boldsymbol{s}_1$. The data symbol s_0 of rate of $(1 - \alpha^{(1)}) \log_2(P)$ bits can be decoded with a vanishing probability of error as its effective SNR can be seen in (19) to scale in $P^{1-\alpha^{(1)}}$. Upon decoding s_0, the quantized estimated interference $(\hat{\boldsymbol{h}}_1^{(1)})^{\mathrm{H}}\boldsymbol{s}_2$ is obtained up to the quantization noise. RX 1 has then decoded

$$\underbrace{(\hat{\boldsymbol{h}}_1^{(1)})^{\mathrm{H}}\boldsymbol{s}_2}_{\doteq P^{\alpha^{(1)}}} = \boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{s}_2 + \underbrace{\bar{P}^{-\alpha^{(1)}}(\boldsymbol{\delta}_1^{(1)})^{\mathrm{H}}\boldsymbol{s}_2}_{\doteq P^0}. \quad (20)$$

Since the quantization noise is at the noise floor, it is neglected in the following. This means that the interference term $\boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{s}_2$ can be suppressed up to the noise floor at RX 1. Proceeding in the same way with $(\hat{\boldsymbol{h}}_1^{(1)})^{\mathrm{H}}\boldsymbol{s}_3$, the remaining signal at RX 1 is

$$y_1 = \boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{s}_1. \quad (21)$$

RX 1 can form a virtual received vector $\boldsymbol{y}_1^{\mathrm{v}} \in \mathbb{C}^3$ by bringing together the signal in (21) and the interference terms $(\hat{\boldsymbol{h}}_2^{(1)})^{\mathrm{H}}\boldsymbol{s}_1$ and $(\hat{\boldsymbol{h}}_3^{(1)})^{\mathrm{H}}\boldsymbol{s}_1$ obtained through s_0. Therefore, $\boldsymbol{y}_1^{\mathrm{v}} \in \mathbb{C}^3$ is defined as

$$\boldsymbol{y}_1^{\mathrm{v}} \triangleq \begin{bmatrix} \boldsymbol{h}_1^{\mathrm{H}} \\ (\hat{\boldsymbol{h}}_2^{(1)})^{\mathrm{H}} \\ (\hat{\boldsymbol{h}}_3^{(1)})^{\mathrm{H}} \end{bmatrix} \boldsymbol{s}_1. \quad (22)$$

Each component of $\boldsymbol{y}_1^{\mathrm{v}}$ has an effective SNR scaling in $P^{\alpha^{(1)}}$ such that RX 1 can decode with a vanishing error probability its 3 data symbols of rate $\alpha^{(1)} \log_2(P)$ bits.

If we take into account the 3 users, $9\alpha^{(1)} \log_2(P)$ bits have been transmitted through the private data symbols and $(1 - 7\alpha^{(1)}) \log_2(P)$ bits through the common data symbol s_0, which yields a sum DoF of $1 + 2\alpha^{(1)}$.

Remark 7. *Interestingly, the above scheme is based on interference estimation, quantization, and retransmission, in a similar fashion as done in the different context of precoding with delayed CSIT (see e.g. [6]–[8]). Yet, we exploit in this work the distributed nature of the CSIT instead of the delayed knowledge of the CSIT, such that the scheme estimates and quantizes the interference before even being generated.* □

V. TRANSMISSION MODE (n, k) WITH n ACTIVES TXS AND k TRANSMITTING TXS

We present in this section the Transmission Mode (n, k) with n Active TXs and k Transmitting TXs. We describe the main structure of the transmission in Section V-A, before describing in detail the precoding scheme in Section V-B. The received signal is then studied in Section V-C, and the achieved DoF is computed in Section V-D and Section V-E. As main insight, a transmission using n *Active TXs* can reduce the interference power received at n RXs by a factor $P^{\alpha^{(n)}}$, i.e., as n increases, it holds that:

- The interference power is reduced by a *smaller* factor.
- The interference power is reduced at *more* RXs.

A. Encoding

The Transmission Mode (n, k) only considers k Transmitting TXs, and thus only k RXs are served simultaneously. Let $\mathcal{U}$ denote the set of RXs served and let us assume w.l.o.g. that the served RXs are the k first users, such that $\mathcal{U} = \{1, \ldots, k\}$. The transmitted signal $\mathbf{x} \in \mathbb{C}^k$ is then given by

$$\mathbf{x} = \begin{bmatrix} 1 \\ \mathbf{0}_{k-1 \times 1} \end{bmatrix} s_0 \; + \; \sum_{i=1}^{k} \mathbf{T}_i^{\text{APZF}} \boldsymbol{s}_i \tag{23}$$

where

- $\boldsymbol{s}_i \in \mathbb{C}^{k-n}$ contains $k - n$ data symbols destined to RX i, which we hence denote as *private*, each one of rate $\alpha^{(n)} \log_2(P)$ bits and power $P^{\alpha^{(n)}}/(k(k-n))$, distributed in an i.i.d. manner. They are precoded with the AP-ZF precoder $\mathbf{T}_i^{\text{APZF}} \in \mathbb{C}^{k \times (k-n)}$ with n active TXs as described in detail in Section V-B.
- $\boldsymbol{s}_0 \in \mathbb{C}$ is a data symbol destined to be decoded at all users and that we hence denote as *common*, transmitted at rate $(1 - \alpha^{(n)}) \log_2(P)$ bits and power $P - P^{\alpha^{(n)}}$.

Note that $k-n$ data streams are sent to each RX, but each RX has only one antenna. This overloaded transmission is necessary to take advantage of the $k - n - 1$ interference terms generated by the RX's symbols at the other RXs, following the intuition from [5] that interference can be used as side information. This is detailed in Section V-C. A total of $k(k-n)\alpha^{(n)} \log_2(P)$ bits are sent in one channel use through the private data symbols. Furthermore, an additional data symbol of data rate $(1 - \alpha^{(n)}) \log_2(P)$ bits is broadcast from TX 1. Importantly, we will show that this common data symbol s_0 does not only contain new information bits, but also side information to enable the successful decoding of the private data symbols.

B. Precoding: AP-ZF with n Active TXs

The proposed precoder can be decoupled such that the precoder for each RX is computed independently of the other RXs up to a power normalization. We describe now the AP-ZF precoder serving a specific RX i with n Active TXs. This precoder allows us to transmit $k - n$ streams to RX i while reducing the interference at the n following RXs, i.e., at RXs $(i + t) \bmod [k] + 1$, $\forall t \in \{1, \ldots, n\}$. For ease of notation, we omit in the following the modulo operation as it is clear what an index bigger than k refers to. The precoder is obtained from distributed precoding at all TXs such that

$$\mathbf{T}_i^{\text{APZF}} = \begin{bmatrix} \boldsymbol{e}_1^{\text{T}} \mathbf{T}_i^{\text{APZF}(1)} \\ \boldsymbol{e}_2^{\text{T}} \mathbf{T}_i^{\text{APZF}(2)} \\ \vdots \\ \boldsymbol{e}_k^{\text{T}} \mathbf{T}_i^{\text{APZF}(k)} \end{bmatrix}, \tag{24}$$

where $\boldsymbol{e}_\ell^{\text{T}}$ refers to the ℓ-th row of the identity matrix $\mathbf{I}_{k \times k}$, and $\mathbf{T}_i^{\text{APZF}(j)}$ denotes the AP-ZF precoder computed at TX j. We will therefore consider the design of $\mathbf{T}_i^{\text{APZF}(j)}$ at TX j.

Remark 8. *Note that, although TX j computes the full precoder $\mathbf{T}_i^{\text{APZF}(j)}$, only some coefficients are effectively used for the transmission due to the distributed precoding configuration, as made clear in (24).* □
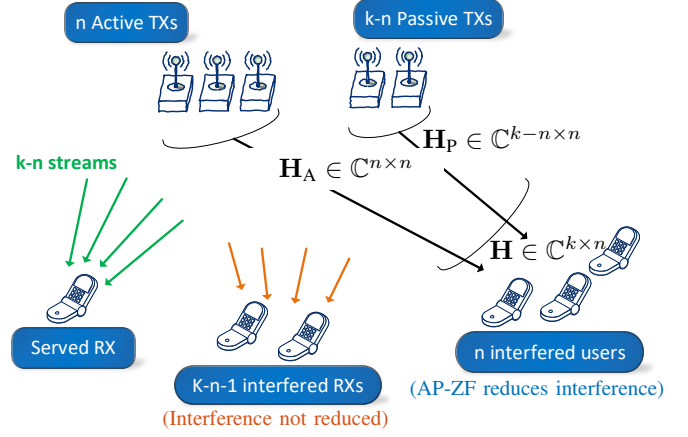

Figure 6: AP-ZF illustration: The number of Active TXs (n) determines the number of RXs at which the interference is reduced, whereas the number of Passive TXs ($k-n$) determines the number of data streams that each RX can receive.

As a preliminary, let us define the *Active Channel* $\mathbf{H}_{\text{A}} \in \mathbb{C}^{n \times n}$ as the channel coefficients from the Active TXs (TX 1 to TX n) to the RXs whose received interference is reduced (RX $i + 1$ to RX $i + n$), i.e.,

$$\mathbf{H}_{\text{A}} \triangleq \mathbf{H}_{i+1:i+n, 1:n}. \tag{25}$$

Similarly, the *Passive Channel* $\mathbf{H}_{\text{P}} \in \mathbb{C}^{n \times (k-n)}$ contains the channel coefficients from the Passive TXs (TX $n+1$ to TX k) to the RXs with reduced interference (RX $i + 1$ to RX $i + n$), i.e.,

$$\mathbf{H}_{\text{P}} \triangleq \mathbf{H}_{i+1:i+n, n+1:k}. \tag{26}$$

The Passive TXs do not use their instantaneous CSIT. Hence, the precoder applied at the Passive TXs (passive precoder) is an arbitrarily chosen deterministic full-rank matrix denoted as $\lambda_i^{\text{APZF}} \mathbf{T}_i^{\text{P}} \in \mathbb{C}^{(k-n) \times (k-n)}$, where λ_i^{APZF} is used to satisfy an average sum power constraint and is detailed further down.

On the other hand, every Active TX j, $\forall j \in \{1, \ldots, n\}$, computes $\mathbf{T}_i^{\text{APZF}(j)} \in \mathbb{C}^{k \times k-n}$ only on the basis of its own available CSIT $\hat{\mathbf{H}}^{(j)}$, such that

$$\mathbf{T}_i^{\text{APZF}(j)} = \lambda_i^{\text{APZF}} \begin{bmatrix} \mathbf{T}_i^{\text{A}(j)} \\ \mathbf{T}_i^{\text{P}} \end{bmatrix}. \tag{27}$$

The precoder applied at the Active TXs (active precoder) is denoted as $\mathbf{T}_i^{\text{A}(j)}$ and computed as

$$\mathbf{T}_i^{\text{A}(j)} = -\left((\hat{\mathbf{H}}_{\text{A}}^{(j)})^{\text{H}} \hat{\mathbf{H}}_{\text{A}}^{(j)} + \frac{1}{P} \mathbf{I}_n \right)^{-1} (\hat{\mathbf{H}}_{\text{A}}^{(j)})^{\text{H}} \hat{\mathbf{H}}_{\text{P}}^{(j)} \mathbf{T}_i^{\text{P}}. \tag{28}$$

Remark 9. *The design of the active precoder in (28) is an extension of the AP-ZF precoder introduced in [19]. Intuitively, the n Active TXs invert the channel to the n chosen RXs so as to cancel the interference generated by the Passive TXs. The number of Passive TXs limits the rank of the transmitted signal, while the number of Active TXs limits the number of users whose received interference is attenuated.* □

The AP-ZF precoder $\mathbf{T}_i^{\mathrm{APZF}} \in \mathbb{C}^{k \times k-n}$ practically used in the transmission and defined in (24) can be written as

$$\mathbf{T}_i^{\mathrm{APZF}} \triangleq \lambda_i^{\mathrm{APZF}} \begin{bmatrix} \boldsymbol{e}_1^{\mathrm{T}} \mathbf{T}_i^{\mathrm{A}(1)} \\ \vdots \\ \boldsymbol{e}_n^{\mathrm{T}} \mathbf{T}_i^{\mathrm{A}(n)} \\ \mathbf{T}_i^{\mathrm{P}} \end{bmatrix}, \quad (29)$$

where the normalization coefficient $\lambda_i^{\mathrm{APZF}}$ is chosen as

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

The normalization constant $\lambda_i^{\mathrm{APZF}}$ requires only statistical CSI and can hence be computed at every TX, even the passive ones. It ensures that an average sum-power normalization constraint is satisfied, i.e., that

$$\mathbb{E}\left[\|\mathbf{T}_i^{\mathrm{APZF}}\|_{\mathrm{F}}^2\right] = 1. \quad (30)$$

The fundamental property of AP-ZF is that it effectively achieves interference reduction at the n RXs up to the worst accuracy across the Active TXs, as stated in the following lemma.

Lemma 2. *The AP-ZF precoder with n Active TXs satisfies*

$$\|\boldsymbol{h}_\ell^{\mathrm{H}} \mathbf{T}_i^{\mathrm{APZF}}\|_{\mathrm{F}}^2 \dot{\leq} P^{-\alpha^{(n)}}, \ \forall \ell \in \{i+1, \ldots, i+n\}. \quad (31)$$

Proof: The proof of Lemma 2 is given in Appendix I along with the derivation of other important properties of AP-ZF precoding. ∎

C. Received Signals

The signal received at RX i is given by

$$y_i = \underbrace{\mathbf{H}_{i,1}s_0}_{\doteq P} + \underbrace{\boldsymbol{h}_i^{\mathrm{H}}\mathbf{T}_i^{\mathrm{APZF}}\boldsymbol{s}_i}_{\doteq P^{\alpha^{(n)}}} \\ + \underbrace{\boldsymbol{h}_i^{\mathrm{H}}\sum_{\ell \in \mathcal{I}_i^{\mathrm{APZF}}}\mathbf{T}_\ell^{\mathrm{APZF}}\boldsymbol{s}_\ell}_{\doteq P^{\alpha^{(n)}}} + \underbrace{\boldsymbol{h}_i^{\mathrm{H}}\sum_{\ell \in \mathcal{U}\setminus\mathcal{I}_i^{\mathrm{APZF}}}\mathbf{T}_\ell^{\mathrm{APZF}}\boldsymbol{s}_\ell}_{\doteq P^0}, \quad (32)$$

where the noise term has been neglected for clarity, and where $\mathcal{I}_i^{\mathrm{APZF}}$ is defined (omitting the modulo operation) as

$$\mathcal{I}_i^{\mathrm{APZF}} \triangleq \{i+1, \ldots, i+k-n-1\}. \quad (33)$$

Intuitively, the set $\mathcal{I}_i^{\mathrm{APZF}}$ contains the interfering signals that have not been attenuated towards RX i. The last term in (32) scales as P^0 due to the attenuation produced by the AP-ZF precoding, as shown in Lemma 2. In Fig. 7, we illustrate the received signal at every RX for $k = 3$ RXs and $n = 1$ Active TX. We can see the improvement with respect to Fig. 5 since the number of significant interference terms is reduced by half thanks to the AP-ZF precoding.

D. Decoding

TX 1 uses its local CSIT $\hat{\mathbf{H}}^{(1)}$ to estimate the interference terms $\boldsymbol{h}_i^{\mathrm{H}}\mathbf{T}_\ell^{\mathrm{APZF}}\boldsymbol{s}_\ell$, $\forall \ell \in \mathcal{I}_i^{\mathrm{APZF}}$. Each interference term scales in $P^{\alpha^{(n)}}$. Therefore, each term can be quantized with $\alpha^{(n)}\log_2(P)$ bits such that the quantization noise lies at the noise floor [49]. Considering all users, this means that $k(k-n-1)\alpha^{(n)}\log_2(P)$ interference bits have to be transmitted. In order to do so, we use the broadcast data symbol s_0. If the quantity of information to be retransmitted exceeds the data rate of s_0, additional broadcast resources will need to be found to enable the successful decoding of the private data symbol. This is the essence of the linear optimization in Theorem 3 and will be discussed further in Section VI. We assume here that all the interference terms can be transmitted using the common data symbol s_0, and we will verify that it is indeed possible for a given RX i to decode its $(k-n)\alpha^{(n)}\log_2(P)$ intended bits.

By using successive decoding, the data symbol s_0 of rate of $(1-\alpha^{(n)})\log_2(P)$ bits can be decoded with a vanishing probability of error as its effective SNR can be seen in (32) to scale as $P^{1-\alpha^{(n)}}$. Upon decoding s_0, we obtain the estimated interferences $(\hat{\boldsymbol{h}}_i^{(1)})^{\mathrm{H}}\mathbf{T}_\ell^{\mathrm{APZF}(1)}\boldsymbol{s}_\ell$, for $\ell \in \mathcal{I}_i^{\mathrm{APZF}}$, up to the quantization noise at the noise floor. It then holds that

$$(\hat{\boldsymbol{h}}_i^{(1)})^{\mathrm{H}}\mathbf{T}_\ell^{\mathrm{APZF}(1)}\boldsymbol{s}_\ell$$
$$= \left(\boldsymbol{h}_i^{\mathrm{H}} + \bar{P}^{-\alpha^{(1)}}(\boldsymbol{\delta}_i^{(1)})^{\mathrm{H}}\right)\mathbf{T}_\ell^{\mathrm{APZF}(1)}\boldsymbol{s}_\ell \quad (34)$$
$$= \boldsymbol{h}_i^{\mathrm{H}}\mathbf{T}_\ell^{\mathrm{APZF}(1)}\boldsymbol{s}_\ell + \bar{P}^{-\alpha^{(1)}}(\boldsymbol{\delta}_i^{(1)})^{\mathrm{H}}\mathbf{T}_\ell^{\mathrm{APZF}(1)}\boldsymbol{s}_\ell \quad (35)$$
$$= \boldsymbol{h}_i^{\mathrm{H}}\mathbf{T}_\ell^{\mathrm{APZF}}\boldsymbol{s}_\ell + \boldsymbol{h}_i^{\mathrm{H}}\left(\mathbf{T}_\ell^{\mathrm{APZF}(1)} - \mathbf{T}_\ell^{\mathrm{APZF}}\right)\boldsymbol{s}_\ell, \quad (36)$$

where (36) is obtained after omitting the second term of (35) because its power scales as $P^{-\alpha^{(1)}}P^{\alpha^{(1)}} = P^0$, i.e., it lies on the noise floor. It holds that, $\forall \ell \in \{1, \ldots, k\}, \forall j \in \{1, \ldots, n\}$, the AP-ZF precoding satisfies the following property (see the proof in Appendix I):

$$\|\mathbf{T}_\ell^{\mathrm{APZF}(j)} - \mathbf{T}_\ell^{\mathrm{APZF}}\|_{\mathrm{F}}^2 \dot{\leq} P^{-\alpha^{(j)}}. \quad (37)$$

It follows from (37) that

$$\boldsymbol{h}_i^{\mathrm{H}}\left(\mathbf{T}_\ell^{\mathrm{APZF}(1)} - \mathbf{T}_\ell^{\mathrm{APZF}}\right)\boldsymbol{s}_\ell \doteq P^0. \quad (38)$$

Once we have subtracted the quantized interference terms, the remaining signal at RX i up to the noise floor is

$$y_i = \boldsymbol{h}_i^{\mathrm{H}}\mathbf{T}_i^{\mathrm{APZF}}\boldsymbol{s}_i. \quad (39)$$

The key point of our approach is that RX i also receives through the broadcast data symbol the interference created by its own intended symbols at the other RXs, i.e., the estimated interference terms $(\hat{\boldsymbol{h}}_\ell^{(1)})^{\mathrm{H}}\mathbf{T}_i^{\mathrm{APZF}(1)}\boldsymbol{s}_i$, for any ℓ such that $i \in \mathcal{I}_\ell^{\mathrm{APZF}}$ (note the swap of indexes i-ℓ with respect to previous expressions). Each of those terms is an independent linear combination of the symbols $\boldsymbol{s}_i$, and thus RX i can form a virtual received vector $\boldsymbol{y}_i^{\mathrm{v}} \in \mathbb{C}^{k-n}$ equal to

$$\boldsymbol{y}_i^{\mathrm{v}} \triangleq \begin{bmatrix} \boldsymbol{h}_i^{\mathrm{H}} \\ (\hat{\boldsymbol{h}}_{i-1}^{(1)})^{\mathrm{H}} \\ \vdots \\ (\hat{\boldsymbol{h}}_{i-(k-n-1)}^{(1)})^{\mathrm{H}} \end{bmatrix}\mathbf{T}_i^{\mathrm{APZF}}\boldsymbol{s}_i. \quad (40)$$
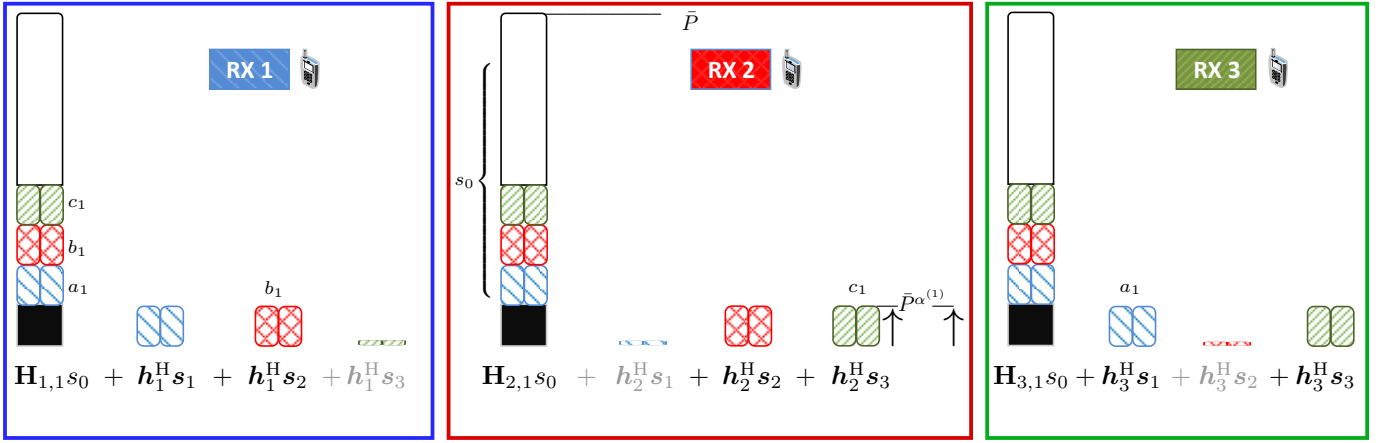
Figure 7: Illustration of the received signals for the Weak-CSIT regime in the case of $k = 3$ Transmitting TXs and $n = 1$ Active TX. Due to the AP-ZF precoding, the interference generated is reduced, and thus extra new information can be sent through s_0 (white).

Each component of $\boldsymbol{y}_y^{\mathrm{v}}$ has a SINR scaling in $P^{\alpha^{(n)}}$ and the AP-ZF precoder is of rank $k - n$ (See Lemma 4 in Appendix I) such that RX i can decode its desired $k - n$ data symbols, each with the rate of $\alpha^{(n)} \log_2(P)$ bits.

Remark 10. *The rank in* (40) *is ensured by the use of the* Passive TXs. *Hence, it is interesting to observe how uninformed TXs prove to be instrumental in the proposed scheme.* □

E. DoF Analysis

The transmission of the data symbols intended to RX i creates $|\mathcal{I}_i^{\mathrm{APZF}}| = k - n - 1$ interference terms, which gives in total $k(k-n-1)\alpha^{(n)} \log_2(P)$ bits that need to be retransmitted. Consequently, we define $\mathrm{DoF}_{n,k}^{\mathrm{Interf}(\text{-})}$ as the DoF *consumed* in order to transmit these interference terms and which is given by

$$\mathrm{DoF}_{n,k}^{\mathrm{Interf}(\text{-})} \triangleq k(k-n-1)\alpha^{(n)}. \tag{41}$$

In contrast, data symbol s_0 carries $(1 - \alpha^{(n)}) \log_2(P)$ bits, i.e., the DoF of the common data symbol $\mathrm{DoF}_{n,k}^{\mathrm{BC}}$ is given by

$$\mathrm{DoF}_{n,k}^{\mathrm{BC}} \triangleq 1 - \alpha^{(n)}. \tag{42}$$

Finally, considering the $(k - n)\alpha^{(n)} \log_2(P)$ private bits for all k users leads to the private DoF denoted by $\mathrm{DoF}_{n,k}^{\mathrm{Priv}}$ and defined as

$$\mathrm{DoF}_{n,k}^{\mathrm{Priv}} \triangleq k(k-n)\alpha^{(n)}, \tag{43}$$

which is the DoF obtained from the private data symbols if all the interference is canceled. Putting (41), (42), and (43) together, the total DoF is

$$\mathrm{DoF}_{n,k} = \mathrm{DoF}_{n,k}^{\mathrm{Priv}} + \mathrm{DoF}_{n,k}^{\mathrm{BC}} - \mathrm{DoF}_{n,k}^{\mathrm{Interf}(\text{-})} \tag{44}$$

at the condition that $\mathrm{DoF}_{n,k}^{\mathrm{BC}} - \mathrm{DoF}_{n,k}^{\mathrm{Interf}(\text{-})} \geq 0$, i.e., that all interference terms could have been retransmitted. If this condition does not hold, the retransmission of the interference is managed through the time-sharing optimization of the different modes as discussed in Section VI. Conversely, the optimal result of Theorem 2 is achieved if the condition $\mathrm{DoF}_{n,k}^{\mathrm{BC}} - \mathrm{DoF}_{n,k}^{\mathrm{Interf}(\text{-})} \geq 0$ is true for the Transmission Mode (m, K), where m is the

number of TXs with $\alpha^{(j)} = \alpha^{(1)}$. By solving the inequality, the maximum value of α_m^{Weak} is obtained as

$$\alpha_m^{\mathrm{Weak}} \triangleq \frac{1}{1 + K(K - m - 1)}. \tag{45}$$

VI. PROOF OF THEOREM 3

For a particular Transmission Mode (n, k), let us start by defining $d_{n,k}$ as the difference between the DoF available in the broadcast symbol s_0 in (42) and the DoF consumed by the interference to be retransmitted in (41), i.e.,

$$d_{n,k} \triangleq \mathrm{DoF}_{n,k}^{\mathrm{BC}} - \mathrm{DoF}_{n,k}^{\mathrm{Interf}(\text{-})} \tag{46}$$

$$= 1 - \alpha^{(n)} - k(k-n-1)\alpha^{(n)}. \tag{47}$$

It is not required that each Transmission Mode leads to the transmission of all interference terms, which would imply that every $d_{n,k}$ satisfies $d_{n,k} > 0$. In fact, it is only necessary that all interference terms were successfully transmitted at the end of the time sharing between all Transmission Modes. Mathematically, this interference retransmission constraint is written as

$$\sum_{k=2}^{K} \sum_{n=1}^{k-1} \gamma_{n,k} d_{n,k} \geq 0, \tag{48}$$

where $\gamma_{n,k}$ is the time-sharing variable, such that $\gamma_{n,k} \geq 0$ and $\sum_{k=2}^{K} \sum_{n=1}^{k-1} \gamma_{n,k} = 1$. By taking into consideration this constraint, the sum DoF can then be rewritten as

$$\sum_{k=2}^{K} \sum_{n=1}^{k-1} \gamma_{n,k} \, \mathrm{DoF}_{n,k}$$

$$= \sum_{k=2}^{K} \sum_{n=1}^{k-1} \mathrm{DoF}_{n,k}^{\mathrm{Priv}} + \mathrm{DoF}_{n,k}^{\mathrm{BC}} - \mathrm{DoF}_{n,k}^{\mathrm{Interf}(\text{-})} \tag{49}$$

$$= \sum_{k=2}^{K} \sum_{n=1}^{k-1} \gamma_{n,k} \left(1 + (k - 1)\alpha^{(n)} \right). \tag{50}$$

The optimal time allocated to each Transmission Mode is obtained by solving the following optimization problem, which

concludes the proof:

$$\underset{\gamma_{n,k}}{\text{maximize}} \quad \sum_{k=2}^{K}\sum_{n=1}^{k-1}\gamma_{n,k}\Big(1+(k-1)\alpha^{(n)}\Big) \quad (51)$$

$$\text{subject to} \quad \sum_{k=2}^{K}\sum_{n=1}^{k-1}\gamma_{n,k}=1, \quad \gamma_{n,k}\geq 0, \quad (52)$$

$$\sum_{k=2}^{K}\sum_{n=1}^{k-1}d_{n,k}\gamma_{n,k}\geq 0. \quad (53)$$

It is important to optimize over both the number of Transmitting TXs and the number of Active TXs. As shown in Fig. 4, using K Transmitting TXs can be detrimental depending on the CSI allocation at each TX. The number of Active TXs (n) controls how many RXs can have its received interference attenuated and up to which level that interference is reduced, while the number of Transmitting TXs (k) controls how many users are served. Furthermore, the difference ($k-n$) impacts on how overloaded the transmission is. For example, setting $k=n+1$ implies a single stream per RX, and thus no overload. In this case, only $n+1$ RXs are served simultaneously, and no interference is generated. Increasing k while fixing n would increase the number of RXs served and the interference generated, which can be beneficial thanks to the fact that the retransmitted interference is useful for two RXs.

VII. CONCLUSION

We have described a novel D-CSIT robust transmission scheme that significantly improves the achieved DoF with respect to state-of-the-art precoding approaches when faced with distributed CSIT. We have first derived an upper bound coined as the *centralized upper bound* and consisting in a genie-aided setting where all the channel estimates are made available at all TXs. We have then shown how this genie-aided upper bound was achieved by the proposed transmission scheme over a range of CSIT configurations, the so-called "Weak-CSIT" regime. Surprisingly, this upper bound can even be achieved with the CSI being handed at a single TX, i.e., with all other TXs not having access to CSIT. The proposed robust precoding scheme relies on new methods, such as the transmission of the estimated interference from a single TX before the interference is generated, as well as a modified ZF precoding allowing for an overloaded transmission. These new methods have a strong potential in other wireless configurations with TXs having access to different qualities of CSI. Converting these new innovative transmission schemes into practical transmission schemes in realistic environments is an interesting and ongoing research direction. Such a robust precoding scheme could yield important gains in practice and make advanced precoding schemes more practical. Deriving tighter distributed upper bounds is also an interesting and challenging research problem.

APPENDIX I
AP-ZF PROPERTIES

We start by showing some simple but important properties of the AP-ZF precoder. We consider in the following the precoder for a specific RX i's data symbols. From symmetry,

the precoder satisfies the same properties for any RX, such that we omit hereinafter the RX's sub-index i for clarity. Let us recall that the AP-ZF precoder aims to cancel the interference out at only a subset of n RXs.

Lemma 3. *Let $\mathbf{H}\in\mathbb{C}^{n\times K}$ denote the channel matrix towards the n RXs whose received interference is reduced. With perfect channel knowledge at all Active TXs, the AP-ZF precoder with n Active TXs and $K-n$ Passive TXs satisfies*

$$\mathbf{H}\mathbf{T}^{\text{APZF}\star} \xrightarrow{P\rightarrow\infty} \mathbf{0}_{n\times(K-n)}, \quad (54)$$

where $\mathbf{T}^{\text{APZF}\star}$ denotes the AP-ZF precoder according to the description in Section V-B but based on perfect CSIT, and it is given as

$$\mathbf{T}^{\text{APZF}\star} \triangleq \lambda^{\text{APZF}}\begin{bmatrix}\mathbf{T}^{\text{A}\star}\\\mathbf{T}^{\text{P}}\end{bmatrix}. \quad (55)$$

Proof: Using the well known Resolvent identity [51, Lemma 6.1], we can write that

$$\left(\mathbf{H}_{\text{A}}^{\text{H}}\mathbf{H}_{\text{A}}+\tfrac{1}{P}\mathbf{I}_n\right)^{-1}-\left(\mathbf{H}_{\text{A}}^{\text{H}}\mathbf{H}_{\text{A}}\right)^{-1}$$
$$=-\left(\mathbf{H}_{\text{A}}^{\text{H}}\mathbf{H}_{\text{A}}\right)^{-1}\tfrac{1}{P}\mathbf{I}_n\left(\mathbf{H}_{\text{A}}^{\text{H}}\mathbf{H}_{\text{A}}+\tfrac{1}{P}\mathbf{I}_n\right)^{-1}. \quad (56)$$

We can then compute the leaked interference as

$$\mathbf{H}\mathbf{T}^{\text{APZF}\star}=\lambda^{\text{APZF}}\mathbf{H}_{\text{A}}\mathbf{T}^{\text{A}\star}+\lambda^{\text{APZF}}\mathbf{H}_{\text{P}}\mathbf{T}^{\text{P}} \quad (57)$$

$$\overset{(a)}{=}\lambda^{\text{APZF}}\mathbf{H}_{\text{A}}\left(\mathbf{H}_{\text{A}}^{\text{H}}\mathbf{H}_{\text{A}}\right)^{-1}\frac{1}{P}\mathbf{I}_n$$
$$\times\left(\mathbf{H}_{\text{A}}^{\text{H}}\mathbf{H}_{\text{A}}+\frac{1}{P}\mathbf{I}_n\right)^{-1}\mathbf{H}_{\text{A}}^{\text{H}}\mathbf{H}_{\text{P}}\mathbf{T}^{\text{P}}, \quad (58)$$

where equality (a) follows from inserting (56) inside the AP-ZF precoder and simplifying. By letting the available power P tend to infinity, the leaked interference tends to zero. ∎

Lemma 4. *The AP-ZF precoder matrix with n Active TXs and $K-n$ Passive TXs is of rank $K-n$.*

Proof: The passive precoder was chosen such that $\mathbf{T}^{\text{P}}$ is full rank, i.e., of rank $K-n$. For any Active TX j, the precoder $\mathbf{T}^{\text{A}(j)}$ is a linear function of $\mathbf{T}^{\text{P}}$, such that the effective AP-ZF precoder $\mathbf{T}^{\text{APZF}}$ resulting from distributed precoding is exactly of rank $K-n$. ∎

A. Proof of Lemma 2

We follow a similar approach as in [19]. We can use once more the resolvent identity [51, Lemma 6.1] to approximate the matrix inverse and show that, for any $j\leq n$,

$$\left\|\mathbf{T}^{\text{APZF}(j)}-\mathbf{T}^{\text{APZF}\star}\right\|_{\text{F}}^2\overset{.}{\leq}P^{-\alpha^{(j)}}. \quad (59)$$

It then follows that

$$\left\|\mathbf{H}\mathbf{T}^{\text{APZF}}\right\|_{\text{F}}^2\overset{.}{\leq}\left\|\mathbf{H}\left(\mathbf{T}^{\text{APZF}}-\mathbf{T}^{\text{APZF}\star}\right)\right\|_{\text{F}}^2 \quad (60)$$

$$\overset{.}{\leq}\left\|\mathbf{H}\right\|_{\text{F}}^2\left\|\mathbf{T}^{\text{APZF}}-\mathbf{T}^{\text{APZF}\star}\right\|_{\text{F}}^2 \quad (61)$$

$$\overset{.}{\leq}\left\|\mathbf{H}\right\|_{\text{F}}^2\sum_{j=1}^{n}\left\|\mathbf{T}^{\text{APZF}(j)}-\mathbf{T}^{\text{APZF}\star}\right\|_{\text{F}}^2 \quad (62)$$

$$\overset{.}{\leq}P^{-\min_{j\in\{1,\dots,n\}}\alpha^{(j)}}, \quad (63)$$

where (60) comes from Lemma 3 and (63) follows from (59).

Remark 11. *The interference attenuation of AP-ZF is only limited by the worst CSI accuracy among the Active TXs, and it does not depend on the CSI accuracy at the Passive TXs.* □

APPENDIX II
PROOF OF LEMMA 1

We prove Lemma 1 (and so Theorem 1) for the broad general case where the estimation noise random variables are mutually independent and they are drawn from continuous distributions with density. We first enunciate some definitions and assumptions that are taken on the random variables and their probability density functions (pdf). Later, we prove the lemma for the $K = 2$ users case and, to conclude, we prove the general $K > 2$ users case by induction. From the independence between different channel coefficients, we can restrict ourselves to an arbitrary link such that we omit the previously used sub-indexes i, k.

A. Proof of Lemma 1

1) Preliminaries:

We recall that the probability density function of a random variable $\mathcal{X}$ is denoted as $f_{\mathcal{X}}(x)$. Next, we introduce three definitions that are necessary for the proof.

D1) *ε-Support:* For any $\varepsilon > 0$, the ε-support of a random variable $\mathcal{X}$ is defined as

$$\mathcal{S}_{\mathcal{X}}^{\varepsilon} \triangleq \{x \mid f_{\mathcal{X}}(x) > \varepsilon\}. \tag{64}$$

D2) *Bounded Support*: A random variable $\mathcal{X}$ is said to have bounded support if $\exists M_{\mathcal{X}} < \infty$ such that $x \leq |M_{\mathcal{X}}|$ for any $x \in \mathcal{S}_{\mathcal{X}}^{\varepsilon}$ and for any $\varepsilon > 0$.

D3) *Bounded Probability Density Function*: A random variable $\mathcal{X}$ is said to have bounded probability density function if there exists a constant $f_{\mathcal{X}}^{\max} < \infty$ such that $f_{\mathcal{X}}(x) \leq f_{\mathcal{X}}^{\max}$ for any x.

Furthermore, the following assumptions are adopted:

H1) $\mathcal{H}$, $\Delta^{(j)}$, $\forall j \in \mathcal{K}$, are continuous random variables with density that satisfy D2) and D3).

H2) $P > 1$ and $1 \geq \alpha^{(1)} \geq \cdots \geq \alpha^{(K)} \geq 0$.

H3) $\mathcal{H}$, $\Delta^{(j)}$, are independent of P, $\alpha^{(j)}$.

H4) $\mathcal{H}_{i,k}$ is independent of $\mathcal{H}_{i',k'}$, $\quad \forall (i, k) \neq (i', k')$.

H5) $\Delta_{i,k}^{(\ell)}$ is independent of $\Delta_{i',k'}^{(j)}$, $\quad \forall (i, k, j) \neq (i', k', \ell)$.

From the above, $\mathcal{H}$ and $\Delta^{(j)}$ satisfy Definition 1, $\forall j \in \mathcal{K}$. The K different estimates of $\mathcal{H}$ are K random variables defined as $\hat{\mathcal{H}}^{(j)} \triangleq \mathcal{H} + \bar{P}^{-\alpha^{(j)}} \Delta^{(j)}$. Further, we denote the samples drawn from the aforementioned variables as $h \sim \mathcal{H}$, $\delta^{(j)} \sim \Delta^{(j)}$, $\hat{h}^{(j)} \sim \hat{\mathcal{H}}^{(j)}$, such that it follows that $\hat{h}^{(j)} \triangleq h + \bar{P}^{-\alpha^{(j)}} \delta^{(j)}$. As a refresher, and because we will make extensive use of it, we recall the well-known formula for the pdf of a random variable multiplied by a positive constant.

Proposition 1. *If $\mathcal{X}$ is a continuous random variable with probability density function $f_{\mathcal{X}}(x)$, then, for $c > 0$, so is $c \cdot \mathcal{X}$ a continuous random variable with probability density function*

$$f_{c\mathcal{X}}(x) = \frac{1}{c} f_{\mathcal{X}} \left(\frac{x}{c} \right). \tag{65}$$

Furthermore, we present a useful lemma on the convergence of the estimate variables $\hat{\mathcal{H}}^{(j)}$ that we apply during the proof.

Lemma 5. *Let $\hat{\mathcal{H}}^{(j)}$, $\forall j \in \mathcal{K}$, be defined from assumptions H1)-H5), and consider that $\alpha^{(j)} > 0$. Then, $f_{\hat{\mathcal{H}}^{(j)}}$ converges almost surely to $f_{\mathcal{H}}$, i.e.,*

$$\lim_{P \to \infty} f_{\hat{\mathcal{H}}^{(j)}}(x) = f_{\mathcal{H}}(x). \tag{66}$$

Corollary 2. *Let $\alpha^{(1)} > 0$. Then,*

$$\lim_{P \to \infty} f_{\Delta^{(1)} \mid \hat{\mathcal{H}}^{(1)}, \dots, \hat{\mathcal{H}}^{(K)}}(y \mid \hat{h}^{(1)}, \dots, \hat{h}^{(K)}) = f_{\Delta^{(1)}}(y). \tag{67}$$

Proof: The proof of both Lemma 5 and Corollary 2 is relegated to Appendix II-B. ∎

Finally, we recall here the Lebesgue's Dominated Convergence Theorem [52].

Theorem 4 ([52, Theorem 16.4]). *Let $\{f_n\}$ be a sequence of functions on the measure space (Ω, Σ, μ), where Ω is a non-empty sample space, Σ is a σ-algebra on the space Ω, and μ a measure on (Ω, Σ). Suppose that*

$$\lim_{n \to \infty} f_n(x) = f(x) \tag{68}$$

almost surely. Further suppose that exists an integrable non-negative function G such that $|f_n(x)| \leq G(x)$, $\forall n$, almost surely. Then $\{f_n\}$ and f are integrable and

$$\lim_{n \to \infty} \int_{\Omega} f_n(x) d\mu(x) = \int_{\Omega} f(x) d\mu(x). \tag{69}$$

2) Proof for the K=2 estimates Case:

We can write

$$f_{\mathcal{H} \mid \hat{\mathcal{H}}^{(1)}, \hat{\mathcal{H}}^{(2)}}(h \mid \hat{h}^{(1)}, \hat{h}^{(2)})$$

$$= \frac{f_{\mathcal{H}, \hat{\mathcal{H}}^{(1)}, \hat{\mathcal{H}}^{(2)}}(h, \hat{h}^{(1)}, \hat{h}^{(2)})}{f_{\hat{\mathcal{H}}^{(1)}, \hat{\mathcal{H}}^{(2)}}(\hat{h}^{(1)}, \hat{h}^{(2)})} \tag{70}$$

$$\stackrel{(a)}{=} \frac{f_{\mathcal{H}, \hat{\mathcal{H}}^{(1)}}(h, \hat{h}^{(1)}) f_{\bar{P}^{-\alpha^{(2)}} \Delta^{(2)}}(\hat{h}^{(2)} - h)}{f_{\hat{\mathcal{H}}^{(1)}}(\hat{h}^{(1)}) f_{\hat{\mathcal{H}}^{(2)} \mid \hat{\mathcal{H}}^{(1)}}(\hat{h}^{(2)} \mid \hat{h}^{(1)})} \tag{71}$$

$$= f_{\mathcal{H} \mid \hat{\mathcal{H}}^{(1)}}(h \mid \hat{h}^{(1)}) \frac{f_{\bar{P}^{-\alpha^{(2)}} \Delta^{(2)}}(\hat{h}^{(2)} - h)}{f_{\hat{\mathcal{H}}^{(2)} \mid \hat{\mathcal{H}}^{(1)}}(\hat{h}^{(2)} \mid \hat{h}^{(1)})}, \tag{72}$$

where (a) comes from the independence between $\mathcal{H}$, $\Delta^{(1)}$, $\Delta^{(2)}$. Furthermore, by applying Bayes' formula we obtain that

$$f_{\mathcal{H} \mid \hat{\mathcal{H}}^{(1)}}(h \mid \hat{h}^{(1)}) = \frac{f_{\bar{P}^{-\alpha^{(1)}} \Delta^{(1)}}(\bar{P}^{-\alpha^{(1)}} \delta^{(1)}) f_{\mathcal{H}}(h)}{f_{\hat{\mathcal{H}}^{(1)}}(\hat{h}^{(1)})} \tag{73}$$

$$= \bar{P}^{\alpha^{(1)}} \frac{f_{\Delta^{(1)}}(\delta^{(1)}) f_{\mathcal{H}}(h)}{f_{\hat{\mathcal{H}}^{(1)}}(\hat{h}^{(1)})}, \tag{74}$$

where (74) comes from Proposition 1. Let us consider separately the cases where $\alpha^{(1)} = 0$ and where $\alpha^{(1)} > 0$.

a) $\alpha^{(1)} = 0$: In this case, (74) does not depend on P, since $P^0 = 1$, $\forall P > 0$. From H1), $f_{\mathcal{H}}$ and $f_{\Delta^{(1)}}$ are bounded away from ∞. Moreover, if $\hat{h}^{(1)} \in \mathcal{S}_{\hat{\mathcal{H}}^{(1)}}^{\varepsilon}$, then $f_{\hat{\mathcal{H}}^{(1)}}$ is also lower-bounded by ε. Thus,

$$\max f_{\mathcal{H} \mid \hat{\mathcal{H}}^{(1)}}(h \mid \hat{h}^{(1)}) = \mathcal{O} \left(\bar{P}^0 \right). \tag{75}$$

b) $\alpha^{(1)} > 0$*:* From Lemma 5, we have that $f_{\hat{\mathcal{H}}^{(1)}}$ converges almost surely (a.s.) to $f_{\mathcal{H}}$, and from H1) that $\max f_{\Delta^{(1)}} < \infty$. Thus, from (74) it holds that

$$\max f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)}}(h \mid \hat{h}^{(1)}) = \mathcal{O}(\bar{P}^{\alpha^{(1)}}). \tag{76}$$

Hence, it follows from (75)-(76) that, in order to prove Lemma 1, i.e., that

$$\max f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)},\hat{\mathcal{H}}^{(2)}} = \mathcal{O}(\bar{P}^{\alpha^{(1)}}), \tag{77}$$

we need to demonstrate that the limit

$$\lim_{P\to\infty} \frac{f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)},\hat{\mathcal{H}}^{(2)}}(h|\hat{h}^{(1)},\hat{h}^{(2)})}{f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)}}(h|\hat{h}^{(1)})} \tag{78}$$

exists and is bounded away from 0 and ∞. First, note that from (72) we have that

$$\frac{f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)},\hat{\mathcal{H}}^{(2)}}(h|\hat{h}^{(1)},\hat{h}^{(2)})}{f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)}}(h|\hat{h}^{(1)})} = \frac{f_{\bar{P}^{-\alpha^{(2)}}\Delta^{(2)}}(\hat{h}^{(2)}-h)}{f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{h}^{(2)} \mid \hat{h}^{(1)})}. \tag{79}$$

Let us focus first on the denominator of the right-hand side of (79). By taking into account again that $\Delta^{(j)}$ is independent of $\mathcal{H}$, and that $\hat{\mathcal{H}}^{(2)} = \hat{\mathcal{H}}^{(1)} - \bar{P}^{-\alpha^{(1)}}\Delta^{(1)} + \bar{P}^{-\alpha^{(2)}}\Delta^{(2)}$, we obtain that

$$f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{h}^{(2)} \mid \hat{h}^{(1)})$$
$$= f_{\bar{P}^{-\alpha^{(2)}}\Delta^{(2)}-\bar{P}^{-\alpha^{(1)}}\Delta^{(1)}|\hat{\mathcal{H}}^{(1)}}(\hat{h}^{(2)} - \hat{h}^{(1)} \mid \hat{h}^{(1)}). \tag{80}$$

Note that $\hat{h}^{(2)} - \hat{h}^{(1)} = \bar{P}^{-\alpha^{(2)}}\delta^{(2)} - \bar{P}^{-\alpha^{(1)}}\delta^{(1)}$. From the independence of $\Delta^{(1)}$ and $\Delta^{(2)}$, (80) can be expressed as a convolution such that

$$f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{h}^{(2)} \mid \hat{h}^{(1)})$$
$$= f_{-\bar{P}^{-\alpha^{(1)}}\Delta^{(1)}|\hat{\mathcal{H}}^{(1)}} * f_{\bar{P}^{-\alpha^{(2)}}\Delta^{(2)}}(\hat{h}^{(2)}-\hat{h}^{(1)}|\hat{h}^{(1)}) \tag{81}$$
$$= \int_{-\infty}^{\infty} \Big(f_{\bar{P}^{-\alpha^{(2)}}\Delta^{(2)}}(\hat{h}^{(2)}-\hat{h}^{(1)}-x)$$
$$\times f_{-\bar{P}^{-\alpha^{(1)}}\Delta^{(1)}|\hat{\mathcal{H}}^{(1)}}(x|\hat{h}^{(1)}) \Big) \mathrm{d}x \tag{82}$$
$$= \int_{-\infty}^{\infty} \Big(\bar{P}^{\alpha^{(2)}} f_{\Delta^{(2)}}(\bar{P}^{\alpha^{(2)}}(\hat{h}^{(2)} - \hat{h}^{(1)} - x))$$
$$\times \bar{P}^{\alpha^{(1)}} f_{-\Delta^{(1)}|\hat{\mathcal{H}}^{(1)}}(\bar{P}^{\alpha^{(1)}} x \mid \hat{h}^{(1)}) \Big) \mathrm{d}x \tag{83}$$
$$= \bar{P}^{\alpha^{(2)}} \int_{-\infty}^{\infty} \Big(f_{\Delta^{(2)}}(\bar{P}^{\alpha^{(2)}}(\hat{h}^{(2)} - \hat{h}^{(1)} - \bar{P}^{-\alpha^{(1)}}y))$$
$$\times f_{-\Delta^{(1)}|\hat{\mathcal{H}}^{(1)}}(y|\hat{h}^{(1)}) \Big) \mathrm{d}y, \tag{84}$$

where (83) follows from applying the change of pdf of Proposition 1 from $f_{-\bar{P}^{-\alpha^{(1)}}\Delta^{(1)}}$ to $f_{-\Delta^{(1)}}$, and (84) comes from changing the integration variable to $y = \bar{P}^{\alpha^{(1)}}x$ (and thus $\mathrm{d}x = \bar{P}^{-\alpha^{(1)}}\mathrm{d}y$).

Note that, by applying $\hat{h}^{(i)} = h + \bar{P}^{-\alpha^{(i)}}\delta^{(i)}$, it follows that $\bar{P}^{\alpha^{(2)}}(\hat{h}^{(2)} - \hat{h}^{(1)} - \bar{P}^{-\alpha^{(1)}}y) = \delta^{(2)} - \bar{P}^{\alpha^{(2)}-\alpha^{(1)}}(\delta^{(1)} + y)$ for any y. Therefore, let us introduce the notation

$$v_P(y) \triangleq f_{\Delta^{(2)}}(\delta^{(2)} - \bar{P}^{\alpha^{(2)}-\alpha^{(1)}}(\delta^{(1)} + y))$$
$$\times f_{-\Delta^{(1)}|\hat{\mathcal{H}}^{(1)}}(y|\hat{h}^{(1)}). \tag{85}$$

We can thus re-write (84) by means of (85) to obtain

$$f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{h}^{(2)} \mid \hat{h}^{(1)}) = \bar{P}^{\alpha^{(2)}} \int_{-\infty}^{\infty} v_P(y)\,\mathrm{d}y. \tag{86}$$

From (86) and by applying again the change of pdf of Proposition 1 to the numerator, the term $\frac{f_{\bar{P}^{-\alpha^{(2)}}\Delta^{(2)}}(\hat{h}^{(2)}-h)}{f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{h}^{(2)}|\hat{h}^{(1)})}$ can be expressed as

$$\frac{f_{\bar{P}^{-\alpha^{(2)}}\Delta^{(2)}}(\hat{h}^{(2)} - h)}{f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{h}^{(2)} \mid \hat{h}^{(1)})} = \frac{f_{\Delta^{(2)}}(\delta^{(2)})}{\int_{-\infty}^{\infty} v_P(y)\,\mathrm{d}y}. \tag{87}$$

From continuity of $f_{\Delta^{(1)}}$ and $f_{\Delta^{(2)}}$, and Corollary 2, we obtain the limit of $v_P(y)$ in (85) when $P \to \infty$. This limit has two possible expressions depending on the relation between $\alpha^{(1)}$ and $\alpha^{(2)}$. Specifically, it holds that

$$\lim_{P\to\infty} v_P(y) = f_{\Delta^{(2)}}(\delta^{(2)}-\delta^{(1)} - y)f_{-\Delta^{(1)}}(y) \tag{88}$$

if $\alpha^{(1)} = \alpha^{(2)}$, and that

$$\lim_{P\to\infty} v_P(y) = f_{\Delta^{(2)}}(\delta^{(2)})f_{-\Delta^{(1)}}(y) \tag{89}$$

if $\alpha^{(1)} > \alpha^{(2)}$. Now we prove separately each of the two possible cases.

a) $\alpha^{(1)} = \alpha^{(2)}$*:* From the Lebesgue's Dominated Convergence Theorem (Theorem 4), D3), and (88), the limit exists and it holds that

$$\lim_{P\to\infty} \int_{-\infty}^{\infty} v_P(y)\,\mathrm{d}y = f_{-\Delta^{(1)}} * f_{\Delta^{(2)}}(\delta^{(2)} - \delta^{(1)}). \tag{90}$$

From (87) and (90), it holds that

$$\lim_{P\to\infty} \frac{f_{\bar{P}^{-\alpha^{(2)}}\Delta^{(2)}}(\hat{h}^{(2)}-h)}{f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{h}^{(2)}|\hat{h}^{(1)})} = \frac{f_{\Delta^{(2)}}(\delta^{(2)})}{f_{-\Delta^{(1)}} * f_{\Delta^{(2)}}(\delta^{(2)}-\delta^{(1)})}. \tag{91}$$

From D3), we have that, for any $i \in \{1,2\}$, $\exists f_{\Delta^{(i)}}^{\max} < \infty$ such that $f_{\Delta^{(i)}}(x) \le f_{\Delta^{(i)}}^{\max} \;\forall x$. Then, it holds that

$$f_{-\Delta^{(1)}} * f_{\Delta^{(2)}}(x) \le \max(f_{\Delta^{(1)}}^{\max}, f_{\Delta^{(2)}}^{\max}). \tag{92}$$

Conversely, let $\mathbb{1}$ be the indicator function and let then τ be

$$\tau \triangleq \int_{-\infty}^{\infty} \mathbb{1}_{x\in\mathcal{S}_{-\Delta^{(1)}}^{\varepsilon}} \times \mathbb{1}_{(\delta^{(2)}-\delta^{(1)}-x)\in\mathcal{S}_{\Delta^{(2)}}^{\varepsilon}}\,\mathrm{d}x. \tag{93}$$

Then,

$$f_{-\Delta^{(1)}} * f_{\Delta^{(2)}}(\delta^{(2)} - \delta^{(1)}) > \epsilon^2 \tau \tag{94}$$

and $\tau > 0$ if $\delta^{(1)} \in \mathcal{S}_{\Delta^{(1)}}^{\varepsilon}$ and $\delta^{(2)} \in \mathcal{S}_{\Delta^{(2)}}^{\varepsilon}$. From (92) and (94), (91) satisfies

$$\frac{\varepsilon}{\max(f_{\Delta^{(1)}}^{\max}, f_{\Delta^{(2)}}^{\max})} < \frac{f_{\Delta^{(2)}}(\delta^{(2)})}{f_{-\Delta^{(1)}} * f_{\Delta^{(2)}}(\delta^{(2)} - \delta^{(1)})} < \frac{f_{\Delta^{(2)}}^{\max}}{\varepsilon^2\tau}.$$

This implies that $\max f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)},\hat{\mathcal{H}}^{(2)}} = \mathcal{O}(\bar{P}^{\alpha^{(1)}})$ (cf. (77)) and thus the proof is concluded for the $\alpha^{(2)} = \alpha^{(1)}$ case.

b) $\alpha^{(1)} > \alpha^{(2)}$: From the Lebesgue's Dominated Convergence Theorem and (89), the limit exists and it holds that

$$\lim_{P\to\infty} v_P(y) = \int_{-\infty}^{\infty} f_{\Delta^{(2)}}(\delta^{(2)}) f_{-\Delta^{(1)}}(y)\,\mathrm{d}y \qquad (95)$$

$$= f_{\Delta^{(2)}}(\delta^{(2)}). \qquad (96)$$

Applying (96) in (87) yields

$$\lim_{P\to\infty} \frac{f_{\bar{P}^{-\alpha^{(2)}}\Delta^{(2)}}(\hat{h}^{(2)} - h)}{f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{h}^{(2)}\mid\hat{h}^{(1)})} = 1, \qquad (97)$$

which concludes the proof of Lemma 1 for the 2-estimate case.

3) Proof for $K > 2$ estimates:

We prove by induction that Lemma 1 also holds for any number K of estimates. We have proved that it is true for the base cases $K = 1$ (trivial) and $K = 2$. In the following, we prove the induction step. We denote the set of estimates as $\bar{\mathcal{G}}_{\mathcal{K}} \triangleq \{\hat{\mathcal{H}}^{(1)}, \dots, \hat{\mathcal{H}}^{(K)}\}$ and, consistently, the set of given values as $\bar{g}_{\mathcal{K}} \triangleq \{\hat{h}^{(1)}, \dots, \hat{h}^{(K)}\}$. Let us assume that Lemma 1 is verified for a given K. We consider $K+1$ estimates. Then, from the mutual independence of the estimation noise variables $\Delta^{(j)}$ and Bayes' formula, we obtain that

$$f_{\mathcal{H}|\bar{\mathcal{G}}_{\mathcal{K}},\hat{\mathcal{H}}^{(K+1)}}\big(h\mid\bar{g}_{\mathcal{K}},\hat{h}^{(K+1)}\big)$$
$$= \underbrace{\frac{f_{\mathcal{H},\bar{\mathcal{G}}_{\mathcal{K}}}(h,\bar{g}_{\mathcal{K}})}{f_{\bar{\mathcal{G}}_{\mathcal{K}}}(\bar{g}_{\mathcal{K}})}}_{f_{\mathcal{H}|\bar{\mathcal{G}}_{\mathcal{K}}}(\hat{h}|\bar{g}_{\mathcal{K}})} \frac{f_{\bar{P}^{-\alpha^{(K+1)}}\Delta^{(K+1)}}\big(\hat{h}^{(K+1)} - h\big)}{f_{\hat{\mathcal{H}}^{(K+1)}|\bar{\mathcal{G}}_{\mathcal{K}}}\big(\hat{h}^{(K+1)}|\bar{g}_{\mathcal{K}}\big)}. \qquad (98)$$

From the induction hypothesis, it holds that

$$\max f_{\mathcal{H}|\bar{\mathcal{G}}_{\mathcal{K}}}(h\mid\bar{g}_{\mathcal{K}}) = \mathcal{O}\big(\bar{P}^{\alpha^{(1)}}\big). \qquad (99)$$

Thus, we need to prove that

$$0 < \lim_{P\to\infty} \frac{f_{\bar{P}^{-\alpha^{(K+1)}}\Delta^{(K+1)}}\big(\hat{h}^{(K+1)} - h\big)}{f_{\hat{\mathcal{H}}^{(K+1)}|\bar{\mathcal{G}}_{\mathcal{K}}}\big(\hat{h}^{(K+1)}|\bar{g}_{\mathcal{K}}\big)} < \infty. \qquad (100)$$

Let us introduce the notation

$$\Delta' \triangleq \bar{P}^{-\alpha^{(K+1)}}\Delta^{(K+1)} - \bar{P}^{-\alpha^{(1)}}\Delta^{(1)}. \qquad (101)$$

By taking into consideration that

$$\hat{h}^{(K+1)} - \hat{h}^{(1)} = \bar{P}^{-\alpha^{(K+1)}}\delta^{(K+1)} - \bar{P}^{-\alpha^{(1)}}\delta^{(1)}, \qquad (102)$$

the denominator of the expression in (100) can be rewritten as

$$f_{\hat{\mathcal{H}}^{(K+1)}|\bar{\mathcal{G}}_{\mathcal{K}}}\big(\hat{h}^{(K+1)}|\bar{g}_{\mathcal{K}}\big)$$
$$= f_{\hat{\mathcal{H}}^{(1)}+\Delta'|\bar{\mathcal{G}}_{\mathcal{K}}}\big(\hat{h}^{(K+1)} - \hat{h}^{(1)} + \hat{h}^{(1)}\mid\bar{g}_{\mathcal{K}}\big) \qquad (103)$$
$$= f_{\Delta'|\bar{\mathcal{G}}_{\mathcal{K}}}\big(\bar{P}^{-\alpha^{(K+1)}}\delta^{(K+1)} - \bar{P}^{-\alpha^{(1)}}\delta^{(1)}\mid\bar{g}_{\mathcal{K}}\big) \qquad (104)$$
$$= f_{\bar{P}^{-\alpha^{(K+1)}}\Delta^{(K+1)}} * f_{-\bar{P}^{-\alpha^{(1)}}\Delta^{(1)}|\bar{g}_{\mathcal{K}}}\big(\hat{h}^{(K+1)}-\hat{h}^{(1)}|\bar{g}_{\mathcal{K}}\big) \qquad (105)$$

where in (105) we have expressed $f_{\Delta'|\bar{\mathcal{G}}_{\mathcal{K}}}$ as a convolution of pdfs. We follow the same steps as in (82)-(87) to obtain

$$\frac{f_{\bar{P}^{-\alpha^{(K+1)}}\Delta^{(K+1)}}\big(\hat{h}^{(K+1)} - h\big)}{f_{\hat{\mathcal{H}}^{(K+1)}|\bar{\mathcal{G}}_{\mathcal{K}}}\big(\hat{h}^{(K+1)}|\bar{g}_{\mathcal{K}}\big)}$$
$$= \frac{f_{\Delta^{(K+1)}}(\delta^{(K+1)})}{\int_{-\infty}^{\infty} f_{\Delta^{(K+1)}}(\delta'_y) f_{-\Delta^{(1)}|\bar{\mathcal{G}}_{\mathcal{K}}}(y|\bar{g}_{\mathcal{K}})\,\mathrm{d}y}, \qquad (106)$$

where we have introduced the notation

$$\delta'_y \triangleq \delta^{(K+1)} - \bar{P}^{\alpha^{(K+1)}-\alpha^{(1)}}(\delta^{(1)} + y) \qquad (107)$$

for ease of reading. We can see that (106) is equivalent to (87) with $\Delta^{(K+1)}$ in place of $\Delta^{(2)}$. Then, by following the same derivation as in the 2-estimate case, i.e., using Corollary 2 and Lebesgue's Dominated Convergence Theorem, we conclude the induction step. From the base case and the induction step, Lemma 1 is proven.

B. Proof of Lemma 5

Assuming $\alpha^{(i)} > 0$, we have that

$$\lim_{P\to\infty} f_{\hat{\mathcal{H}}^{(i)}}(\hat{h}^{(i)})$$
$$= \lim_{P\to\infty} f_{\mathcal{H}+\bar{P}^{-\alpha^{(i)}}\Delta^{(i)}}(h + \bar{P}^{-\alpha^{(i)}}\delta^{(i)}) \qquad (108)$$
$$= \lim_{P\to\infty} \int_{-\infty}^{\infty} f_{\mathcal{H}}(h + \bar{P}^{-\alpha^{(i)}}\delta^{(i)} - x) f_{\bar{P}^{-\alpha^{(i)}}\Delta^{(i)}}(x)\,\mathrm{d}x \qquad (109)$$
$$= \lim_{P\to\infty} \int_{-\infty}^{\infty} f_{\mathcal{H}}(h + \bar{P}^{-\alpha^{(i)}}\delta^{(i)} - \bar{P}^{-\alpha^{(i)}}y) f_{\Delta^{(i)}}(y)\,\mathrm{d}y \qquad (110)$$
$$= \int_{-\infty}^{\infty} f_{\mathcal{H}}(h) f_{\Delta^{(i)}}(y)\,\mathrm{d}y \qquad (111)$$
$$= f_{\mathcal{H}}(h), \qquad (112)$$

where (110) comes from applying the relation between $f_{\bar{P}^{-\alpha_i}\Delta_i}$ and f_{Δ_i} (see Proposition 1), and from the change of integration variable $y = \bar{P}^{\alpha_i}x$. Finally, (111) follows from applying Lebesgue's Dominated Convergence Theorem. Therefore, $f_{\hat{\mathcal{H}}^{(i)}}$ converges almost surely to $f_{\mathcal{H}}$ and hence Lemma 5 is proven. Then, in order to prove Corollary 2, i.e., that

$$\lim_{P\to\infty} f_{\Delta^{(1)}|\hat{\mathcal{H}}^{(1)},\dots,\hat{\mathcal{H}}^{(K)}}(y|\hat{h}^{(1)},\dots,\hat{h}^{(K)}) = f_{\Delta^{(1)}}(y), \qquad (113)$$

we apply Bayes' formula to obtain

$$f_{\Delta^{(1)}|\hat{\mathcal{H}}^{(1)},\dots,\hat{\mathcal{H}}^{(K)}}(y|\hat{h}^{(1)},\dots,\hat{h}^{(K)})$$
$$= \frac{f_{\hat{\mathcal{H}}^{(1)},\dots,\hat{\mathcal{H}}^{(K)}|\Delta^{(1)}}(\hat{h}^{(1)},\dots,\hat{h}^{(K)}|y) f_{\Delta^{(1)}}(y)}{f_{\hat{\mathcal{H}}^{(1)},\dots,\hat{\mathcal{H}}^{(K)}}(\hat{h}^{(1)},\dots,\hat{h}^{(K)})} \qquad (114)$$
$$= \frac{f_{\mathcal{H},\hat{\mathcal{H}}^{(2)},\dots,\hat{\mathcal{H}}^{(K)}}(h, h^{(2)},\dots,h^{(K)})}{f_{\hat{\mathcal{H}}^{(1)},\dots,\hat{\mathcal{H}}^{(K)}}(h^{(1)},\dots,h^{(K)})} f_{\Delta^{(1)}}(y). \qquad (115)$$

The fact that $\alpha^{(1)} > 0$ and (112) yield

$$\lim_{P\to\infty} \frac{f_{\mathcal{H},\hat{\mathcal{H}}^{(2)},\dots,\hat{\mathcal{H}}^{(K)}}(h, h^{(2)},\dots,h^{(K)})}{f_{\hat{\mathcal{H}}^{(1)},\dots,\hat{\mathcal{H}}^{(K)}}(h^{(1)},\dots,h^{(K)})} = 1. \qquad (116)$$

By taking the limit on (115) and applying (116), it holds that

$$\lim_{P\to\infty} f_{\Delta^{(1)}|\hat{\mathcal{H}}^{(1)},\dots,\hat{\mathcal{H}}^{(K)}}(y|h_1,\dots,h_K) = f_{\Delta^{(1)}}(y), \qquad (117)$$

which concludes the proof. $\qquad\square$

APPENDIX III
PROOF OF COROLLARY 1

In this appendix we prove Corollary 1, i.e., that the solution of Theorem 3 is composed of at most two phases

(Transmission Modes). This is equivalent to prove that there exist $k_1, k_2 \in \mathcal{K}$, $n_1 < k_1$, $n_2 < k_2$ such that

$$
\begin{aligned}
\gamma_{n_1,k_1} &> 0, \\
\gamma_{n_2,k_2} &\geq 0, \\
\gamma_{n,k} &= 0, \qquad \forall (n,k) \neq (n_1,k_1),(n_2,k_2),
\end{aligned}
\tag{118}
$$

is always an optimal solution of the maximization problem stated in Theorem 3.

Let us denote the number of variables $\gamma_{n,k}$ in the optimization problem as D. From the problem definition, we know that $D = \sum_{i=2}^{K}(i-1) = \frac{K(K-1)}{2}$. In order to simplify the notation, we denote the aforementioned variables with a unique sub-index. Therefore, the variables $\{\gamma_{n,k} \mid k, n \in \mathcal{K}, \ n < k\}$ become $\{\bar{\gamma}_i \mid i \in \{1, \ldots, D\}\}$. For a given bijective function f, the index i is defined as $i \triangleq f(n,k)$. We can choose for example the function $f(n,k) = \frac{(k-1)(k-2)}{2} + n$.

We recall the optimization problem of Theorem 3 but, for sake of clarity, we present it in vector notation. For that, let the vector containing the time-sharing variables $\bar{\gamma}_i$ be denoted by $\boldsymbol{\gamma}$, i.e., $\boldsymbol{\gamma} \triangleq [\bar{\gamma}_1, \bar{\gamma}_2, \ldots, \bar{\gamma}_D]$. Similarly, we define the vector $\mathbf{F}^\alpha$ as the concatenation of the effective DoF of each Transmission Mode (see (15)), such that $\mathbf{F}^\alpha \triangleq [\mathbf{F}_1^\alpha, \ldots, \mathbf{F}_D^\alpha]$, and $\mathbf{F}_i^\alpha = 1 + (k-1)\alpha^{(n)}$, with k, n given by $(n,k) = f^{-1}(i)$. Finally, the vector of terms $d_{n,k} \triangleq 1 - \alpha^{(n)} - k(k-n-1)\alpha^{(n)}$ for the constraint (17) is denoted as $\boldsymbol{d}$. Hence, the optimization problem of Theorem 3 can be expressed as

$$
\text{DoF}^{\text{APZF}}(\alpha) = \underset{\boldsymbol{\gamma}}{\text{maximize}} \quad \mathbf{F}^\alpha \boldsymbol{\gamma} \tag{119}
$$

$$
\text{subject to} \quad \|\boldsymbol{\gamma}\|_1 = 1 \tag{120}
$$

$$
\boldsymbol{\gamma} \succeq 0 \tag{121}
$$

$$
\boldsymbol{d}\boldsymbol{\gamma} \geq 0, \tag{122}
$$

where $\mathbf{F}^\alpha$ and $\boldsymbol{d}$ are constant vectors. Let us remind that if a linear programming problem has an optimal solution then it is an extreme point of the feasible set [53].

The feasible set given by conditions (120)-(121), which is denoted by $\mathcal{C}$, is the *probability simplex* [54] determined by the unit vectors $e_1, \ldots, e_D \in \mathbb{R}^D$, and consequently it is a $(D-1)$-dimensional simplex. On the other hand, (122) represents a half-space determined by the vector hyperplane [54] given by $\boldsymbol{d}\boldsymbol{\gamma} = 0$. The vector hyperplane $\boldsymbol{d}\boldsymbol{\gamma} = 0$ is denoted as $\mathcal{V}$. We can have different cases depending on how the *probability simplex* $\mathcal{C}$ and the half-space determined by the hyperplane $\mathcal{V}$ intersect. Namely:

1) If $\mathcal{C} \cap \{\boldsymbol{\gamma} \mid \boldsymbol{d}\boldsymbol{\gamma} \geq 0\} = \mathcal{C}$ (i.e., $\mathcal{C}$ is a subset of the half-space), the feasible region is $\mathcal{C}$ and the extreme points are the unit vectors e_i. Therefore, the solution of the problem uses only a single mode because the only non-zero variable in a unit vector e_i is the i-th variable.

2) If $\mathcal{C} \cap \{\boldsymbol{\gamma} \mid \boldsymbol{d}\boldsymbol{\gamma} \geq 0\} = \emptyset$, there is not feasible solution. However, this is not possible since we have shown that this linear program is always feasible, just choosing $\gamma_{k-1,k} = 1$, with $k \in \{2, \ldots, K\}$.

3) If $\mathcal{C} \cap \{\boldsymbol{\gamma} \mid \boldsymbol{d}\boldsymbol{\gamma} \geq 0\} \subset \mathcal{C}$, we need to prove that all the extreme points of the resulting set satisfy (118). Those extreme points will either be the extreme points of $\mathcal{C}$ or belong to the intersection between $\mathcal{C}$ and $\mathcal{V}$.

From linear algebra, we know that the intersection of an l-dimensional and an m-dimensional sub-space in the n-dimensional space $\mathbb{R}^n$ has dimension p_i such that

$$
p_i \geq l + m - n. \tag{123}
$$

Thus, in order to obtain the extreme points ($p_i = 0$) of the feasible set, we must obtain the intersection in the space $\mathbb{R}^D$ between $\mathcal{V}$ ($m = D - 1$) and the *edges* of $\mathcal{C}$, i.e., the 1-faces (segments) that define $\mathcal{C}$. The edges of $\mathcal{C}$ are segments that connect two points with a single non-zero variable (the unit vectors), and therefore they belong to a line of only two non-zero variables. Given that the intersection of $\mathcal{V}$ with one edge must be a point of the edge, it holds that all the extreme points have at most two non-zero variables, what means that they satisfy (118). Therefore, Corollary 1 is proven. From the previous analysis, it follows that the feasibility set is convex.

Moreover, as Theorem 3 is always composed of at most two Transmission Modes, it can be expressed as the following integer linear program:

$$
\underset{\substack{k_1,n_1, \\ k_2,n_2}}{\text{maximize}} \ 1 + \rho(k_1-1)\alpha^{(n_1)} + (1-\rho)(k_2-1)\alpha^{(n_2)}
$$

$$
\begin{aligned}
\text{subject to} \quad & k_1, k_2 \in \{2, \ldots, K\}, \\
& n_1 \in \{1, \ldots, k_1 - 1\}, \\
& n_2 \in \{1, \ldots, k_2 - 1\} \mid d_{n_2,k_2} \geq 0,
\end{aligned}
$$

where ρ is defined as

$$
\rho \triangleq \begin{cases} 1 & \text{if } d_{n_1,k_1} \geq 0 \\ \dfrac{d_{n_2,k_2}}{d_{n_2,k_2} - d_{n_1,k_1}} & \text{otherwise.} \end{cases} \tag{124}
$$

REFERENCES

[1] P. de Kerret and D. Gesbert, "Network MIMO: Transmitters with no CSI can still be very useful," in *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016, pp. 1466–1470.

[2] A. Bazco, P. de Kerret, D. Gesbert, and N. Gresset, "Méthode de transmission robuste au partage imparfait de l'information de canal entre transmetteurs," in *2017 GRETSI*, Septembre 2017.

[3] S. A. Jafar and M. J. Fakhereddin, "Degrees of Freedom for the MIMO Interference Channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2637–2642, July 2007.

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

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

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

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

[8] P. de Kerret, D. Gesbert, J. Zhang, and P. Elia, "Optimally bridging the gap from delayed to perfect CSIT in the K-user MISO BC," in *Proc. IEEE Information Theory Workshop (ITW)*, Sep. 2016, pp. 300–304.

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

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

[11] Z. Becvar, M. Vondra, P. Mach, J. Plachy, and D. Gesbert, "Performance of mobile networks with UAVs: Can flying base stations substitute ultra-dense small cells?" in *Proc. 23th European Wireless Conference (EW)*, May 2017, pp. 1–7.

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

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

[14] C. Hao and B. Clerckx, "Degrees-of-Freedom of the K-user MISO interference channel with delayed local CSIT," in *Proc. IEEE International Conference on Communications (ICC)*, June 2015, pp. 4217–4222.

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

[16] A. Vahid and R. Calderbank, "Two-user erasure Interference Channels with local delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 4910–4923, Sept. 2016.

[17] A. Lapidoth and Y. Steinberg, "The multiple access channel with two independent states each known causally to one encoder," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2010, pp. 480–484.

[18] R. Kolte, A. Özgür, and H. Permuter, "Cooperative binning for semideterministic channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1231–1249, Mar. 2016.

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

[20] B. Li, Z. Fei, and Y. Zhang, "UAV communications for 5G and Beyond: Recent advances and future trends," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2241–2263, Apr. 2019.

[21] Veniam. (2019, Jul.) Moving Terabytes of data between vehicles and the cloud. [Online]. Available: https://veniam.com/

[22] C. T. K. Ng, D. Gunduz, A. J. Goldsmith, and E. Erkip, "Distortion minimization in Gaussian layered broadcast coding with successive Refinement," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5074–5086, 2009.

[23] O. E. Ayach and R. W. Heath, "Interference alignment with analog channel state feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 626–636, Feb. 2012.

[24] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, Dec 2013.

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

[26] N. Golrezaei, A. G. Dimakis, and A. F. Molisch, "Scaling behavior for device-to-device communications with distributed caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4286–4298, 2014.

[27] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proc. Workshop on Mobile Cloud Computing (MCC)*, 2012, pp. 13–16.

[28] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: Insights and challenges," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 152–160, Apr. 2015.

[29] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug 2014.

[30] J. Zhang and P. Elia, "Fundamental limits of cache-aided wireless BC: Interplay of coded-caching and CSIT feedback," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 3142–3160, May 2017.

[31] M. Chen, Y. Qian, Y. Hao, Y. Li, and J. Song, "Data-driven computing and Caching in 5G networks: Architecture and delay analysis," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 70–75, Feb. 2018.

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

[33] Q. Li, P. de Kerret, D. Gesbert, and N. Gresset, "Robust Regularized ZF in cooperative Broadcast Channel under distributed CSIT," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1845–1860, Mar. 2020.

[34] P. de Kerret, R. Fritzsche, D. Gesbert, and U. Salim, "Robust precoding for network MIMO with hierarchical CSIT," in *2014 11th Int. Symp. Wireless Commun. Systems (ISWCS)*, Aug 2014, pp. 987–991.

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

[36] A. Bazco-Nogueras, P. de Kerret, D. Gesbert, and N. Gresset, "Distributed CSIT does not reduce the Generalized DoF of the 2-user MISO Broadcast Channel," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 685–688, June 2019.

[37] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[38] P. de Kerret and D. Gesbert, "Robust decentralized joint precoding using team deep neural network," in *Proc. IEEE Int. Symp. on Wireless Commun. Systems (ISWCS)*, Aug. 2018.

[39] M. Kim, P. de Kerret, and D. Gesbert, "Learning to cooperate in decentralized wireless networks," in *Proc. IEEE Asilomar Conf. on Signals, Systems and Computers (ACSSC)*, Nov. 2018.

[40] H. Lee, S. Hyun, and T. Q. S. Quek, "Deep learning for distributed optimization: Applications to wireless resource management," *IEEE J. Sel. Areas Commun.*, 2019.

[41] D. Gunduz, P. de Kerret, N. D. Sidiropoulos, D. Gesbert, C. Murthy, and M. van der Schaar, "Machine Learning in the Air," *IEEE J. Sel. Areas Commun.*, 2019.

[42] R. Radner, "Team decision problems," *The Annals of Mathematical Statistics*, 1962.

[43] S. Lin and I.-H. Wang, "Gaussian Broadcast Channels with intermittent connectivity and hybrid state information at the transmitter," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6362–6383, Sept 2018.

[44] S. Lashgari, R. Tandon, and S. Avestimehr, "MISO Broadcast Channel with hybrid CSIT: Beyond two users," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7056–7077, Dec 2016.

[45] A. G. Davoodi and S. A. Jafar, "CSIT thresholds for collapse of Degrees of Freedom in wireless networks," in *2018 IEEE International Conference on Communications (ICC)*, May 2018, pp. 1–6.

[46] E. Piovano, H. Joudeh, and B. Clerckx, "Overloaded multiuser MISO transmission with imperfect CSIT," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Nov 2016, pp. 34–38.

[47] C. Hao, B. Rassouli, and B. Clerckx, "Achievable DoF regions of MIMO networks with imperfect CSIT," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6587–6606, Oct 2017.

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

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

[50] C. S. Vaze and M. K. Varanasi, "The Degree-of-Freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5354–5374, Aug. 2012.

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

[52] P. Billingsley, *Probability and Measure*, ser. Wiley Series in Probability and Statistics. Wiley, 1995.

[53] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming.* Springer Publishing Company, Incorporated, 2015.

[54] S. Boyd and L. Vandenberghe, *Convex Optimization.* New York, NY, USA: Cambridge University Press, 2004.

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

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

David Gesbert (F'10) received the Ph.D. degree from Ecole Nationale Superieure des Telecommunications, France, in 1997. From 1997 to 1999, he was with the Information Systems Laboratory, Stanford University. He was then a Founding Engineer of Iospan Wireless Inc., a Stanford spinoff pioneering MIMO-OFDM (now Intel). Before joining EURECOM in 2004, he was with the Department of Informatics, University of Oslo, as an Adjunct Professor. He is currently a Professor and the Head of the Communication Systems Department, EURECOM.

He has published about 300 articles and 25 patents, some of them winning the 2019 ICC Best Paper Award, the 2015 IEEE Best Tutorial Paper Award (Communications Society), the 2012 SPS Signal Processing Magazine Best Paper Award, the 2004 IEEE Best Tutorial Paper Award (Communications Society), the 2005 Young Author Best Paper Award for Signal Processing Society journals, and paper awards at conferences 2011 IEEE SPAWC and 2004 ACM MSWiM. He was the Technical Program Co-Chair of ICC2017. He was named a Thomson-Reuters Highly Cited Researchers in Computer Science. Since 2015, he has held the ERC Advanced grant PERFUME on the topic of smart device communications in future wireless networks. He is a Board Member of the OpenAirInterface (OAI) Software Alliance. Since 2019, he has been heading the Huawei-funded Chair on Advanced Wireless Systems Towards 6G Networks. He sits on the Advisory Board of the HUAWEI European Research Institute. He held visiting professor positions in KTH in 2014 and TU Munich in 2016. Since 2017, he has also been a visiting Academic Master within the Program 111 with the Beijing University of Posts and Telecommunications as well as a member of the Joint BUPT-EURECOM Open5G Lab.

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