Degrees of Freedom of the Network MIMO Channel With Distributed CSI

Paul de Kerret, *Student Member, IEEE* and David Gesbert, *Fellow, IEEE,*

Abstract—In this work, we discuss the joint precoding with finite rate feedback in the so-called network MIMO where the TXs share the knowledge of the data symbols to be transmitted. We introduce a distributed channel state information (DCSI) model where each TX has its own local estimate of the overall multi-user MIMO channel and must make a precoding decision solely based on the available local CSI. We refer to this channel as the *DCSI-MIMO channel* and the precoding problem as *distributed precoding*. We extend to the DCSI setting the work from Jindal in [1] for the conventional MIMO Broadcast Channel (BC) in which the number of Degrees of Freedom (DoFs) achieved by Zero Forcing (ZF) was derived as a function of the scaling in the logarithm of the Signal-to-Noise Ratio (SNR) of the number of quantizing bits. Particularly, we show the seemingly pessimistic result that the number of DoFs *at each user* is limited by the worst CSI across all users and across all TXs. This is in contrast to the conventional MIMO BC where the number of DoFs at one user is solely dependent on the quality of the estimation of his own feedback. Consequently, we provide precoding schemes improving on the achieved number of DoFs. For the two-user case, the derived novel precoder achieves a number of DoFs limited by the best CSI accuracy across the TXs instead of the worst with conventional ZF. We also advocate the use of *hierarchical quantization* of the CSI, for which we show that considerable gains are possible. Finally, we use the previous analysis to derive the DoFs optimal allocation of the feedback bits to the various TXs under a constraint on the size of the aggregate feedback in the network, in the case where conventional ZF is used.

Index Terms—Network MIMO, Degrees of freedom, finite rate feedback, Zero Forcing, Broadcast Channel, Distributed, Imperfect CSI

I. Introduction

Network MIMO channel, or multicell MIMO channels, whereby multiple interfering transmitters (TXs) share user messages and allow for joint precoding (downlink), are currently considered for next generation wireless networks [2]–[4]. With perfect message and channel state information (CSI) sharing, the different TXs can be seen as a unique virtual multiple-antenna array serving all receivers (RXs), in a multiple-antenna broadcast channel (BC) fashion.

Although the sharing of user data symbols can be made possible in certain situations, such as cellular networks with a pre-existing backbone infrastructure where user packets can be routed to several base stations simultaneously, the obtaining of accurate CSI at the TXs is made difficult due to the finite quantizing effects over the feedback channels and the

limited capability of signaling between TXs to exchange the CSI. In addition, CSI exchange necessarily introduces further degradation due to latency effects over inter-TX links [5].

This situation gives rise to an interesting information theoretic framework whereby a MIMO broadcast channel is formed (due to the assumed perfect user message sharing among the various TXs), yet the individual TXs composing the distributed multiple-antenna array have access to individual CSI estimates, possibly different from each other, and possibly of different quality (statistically). In this paper, we refer to this channel as the *distributed CSI (DCSI)-MIMO channel*. We emphasize the difference between this CSI model and the previously studied CSI models such as the so-called imperfect limited CSI [1] or the delayed CSI model [6] where the TX antennas are assumed to share ideally the *same* imperfect channel knowledge.

Note that the sharing of the symbols via finite capacity links between the cooperating TXs has been discussed in recent works [7]–[10]. This problem represents in itself a challenging topic, and we consider in the sequel perfect sharing of the users symbols.

For the conventional MIMO BC, the impact of limited feedback [1], [11]–[16] and the derivation of robust solutions [17], [18] have been investigated, with later extensions to the multicell coordinated beamforming case [19] and the multicell MIMO case [20]–[22].

More recently, the optimization of the feedback allocations to the different users has been the focus of a large interest. It has been studied in conventional MIMO BCs [23], in multicell settings with coordinated beamforming [24]–[27], in multicell MIMO networks [28]–[30] and in interfering BCs [31], [32].

Yet, as mentioned before, these papers always consider *perfect sharing* between the TXs precoding jointly the signal. In contrast, we consider here that each TX has its own imperfect estimation of the multi-user channel but all the TXs jointly precode the user's data symbols. This gives rise to a very different transmission setting which can be seen as a *team decision problem* [33]. Indeed, the precoder must cope not only with the inaccuracy of the CSI due to the limited feedback channel capacity but also with the distributiveness of the CSI and the precoding. Each TX emits one component of the transmit signal vector which it computes based on its *own* channel estimate. As is pointed out in this work, the *discrepancies* between the channel estimates obtained by the different TXs are particularly detrimental to the channel capacity, and even to the Degrees of Freedom (DoFs), if not accounted for in the precoding design.

The DCSI-MIMO scenario has meaningful applications to network MIMO schemes in cellular networks or MIMO based multi-TX cooperation in general. It was first studied in [34],

and a tractable discrete optimization at finite SNR was derived. However, the approach in [34] does not lend itself to a more general performance analysis, thus giving limited insight for an improved design.

In this paper, we consider the performance of precoding schemes over the DCSI-MIMO channel from a DoFs perspective. The number of DoFs represents the slope with which the rate increases with the SNR in the high SNR regime. Even though it is based on the high SNR analysis, it has been used widely used to gain insight into the wireless transmission thanks to its analytical tractability [6], [35]. By essence, the DoFs analysis is not impacted by the unequal pathloss, which can put in question its practical signification in some settings. When all the wireless links present the same pathloss as it is the case in this work, this does not represent an issue. To extend the DoFs analysis to settings with large pathloss differences, it is then more adequate to use the notion of *generalized DoFs* [36] which takes the pathloss differences into account. We also assume that our system model is isolated from the rest of the world. In a practical scenario, it follows from the impossibility to serve jointly all the users that there is inevitably interference coming from outside the cooperation area. This implies that the number of DoFs is always zero [37]. The number of DoFs derived inside the cooperation cluster is then representative solely up to an SNR at which point the interference from outside the cooperation area leads to the saturation of the rate.

Our work generalizes to the case of distributed CSI the finite rate feedback study by Jindal [1] for the conventional multiple-antenna BC. In [1], the author derives the number of DoFs as a function of the number of feedback (quantizing) bits exploited by each RX and shows that the number of bits must grow with the logarithm of the SNR in order to preserve the full number of DoFs, using ZF precoding arguments. We also consider ZF schemes as they are known to achieve maximum number of DoFs in wide settings[1]. Particularly, a necessary and sufficient feedback of the CSI estimation error for achieving the maximum number of DoFs is derived in [11] for the compound multiple-antenna BC. This condition is the same as the sufficient condition provided in [1]. Thus, no other precoding scheme can achieve the maximal number of DoFs with a lower feedback scaling. This confirms the efficiency of ZF in terms of number of DoFs. As a consequence, we aim in this paper at answering the fundamental questions *"Does conventional ZF also perform well in the distributed MIMO setting?"*, and *"How can we make it more robust in that setting?"*

Specifically, the main contributions read as follows. Let the number of bits quantizing the estimate at TX j of the normalized channel $\tilde{h}_i^{\mathrm{H}}$ of user i be $\alpha_i^{(j)}(K-1)\log_2(P)$ with $\alpha_i^{(j)} \in [0,1]$ and K the number of users. Then, we show that in a block fading Rayleigh channel:

- The number of DoFs achieved at RX i with conventional

ZF is equal to $\min_{i,j\in\{1,\dots,K\}} \alpha_i^{(j)}$. Hence, the worst accuracy across all the estimates limits the number of DoFs at each user. This is a pessimistic result and shows a different behavior compared to the conventional MIMO BC.

- We provide a precoding scheme improving the number of DoFs. In the two-user case, the number of DoFs with the novel precoding scheme is limited by the best accuracy of the CSI across the two TXs instead of the lowest with conventional ZF.

- To improve the number of DoFs achieved with more users, we introduce a concept of hierarchical quantization of the CSI and we show that this leads to a dramatic improvement of the number of DoFs.

- Under a total feedback constraint and with ZF schemes, we derive the number of DoFs maximizing allocation of the feedback bits toward each TX.

Note that this paper serves to generalize preliminary results that were presented in [38].

Notations: We denote by $\Pi_{\mathbf{A}}(\bullet)$ and $\Pi_{\mathbf{A}}^{\perp}(\bullet)$ the orthogonal projectors over the subspace spanned by the matrix $\mathbf{A}$ and over its orthogonal complement, respectively. $\bar{i}$ denotes the complementary indice of i when only two users are considered, i.e., $\bar{i} = i \mod 2 + 1$. $\| \bullet \|_{\mathrm{F}}$ designates the Frobenius norm while $\mathcal{N}(\mu, \sigma^2)$ denotes the complex circularly symmetric Gaussian distribution with mean μ and variance σ^2. We also denote the ith element of a vector $\boldsymbol{a}$ by $\{\boldsymbol{a}\}_i$ and the (i,j)th element of a matrix $\mathbf{A}$ by $\{\mathbf{A}\}_{ij}$. Additionally, we use the notation $\lesssim$ to denote a relation of order which holds true asymptotically. We also write $f(x) = o(g(x))$ (resp. $f(x) = O(g(x))$) to represent the fact that $\lim_{x\to\infty} f(x)/g(x) = 0$ (resp. $\lim_{x\to\infty} |f(x)|/|g(x)| \leq a$, with $a > 0$). We also write $f(x) \sim g(x)$ to denote the fact that $f(x) = g(x) + o(g(x))$.

II. SYSTEM MODEL

A. Multicell MIMO

We consider a joint downlink transmission from K TXs to K RXs using linear precoding and single user decoding. For ease of exposition, the TXs and the RXs are equipped with only one antenna, but the principal elements of our approach could extend in principle to more antennas at the TXs. Similarly, we consider a Rayleigh fading scenario but the approach derived should be valid in many other fading scenarios. The transmission can be described as

$$
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_K \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}_1^{\mathrm{H}} \\ \boldsymbol{h}_2^{\mathrm{H}} \\ \vdots \\ \boldsymbol{h}_K^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_K \end{bmatrix} \tag{1}
$$

where $\boldsymbol{y} \triangleq [y_1, \dots, y_K]^{\mathrm{T}} \in \mathbb{C}^{K\times 1}$ contains the received signals at the RXs, the vector $\boldsymbol{x} \triangleq [x_1, \dots, x_K]^{\mathrm{T}} \in \mathbb{C}^{K\times 1}$ is defined such that x_j is the signal transmitted by TX j, and $\boldsymbol{\eta} \triangleq [\eta_1, \dots, \eta_K]^{\mathrm{T}} \in \mathbb{C}^{K\times 1}$ contains the noise realizations at the RXs and has its entries i.i.d. as $\mathcal{N}(0, 1)$.

The vector $\boldsymbol{h}_i^{\mathrm{H}} \in \mathbb{C}^{1\times K}$ is the channel from all TXs to the i-th RX and define the normalized channel to user i

---

[1]Note that the selection of the set of users actually transmitting during one time slot is not considered in this work. In fact the formula for the number of DoFs provided in this work can be used to derive a set of transmitting TX achieving a good number of DoFs, i.e., to use a good combination of ZF precoding and time sharing.

as $\tilde{\boldsymbol{h}}_i \triangleq \boldsymbol{h}_i/\|\boldsymbol{h}_i\|$. We also define the multi-user channel matrix $\mathbf{H} \triangleq [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]^{\mathrm{H}}$ and its normalized counterpart $\tilde{\mathbf{H}} \triangleq [\tilde{\boldsymbol{h}}_1, \ldots, \tilde{\boldsymbol{h}}_K]^{\mathrm{H}}$.

The channel is assumed to be block fading and the entries of the channel matrix $\mathbf{H}$ to be i.i.d. as $\mathcal{N}(0, 1)$, modeling a Rayleigh fading channel. The transmitted signal $\boldsymbol{x}$ is obtained from the vector of transmit symbols $\boldsymbol{s} \triangleq [s_1, \ldots, s_K]^{\mathrm{T}} \in \mathbb{C}^{K \times 1}$ (whose entries are taken as i.i.d. $\mathcal{N}(0, 1)$) as

$$\boldsymbol{x} = \mathbf{T}\boldsymbol{s} = \begin{bmatrix} \boldsymbol{t}_1 & \ldots & \boldsymbol{t}_K \end{bmatrix} \begin{bmatrix} s_1 \\ \vdots \\ s_K \end{bmatrix} \qquad (2)$$

where $\mathbf{T} \in \mathbb{C}^{K \times K}$ is the multi-user precoding matrix and $\boldsymbol{t}_i \in \mathbb{C}^{K \times 1}$ is the beamforming vector used to transmit s_i. Even though a per-TX power constraint is the most relevant power constraint in the multicell setting, we consider a sum power constraint $\|\mathbf{T}\|_{\mathrm{F}}^2 = P$. We also assume for simplicity and symmetry that all data streams are allocated with an equal amount of power so that $\boldsymbol{t}_i = \sqrt{P/K}\boldsymbol{u}_i$ with $\|\boldsymbol{u}_i\|^2 = 1$. These choices can be done without restricting the scope of this work because they do not have any impact on the number of DoFs[2]. We will study the ergodic rate averaged over the random codebooks $\mathcal{W}_i^{(j)}$ used for the CSI Random Vector Quantization (RVQ), as detailed in Subsection II-B. The ergodic rate for RX i reads then as

$$R_i(P) \triangleq \mathrm{E}_{\mathbf{H}, \{\mathcal{W}_i^{(j)}\}_{i,j}} \left[\log_2 \left(1 + \frac{|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_i|^2}{1 + \sum_{\ell=1, \ell \neq i}^K |\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_\ell|^2} \right) \right].$$
$$(3)$$

To achieve the maximal number of DoFs we aim at removing completely the interference at all the RXs, i.e., at having

$$\forall i \in \{1, \ldots, K\}, \qquad \sum_{\ell=1, \ell \neq i}^K |\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_\ell|^2 = 0. \qquad (4)$$

From (4) and the equal power allocation, there is no coupling between the optimizations of the beamforming vectors $\boldsymbol{t}_i$ which can then be carried out in parallel. The number of DoFs achieved *at RX* i is defined as

$$\mathrm{DoF}_i \triangleq \lim_{P \to \infty} \frac{R_i(P)}{\log_2(P)}. \qquad (5)$$

and the total number of DoFs is $\mathrm{DoF} \triangleq \sum_{i=1}^K \mathrm{DoF}_i$. From the above definition of the number of DoFs and definition (3), we can directly obtain that $\forall i \in \{1, \ldots, K\}$,

$$\mathrm{DoF}_i = 1 - \lim_{P \to \infty} \mathrm{E}_{\mathbf{H}, \{\mathcal{W}_i^{(j)}\}_{i,j}} \left[\frac{\log_2 \left(\sum_{\ell \neq i} |\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_\ell|^2 \right)}{\log_2(P)} \right].$$
$$(6)$$

B. Distributed CSI

1) CSI Scaling Coefficients: We assume a limited CSI setting where channel estimate inaccuracies are modeled using

---

quantized feedback. Furthermore, a *distributed* CSI model is defined here in the sense that each TX has its own individual estimate of the normalized channel $\tilde{\boldsymbol{h}}_i$ to RX i. Moreover, the estimates for the different channel vectors $\tilde{\boldsymbol{h}}_i$ are also a priori of different qualities at each TX, i.e., quantized with codebooks of different sizes. We denote by $\tilde{\boldsymbol{h}}_i^{(j)}$ the estimate of the normalized channel vector $\tilde{\boldsymbol{h}}_i$ acquired at TX j. The quantized feedback consists of $B_i^{(j)}$ bits which are used to index a vector in the codebook $\mathcal{W}_i^{(j)}$ made of $2^{B_i^{(j)}}$ elements. We also define $\tilde{\mathbf{H}}^{(j)} \triangleq [\tilde{\boldsymbol{h}}_1^{(j)}, \ldots, \tilde{\boldsymbol{h}}_K^{(j)}]^{\mathrm{H}}$ as the estimate of the total normalized multi-user channel at TX j.

This setting arises in the context of multi-TX cooperation (e.g. Network MIMO [4]) where either (i) all TXs obtain a version of the whole CSI matrix through independent feedback channels (in which case the quality of the uplink feedback channel determines the quality of the individual CSI estimates) or (ii) each TX obtains some portion of the CSI and exchange it through limited rate links or/and with some latency to the other TXs.

In the conventional MIMO BC, it is shown in [1] that the number of quantization bits should scale indefinitely with the logarithm of the SNR in order to achieve a strictly positive number of DoFs when using ZF precoding. Thus, we also focus on the *scaling in the logarithm of the SNR* of the number of quantization bits of all the channel estimates. We introduce the *CSI scaling matrix* $\boldsymbol{\alpha} \in \mathbb{R}^{K \times K}$ with its (i, j)-th element defined as

$$\alpha_i^{(j)} \triangleq \lim_{P \to \infty} \frac{B_i^{(j)}}{(K-1)\log_2(P)}. \qquad (7)$$

Hence, $\alpha_i^{(j)}$ denotes the scaling of the number of bits used to describe the channel of user i at TX j. Since $B_i^{(j)}$ is a design parameter, the limit in (7) can be seen to always exist. We furthermore assume that the CSI scaling matrix $\boldsymbol{\alpha}$ is known to all the TXs.

Remark: We will always consider for notational clarity $\alpha_i^{(j)} \in [0, 1]$ as the range of interest. This follows from the fact that if $\alpha_i^{(j)} = 1$, it then holds $|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_\ell|^2 = O(1)$ for $\ell \neq i$ [1]. The accuracy of CSI resulting from a CSI scaling coefficient equal to one is sufficient for the interference to remain bounded. Thus, increasing the number of CSI feedback bits to get $\alpha_i^{(j)} > 1$ does not increase the number of DoFs. This corresponds to a well known result for the conventional MIMO BC in [1]. It follows that in all the subsequent results, the scaling coefficients $\alpha_i^{(j)}$ should be replaced by $\min(\alpha_i^{(j)}, 1)$ so as to be valid for arbitrary values for the CSI scaling coefficients. This is not done to keep the notations as clear as possible.

2) Random Vector Quantization for the DCSI-MIMO Channel: We consider RVQ where random codebooks are used to quantize the channels. This follows a result in [1] for the conventional MIMO BC, stating that in the case of two antennas at the TX, no codebook can achieve a better number of DoFs than the number of DoFs achieved with RVQ. RVQ is also shown to be optimal for the point-to-point MIMO link as the number of antennas tends to infinity both at the TX and the RX [39]. Finally, RVQ is interesting because it gives an

---

[2]Indeed, it is always possible to scale the total power used when considering the sum power constraint so as to fulfill the per-TX power constraint without impacting the number of DoFs. Similarly, optimally allocating the power does not change the number of DoFs.

achievable lower bound.

In most of the works regarding the conventional MIMO BC, a codeword w is selected for quantizing the unit-norm vector $\tilde{h}_i$ if it maximizes the amplitude of the inner product $|\tilde{h}_i^{\mathrm{H}} w|$. However, in the DCSI-MIMO channel, this quantization scheme is less adequate because the objective is invariant by multiplication of the codeword by a unit-norm complex number. This represents a problem since a different estimate is received at each TX, and this phase invariance creates an ambiguity between the estimates. This is very harmful for the transmission scheme and, in fact, if such a quantization scheme is used, it can be easily shown that the channel estimate obtained is essentially useless for joint precoding.

Thus, another quantization scheme is preferred and the quantized channel $\tilde{h}_i^{(j)}$ is instead obtained in the optimum L_2 norm sense:

$$\tilde{h}_i^{(j)} = \underset{w \in \mathcal{W}_i^{(j)}}{\operatorname{argmin}} \|w - \tilde{h}_i\|. \tag{8}$$

Using directly (8) leads to lower performance as the phase of the channel also impacts the performance, which is different from the quantization in a Grassmannian space. To recover similar performance as the quantization scheme conventionally used, we multiply each element of the codebooks as well as each normalized channels by a complex unit-norm number in order to let their first coefficient be real valued. A detailed analysis of this quantization scheme is provided in Appendix X-A.

C. Distributed Precoding

In the DCSI-MIMO channel, each TX has a different estimate of the multi-user channel $\mathbf{H}$ and controls only one antenna. Thus, each TX uses its CSI to compute a certain precoding matrix from which it extracts the coefficient corresponding to its antenna. We denote the overall multi-user precoder computed at TX j as $\mathbf{T}^{(j)} \triangleq \left[t_1^{(j)} \quad \ldots \quad t_K^{(j)} \right]$ where $t_i^{(j)}$ is the beamforming vector designed to transmit symbol s_i.

Note that although a given TX j may compute the whole precoding matrix $\mathbf{T}^{(j)}$, only the j-th row $e_j^{\mathrm{T}} \mathbf{T}^{(j)}$ will be used in practice, since TX j transmits only $x_j = e_j^{\mathrm{T}} \mathbf{T}^{(j)} s$. Finally, the effective precoder is then given by

$$\mathbf{T} \triangleq \begin{bmatrix} t_1 & \ldots & t_K \end{bmatrix} \triangleq \begin{bmatrix} e_1^{\mathrm{T}} \mathbf{T}^{(1)} \\ e_2^{\mathrm{T}} \mathbf{T}^{(2)} \\ \vdots \\ e_K^{\mathrm{T}} \mathbf{T}^{(K)} \end{bmatrix}. \tag{9}$$

The main elements of the transmission in the distributed CSI MIMO channel are illustrated in Fig. 1.

III. REVIEW OF THE RESULTS IN THE CONVENTIONAL MIMO BC

In this section, we recall briefly the main results from [1] on the number of DoFs achieved with finite rate feedback in the conventional MIMO BC. This will be helpful to understand the differences between the conventional MIMO BC and the distributed CSI setting which is the main focus of this work.
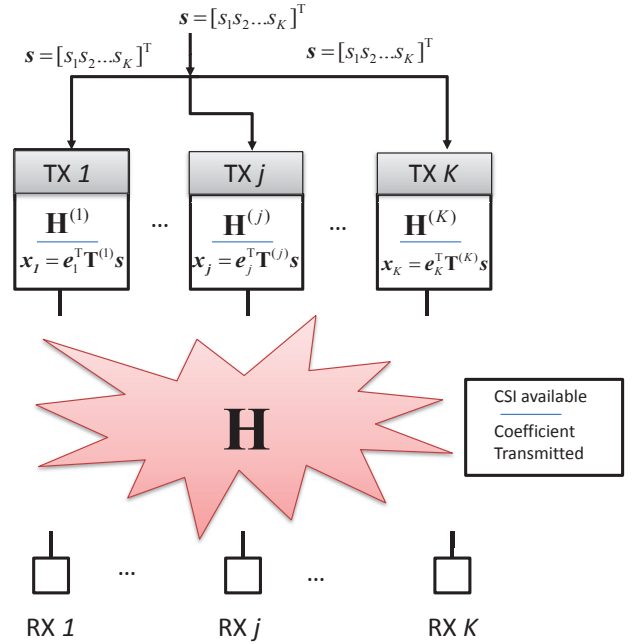

Fig. 1. Distributed precoding in the DCSI-MIMO channel.

Hence, we consider in this section a conventional MIMO BC where M TXs are colocated and share the *same* channel estimate. For this setting, we need to use notations which are different from the ones previously introduced for the DCSI-MIMO channel. We denote by $\hat{h}_i$ the channel estimate of $\tilde{h}_i$ obtained with B_i bits. Following [1], the channel estimate is obtained from

$$\hat{h}_i = \underset{w \in \mathcal{W}_i^{\mathrm{BC}}}{\operatorname{argmax}} |w^{\mathrm{H}} \tilde{h}_i|^2 \tag{10}$$

where $\mathcal{W}_i^{\mathrm{BC}}$ is a random codebook containing 2^{B_i} unit-norm vectors isotropically distributed in $\mathbb{C}^{K \times 1}$. We provide now the main result.

Theorem 1. *[1] In the MIMO BC with M antennas, if the channel estimate $\hat{h}_i$ is obtained from the quantization scheme* (10) *with $B_i = \alpha_i(M-1)\log_2(P)$, the number of DoFs achieved with ZF is given by*

$$\mathrm{DoF}^{BC} = \sum_{i=1}^{M} \alpha_i. \tag{11}$$

This result was given in [1] for $\alpha_i = \alpha$ but the extension to different α_i follows directly from the proof in [1]. The extension to Theorem 1 has been suggested in [40] where the same formula for the number of DoFs is derived in the case where DPC is used instead of ZF.

We will now derive the equivalent result of Theorem 1 for the DCSI-MIMO channel where the TXs do not share the same channel estimates.

IV. Zero Forcing in the DCSI-MIMO Channel with Two Users

As a starting point we consider the particular configuration with only two users. This setting is interesting for two main reasons. Firstly, the exposition is simpler in that case while most of the insights are the same as in the general case, and secondly this scenario makes it possible to obtain stronger results.

In the conventional multiple-antenna BC with imperfect CSI, the number of DoFs with ZF has been derived and shown to be defined by the CSI scaling. In the DCSI-MIMO channel, the CSI scaling of each channel vector $\tilde{\boldsymbol{h}}_i$ is different at each TX. One central goal of our work consist in determining how the formula for the number of DoFs in the conventional MIMO BC generalizes to the DCSI-MIMO channel. This would then lead us to evaluate whether ZF is in that case a performing solution and if not, whether one can find better solutions.

A. Conventional Zero Forcing

In the DCSI-MIMO channel, the conventional ZF precoder is made of the beamformer $\boldsymbol{t}_i^{\mathrm{cZF}} \triangleq [\boldsymbol{e}_1^{\mathrm{T}} \boldsymbol{t}_i^{\mathrm{cZF}(1)}, \boldsymbol{e}_2^{\mathrm{T}} \boldsymbol{t}_i^{\mathrm{cZF}(2)}]^{\mathrm{T}}$ to transmit s_i, with its elements defined in an intuitive way as

$$\boldsymbol{t}_i^{\mathrm{cZF}(j)} \triangleq \sqrt{\frac{P}{2}} \frac{\Pi_{\tilde{\boldsymbol{h}}_{\bar{i}}^{(j)}}^{\perp}\left(\tilde{\boldsymbol{h}}_i^{(j)}\right)}{\|\Pi_{\tilde{\boldsymbol{h}}_{\bar{i}}^{(j)}}^{\perp}\left(\tilde{\boldsymbol{h}}_i^{(j)}\right)\|}, \quad j \in \{1, 2\}. \tag{12}$$

The interpretation behind conventional ZF is that each TX applies ZF using its own CSI implicitly assuming that the other TX shares the same CSI estimate. Our first result given in the following theorem relates the number of DoFs achieved with such a precoding strategy.

Theorem 2. *Conventional ZF achieves the number of DoFs*

$$\mathrm{DoF}^{cZF} = 2 \min_{i,j \in \{1,2\}} \alpha_i^{(j)}. \tag{13}$$

Proof: A detailed proof is provided in Appendix X-B. ∎

We can observe that in the case of distributed CSI, the number of DoFs is limited by the worst quality of the CSI across the channels to the RXs and across the TXs. Comparing this result with the number of DoFs achieved in a conventional MIMO BC given in Theorem 1, it is remarkable that the number of DoFs at *both* users is limited by the worst estimation error whether it is done relative to $\tilde{\boldsymbol{h}}_1$ or $\tilde{\boldsymbol{h}}_2$. This is contrast to the formula for the conventional MIMO BC in (13) where the accuracy of the estimation of $\tilde{\boldsymbol{h}}_i$ impacts only the number of DoFs of RX i.

Note that when all the CSI scaling coefficients are equal, the setting considered is still different from the conventional multiple-antenna BC. Indeed, the estimates at the different TXs have statistically the same accuracy since the CSI scaling coefficients are equal, but the realizations of the estimation errors are still different.

One can conclude from Theorem 2 that the additional interference due to the CSI inconsistency between the TXs does not lead to any loss in number of DoFs compared to the conventional multiple-antenna BC if and only if the channel estimates are of the same quality.

B. Robust Zero Forcing

Robust precoding schemes have been derived in the literature either as statistical robust ZF precoder or precoder optimizing the worst case performance to reduce the harmful effect of the imperfect CSI. Since we consider the average sum rate, the most relevant approach is the statistical one. Thus, we model the quantization error at TX j by an additive white Gaussian noise $\boldsymbol{\Delta}^{(j)} \triangleq [\boldsymbol{\delta}_1^{(j)}, \boldsymbol{\delta}_2^{(j)}]^{\mathrm{H}}$ of variance equal to $P^{-\alpha_i^{(j)}}$ for the estimation error $\boldsymbol{\delta}_i^{(j)}$ resulting from the quantization of $\tilde{\boldsymbol{h}}_i$ at TX j. The variance $P^{-\alpha_i^{(j)}}$ is obtained from the analysis of the scaling of the estimation error which is given in Appendix X-A.

The covariance matrix of the estimation error at TX j is then $\mathbf{R}_{\boldsymbol{\Delta}}^{(j)} \triangleq \mathrm{E}[\boldsymbol{\Delta}^{(j)}(\boldsymbol{\Delta}^{(j)})^{\mathrm{H}}] = \mathrm{diag}([P^{-\alpha_1^{(j)}}, P^{-\alpha_2^{(j)}}])$. Using this model, we can extend the approach from [17] and the beamformer transmitting symbol s_i at TX j is obtained from solving the following minimization:

$$\underset{\boldsymbol{t}_i}{\mathrm{argmin}}\, \mathrm{E}_{\boldsymbol{\Delta}^{(j)}}[\|\boldsymbol{e}_i - \tilde{\mathbf{H}}^{(j)}\boldsymbol{t}_i\|^2], \quad \text{subject to } \|\boldsymbol{t}_i\|^2 = \frac{P}{K}. \tag{14}$$

Writing the Lagrangian of the minimization problem with the Lagrange variable λ for the power constraint and taking the derivative according to $\boldsymbol{t}_i^*$ yields the equation

$$\left(\mathbf{R}_{\boldsymbol{\Delta}}^{(j)} + \mathbf{H}^{(j)\mathrm{H}}\mathbf{H}^{(j)} + \lambda\mathbf{I}\right)\boldsymbol{t}_i - \mathbf{H}^{(j)\mathrm{H}}\boldsymbol{e}_i = \mathbf{0}. \tag{15}$$

The factor λ improves the performance at intermediate SNR by striking a compromise between the orthogonality constraint and the power consumption but it cannot improve the number of DoFs. Thus, we can let λ be equal to zero and normalize the beamformer to fulfill the power constraint. The robust ZF beamformer transmitting symbol s_i is denoted by $\boldsymbol{t}_i^{\mathrm{rZF}} \triangleq [\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{t}_i^{\mathrm{rZF}(1)}, \boldsymbol{e}_2^{\mathrm{T}}\boldsymbol{t}_i^{\mathrm{rZF}(2)}]^{\mathrm{T}}$ and $\forall j \in \{1, 2\}$

$$\boldsymbol{t}_i^{\mathrm{rZF}(j)} \triangleq \sqrt{\frac{P}{K}} \frac{(\mathbf{R}_{\boldsymbol{\Delta}}^{(j)} + \mathbf{H}^{(j)\mathrm{H}}\mathbf{H}^{(j)})^{-1}\mathbf{H}^{(j)\mathrm{H}}\boldsymbol{e}_i}{\left\|(\mathbf{R}_{\boldsymbol{\Delta}}^{(j)} + \mathbf{H}^{(j)\mathrm{H}}\mathbf{H}^{(j)})^{-1}\mathbf{H}^{(j)\mathrm{H}}\boldsymbol{e}_i\right\|}. \tag{16}$$

We then derive the number of DoFs achieved by this robust precoder.

Proposition 1. *The robust ZF precoder defined in (16) achieves the same number of DoFs as conventional ZF.*

Proof: Considering strictly positive CSI scaling coefficients, the variances of the estimation errors tend to zero so that the inverse term in (16) can be approximated and we can write at RX $\bar{i}$:

$$|\tilde{\boldsymbol{h}}_{\bar{i}}^{\mathrm{H}}\boldsymbol{t}_i^{(j)}|^2 = \frac{P}{K}\frac{|\tilde{\boldsymbol{h}}_{\bar{i}}^{\mathrm{H}}(\mathbf{R}_{\boldsymbol{\Delta}}^{(j)} + \mathbf{H}^{(j)\mathrm{H}}\mathbf{H}^{(j)})^{-1}\mathbf{H}^{(j)\mathrm{H}}\boldsymbol{e}_i|^2}{\left\|(\mathbf{R}_{\boldsymbol{\Delta}}^{(j)} + \mathbf{H}^{(j)\mathrm{H}}\mathbf{H}^{(j)})^{-1}\mathbf{H}^{(j)\mathrm{H}}\boldsymbol{e}_i\right\|^2} \tag{17}$$

$$= \frac{P}{K}\frac{|\tilde{\boldsymbol{h}}_{\bar{i}}^{\mathrm{H}}(\mathbf{H}^{(j)})^{-1}((\mathbf{H}^{(j)\mathrm{H}})^{-1}\mathbf{R}_{\boldsymbol{\Delta}}^{(j)}(\mathbf{H}^{(j)})^{-1} + \mathbf{I})^{-1}\boldsymbol{e}_i|^2}{\left\|(\mathbf{R}_{\boldsymbol{\Delta}}^{(j)} + \mathbf{H}^{(j)\mathrm{H}}\mathbf{H}^{(j)})^{-1}\mathbf{H}^{(j)\mathrm{H}}\boldsymbol{e}_i\right\|^2} \tag{18}$$

$$= \frac{P}{K}\left(\frac{|\tilde{\boldsymbol{h}}_{\bar{i}}^{\mathrm{H}}(\mathbf{H}^{(j)})^{-1}(\mathbf{I} - (\mathbf{H}^{(j)\mathrm{H}})^{-1}\mathbf{R}_{\boldsymbol{\Delta}}^{(j)}(\mathbf{H}^{(j)})^{-1})\boldsymbol{e}_i|^2}{\left\|(\mathbf{R}_{\boldsymbol{\Delta}}^{(j)} + \mathbf{H}^{(j)\mathrm{H}}\mathbf{H}^{(j)})^{-1}\mathbf{H}^{(j)\mathrm{H}}\boldsymbol{e}_i\right\|^2} + o(\|\mathbf{R}_{\boldsymbol{\Delta}}^{(j)}\|_{\mathrm{F}}^2)\right) \tag{19}$$

The difference with conventional ZF is the term $(\mathbf{H}^{(j)\mathrm{H}})^{-1}\mathbf{R}_{\boldsymbol{\Delta}}^{(j)}(\mathbf{H}^{(j)})^{-1}$ which can be shown to lead to no reduction of the interference and introduces actually an additional error term. Yet, it converges to zero as $P^{-\min(\alpha_1^{(j)},\alpha_2^{(j)})}$ since $\mathbf{R}_{\boldsymbol{\Delta}}^{(j)} = \mathrm{diag}([P^{-\alpha_1^{(j)}}, P^{-\alpha_2^{(j)}}])$. This is also the rate at which the remaining interference tends to zero when using conventional ZF. Thus, the regularizing term vanishes and the number of DoFs achieved is the same as conventional ZF. ∎

Hence, even the existing designs of robust ZF precoders do not improve the number of DoFs in the DCSI-MIMO channel. Note that the extension of the definition of the statistical robust precoder as well as the extension of proposition 1 to the general setting with K users is trivial and will not be given explicitly.

C. Beacon Zero Forcing

Robust ZF schemes from the literature do not bring any DoFs improvement which leads to investigate other alternative schemes more adapted to the DCSI-MIMO channel. As a result, we now propose a modification of the conventional ZF scheme which improves the number of DoFs when the estimates for $\tilde{h}_1$ and $\tilde{h}_2$ are of different qualities. We call it *Beacon ZF* (bZF) because it makes use of an arbitrary channel-independent vector known beforehand at both TXs (a *beacon signal*).

The beamformer used to transmit symbol s_i is then $\boldsymbol{t}_i^{\mathrm{bZF}} \triangleq [\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{t}_i^{\mathrm{bZF}(1)}, \boldsymbol{e}_2^{\mathrm{T}}\boldsymbol{t}_i^{\mathrm{bZF}(2)}]^{\mathrm{T}}$, with its elements defined from

$$\boldsymbol{t}_i^{\mathrm{bZF}(j)} \triangleq \sqrt{\frac{P}{2}} \frac{\Pi_{\tilde{\boldsymbol{h}}_{\bar{i}}^{(j)}}^{\perp}(\boldsymbol{c}_i)}{\|\Pi_{\tilde{\boldsymbol{h}}_{\bar{i}}^{(j)}}^{\perp}(\boldsymbol{c}_i)\|} \tag{20}$$

where $\boldsymbol{c}_i$ is any non-zero vector chosen beforehand and known at the TXs. Due to the isotropy of the channel, the choice of $\boldsymbol{c}_i$ does not influence the performance of the precoder.

Corollary 1. *The number of DoFs achieved with beacon ZF is*

$$\mathrm{DoF}^{\mathrm{bZF}} = \min_{j\in\{1,2\}} \alpha_1^{(j)} + \min_{j\in\{1,2\}} \alpha_2^{(j)}. \tag{21}$$

Proof: The number of DoFs follows easily from Theorem 2. Indeed, when using beacon ZF, no error is induced by the projection of the direct channel which is replaced by a fixed given vector. In terms of number of DoFs, there is no difference between projecting the direct channel or any given vector. Thus, it is possible to apply the formula for the number of DoFs in Theorem 2 considering that the direct channel is perfectly known, which yields the result. ∎

The key idea behind beacon ZF is to reduce the impact of the differences in CSI quality by using only the CSI necessary to fulfill the orthogonality constraint. Thus, the direct channel, which does not change the number of DoFs but only improves the finite SNR performance, is not used. It follows then that $\boldsymbol{t}_1^{\mathrm{bZF}}$ does no depend on the estimates of $\tilde{\boldsymbol{h}}_1$, and symmetrically $\boldsymbol{t}_2^{\mathrm{bZF}}$ does not depend on the estimates of $\tilde{\boldsymbol{h}}_2$.

D. Active-Passive Zero Forcing

Beacon ZF improves the number of DoFs but it is still the worst CSI scaling across the TXs (although no longer across the RXs) which defines the number of DoFs. To improve further the number of DoFs, we propose a scheme called *Active-Passive Zero Forcing (AP ZF)*. Assuming w.l.o.g. that $\alpha_{\bar{i}}^{(2)} \geq \alpha_{\bar{i}}^{(1)}$, AP ZF consists in the precoder whose beamformer $\boldsymbol{t}_i^{\mathrm{APZF}}$ transmitting symbol s_i is given by

$$\boldsymbol{t}_i^{\mathrm{APZF}} \triangleq \sqrt{\frac{P}{2\log_2(P)}} \begin{bmatrix} 1 \\ -\frac{\{\tilde{\boldsymbol{h}}_{\bar{i}}^{(2)}\}_1}{\{\tilde{\boldsymbol{h}}_{\bar{i}}^{(2)}\}_2} \end{bmatrix} \tag{22}$$

$$= \sqrt{\frac{P(1+\rho_i^{(2)})}{2\log_2(P)}} \boldsymbol{u}_i^{\mathrm{APZF}} \tag{23}$$

where

$$\boldsymbol{u}_i^{\mathrm{APZF}} \triangleq \frac{\begin{bmatrix} 1 & -\frac{\{\tilde{\boldsymbol{h}}_{\bar{i}}^{(2)}\}_1}{\{\tilde{\boldsymbol{h}}_{\bar{i}}^{(2)}\}_2} \end{bmatrix}^{\mathrm{T}}}{\left\| \begin{bmatrix} 1 & -\frac{\{\tilde{\boldsymbol{h}}_{\bar{i}}^{(2)}\}_1}{\{\tilde{\boldsymbol{h}}_{\bar{i}}^{(2)}\}_2} \end{bmatrix}^{\mathrm{T}} \right\|} \tag{24}$$

and $\rho_i^{(2)} \triangleq |\{\tilde{\boldsymbol{h}}_{\bar{i}}^{(2)}\}_1|^2/|\{\tilde{\boldsymbol{h}}_{\bar{i}}^{(2)}\}_2|^2$.

AP ZF is based on the idea that each beamforming vector has to fulfill only one orthogonality constraint so that only one free variable is necessary. Thus, one coefficient can be set to a constant while still fulfilling the ZF constraints. Moreover, the only way to achieve the number of DoFs stemming from the best CSI estimate is if TX 2 (which has the best knowledge of $\tilde{h}_1$) can adapt to the coefficient transmitted at TX 1 to adjust its beamforming vector and improves the accuracy with which the interference are suppressed. This is possible only if TX 2 knows the transmit coefficient at TX 1.

Using this precoding scheme, the number of DoFs is then given in the following proposition.

Proposition 2. *Active-Passive ZF achieves the number of DoFs:*

$$\mathrm{DoF}^{\mathrm{APZF}} \geq \max_{j\in[1,2]} \alpha_1^{(j)} + \max_{j\in[1,2]} \alpha_2^{(j)}. \tag{25}$$

Proof: By symmetry, we consider w.l.o.g. the number of DoFs at RX 1, and we assume that the beamformers $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ are given by (23). We still assume w.l.o.g. that $\alpha_1^{(2)} \geq \alpha_1^{(1)}$, i.e., TX 2 has the best CSI over $\tilde{h}_1$. From (6), the number of DoFs at RX 1 is

$$\mathrm{DoF}_1 = 1 - \lim_{P\to\infty} \frac{\mathrm{E}_{\mathbf{H},\{\mathcal{W}_{i,j}\}}\left[\log_2(|\boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{t}_2|^2)\right]}{\log_2(P)} \tag{26}$$

We now focus on the interference term:

$$|\boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{t}_2|^2 = \frac{P}{2\log_2(P)} \left| \boldsymbol{h}_1^{\mathrm{H}} \begin{bmatrix} 1 \\ -\frac{\{\tilde{\boldsymbol{h}}_1^{(2)}\}_1}{\{\tilde{\boldsymbol{h}}_1^{(2)}\}_2} \end{bmatrix} \right|^2. \tag{27}$$

By construction, $\boldsymbol{t}_2$ is orthogonal to $\boldsymbol{h}_1^{(2)}$, so that

$$|\boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{t}_2|^2 = \frac{P(1+\rho_2^{(2)})}{2\log_2(P)} \|\boldsymbol{h}_1\|^2 \left| \Pi_{\tilde{\boldsymbol{h}}_1^{(2)}}^{\perp}(\tilde{\boldsymbol{h}}_1)^{\mathrm{H}}\boldsymbol{u}_2 \right|^2 \tag{28}$$

$$= \frac{P(1+\rho_2^{(2)})}{2\log_2(P)} \|\boldsymbol{h}_1\|^2 \sin^2(\angle(\tilde{\boldsymbol{h}}_1, \tilde{\boldsymbol{h}}_1^{(2)})). \tag{29}$$

Inserting (29) in the DoFs expression (26) and using Proposition 11 from Appendix X-A to bound the expectation of the sinus, we obtain

$$\text{DoF}_1 \geq \lim_{P\to\infty} \frac{E_{\mathbf{H},\{\mathcal{W}_{i,j}\}}\left[-\log_2\left(\sin^2(\tilde{\boldsymbol{h}}_1,\tilde{\boldsymbol{h}}_1^{(2)})\right)\right]}{\log_2(P)} \quad (30)$$

$$\geq \lim_{P\to\infty} \frac{B_1^{(2)}}{\log_2(P)} \quad (31)$$

$$= \alpha_1^{(2)} \quad (32)$$

which is the best scaling across the TXs. ∎

Comparing the number of DoFs achieved with AP ZF with the number of DoFs achieved when both TXs share the estimate of a channel vector with the highest accuracy gives the following result.

Theorem 3. *Active-Passive ZF achieves the same number of DoFs in the 2-user DCSI MIMO channel as in the conventional MIMO BC where both TXs share the estimates with the highest CSI accuracy.*

Improved scheme at finite SNR: AP ZF allows to recover the number of DoFs which would have been achieved with the best CSI across the TXs. However, the choice of the coefficient used to transmit at TX 1 (with the lowest accuracy of the CSI) remains to be discussed. In fact, the beamformer can be multiplied arbitrarily by any unit-norm complex number without impacting the rate achieved so that only the power used at TX 1 needs to be decided. According to (23), the power used at TX 1 is set to $P/(2\log_2(P))$.

The normalization by $\log_2(P)$ is done because the fading coefficient $\{\tilde{\boldsymbol{h}}_1\}_2$ might have a very small amplitude. In this case it would be necessary for TX 2 to transmit with a very large power to fulfill the orthogonality constraint. To ensure that the interference are canceled for all channel realizations while respecting the power constraint, it is necessary to have the ratio between the power used at TX 1 and the sum power constraint tending to zero. The factor $\log_2(P)$ is used because it fulfills this property while not reducing the number of DoFs due to the partial power consumption.

However, this comes at the cost of using only a small share of the available power, which is clearly inefficient and leads to a rate offset tending to minus infinity. To avoid this behavior, we propose that the TX with the worst CSI accuracy adapts its power consumption with respect to the channel realizations. In the following, we propose two possible solutions to improve the performance at finite SNR:

- Firstly, TX 1 can use its local CSI to normalize the beamformer which is then given by

$$\boldsymbol{t}_i^{\text{APZF}} = \sqrt{\frac{P}{2}} \begin{bmatrix} \frac{1}{\sqrt{1+\rho_i^{(1)}}} \\ -\frac{\{\tilde{\boldsymbol{h}}_{\bar{i}}^{(2)}\}_1}{\sqrt{1+\rho_i^{(2)}}\{\tilde{\boldsymbol{h}}_{\bar{i}}^{(2)}\}_2} \end{bmatrix} \quad (33)$$

with $\rho_i^{(j)} \triangleq |\{\tilde{\boldsymbol{h}}_{\bar{i}}^{(j)}\}_1|^2/|\{\tilde{\boldsymbol{h}}_{\bar{i}}^{(j)}\}_2|^2$, for $j = 1, 2$. This beamformer is not DoFs maximizing because the local CSI is used at TX 1 so that TX 2 does not any longer have an exact knowledge of the coefficient used to transmit

at TX 1. Consequently, beamformer $\boldsymbol{t}_i^{\text{APZF}}$ is not any longer orthogonal to $\tilde{\boldsymbol{h}}_{\bar{i}}^{(2)}$. Yet, this solution achieves good performance at intermediate SNR.

- Another possibility is to assume that TX 1 receives the scalar $\rho_i^{(2)}$ (or ρ_i) and use it to control its power. This means that TX 2 needs to share this scalar. This requires an additional feedback, but only a few bits are necessary to improve the performance at practical SNR.

V. ZERO FORCING IN THE DCSI-MIMO CHANNEL FOR ARBITRARY NUMBER OF USERS

In this section, we will show how the main results can be generalized to arbitrary number of users. The same approach as in the case $K = 2$ can be followed and we start by briefly generalizing to arbitrary number of users the precoding schemes previously described.

A. Conventional Zero Forcing

The conventional ZF precoder will be denoted as $\mathbf{T}^{\text{cZF}} \triangleq [\boldsymbol{t}_1^{\text{cZF}}, \dots, \boldsymbol{t}_K^{\text{cZF}}]$ with $\boldsymbol{t}_i^{\text{cZF}} \triangleq [\boldsymbol{e}_1^T \boldsymbol{t}_i^{\text{cZF}(1)}, \boldsymbol{e}_2^T \boldsymbol{t}_i^{\text{cZF}(2)}, \dots, \boldsymbol{e}_K^T \boldsymbol{t}_i^{\text{cZF}(K)}]^T$ transmitting symbol s_i, and the beamformer $\boldsymbol{t}_i^{\text{cZF}(j)}$ computed at TX j to transmit symbol i given by

$$\boldsymbol{t}_i^{\text{cZF}(j)} \triangleq \sqrt{\frac{P}{K}} \frac{\Pi_{\bar{\mathbf{H}}_i^{(j)}}^\perp(\tilde{\boldsymbol{h}}_i^{(j)})}{\|\Pi_{\bar{\mathbf{H}}_i^{(j)}}^\perp(\tilde{\boldsymbol{h}}_i^{(j)})\|} \quad (34)$$

with $\bar{\mathbf{H}}_i^{(j)} \triangleq [\tilde{\boldsymbol{h}}_1^{(j)}, \dots, \tilde{\boldsymbol{h}}_{i-1}^{(j)}, \tilde{\boldsymbol{h}}_{i+1}^{(j)}, \dots, \tilde{\boldsymbol{h}}_K^{(j)}]$.

We can then generalize the results from Theorem 2 to an arbitrary number of users.

Theorem 4. *In the DCSI-MIMO channel, the number of DoFs achieved with conventional ZF is equal to*

$$\text{DoF}^{\text{cZF}} = K \min_{i,j\in\{1,\dots,K\}} \alpha_i^{(j)}. \quad (35)$$

Proof: A detailed proof is provided in Appendix X-B. ∎

In Theorem 4, we have shown that the results concerning conventional ZF can be exactly generalized and the number of DoFs scales with the worst CSI accuracy across the TXs and the RXs. Indeed, the bad estimation of the channel to *one* user at *one* TX reduces the number of DoFs of *all* the users. This is very pessimistic and represents a different behavior as in the conventional multiple-antennas BC. This can be observed by comparing the number of DoFs for the conventional MIMO BC in (11) with the formula for the number of DoFs in the DCSI-MIMO channel given in (35) when $\forall i, j = 1, \dots, K, \alpha_i^{(j)} = \alpha_i$, i.e., the CSI qualities are the same at all the TXs

B. Beacon Zero Forcing

The beacon ZF precoder is denoted as $\mathbf{T}^{\text{bZF}} \triangleq [\boldsymbol{t}_1^{\text{bZF}}, \boldsymbol{t}_2^{\text{bZF}}, \dots, \boldsymbol{t}_K^{\text{bZF}}]$ with the beamformer $\boldsymbol{t}_i^{\text{bZF}} \triangleq [\boldsymbol{e}_1^T \boldsymbol{t}_i^{\text{bZF}(1)}, \boldsymbol{e}_2^T \boldsymbol{t}_i^{\text{bZF}(2)}, \dots, \boldsymbol{e}_K^T \boldsymbol{t}_i^{\text{bZF}(K)}]^T$ transmitting

symbol s_i. The beamformer $\boldsymbol{t}_i^{\mathrm{bZF}(j)}$ computed at TX j to transmit symbol s_i is given by

$$\boldsymbol{t}_i^{\mathrm{bZF}(j)} \triangleq \sqrt{\frac{P}{K}} \frac{\Pi_{\bar{\mathbf{H}}_i^{(j)}}^{\perp}(\boldsymbol{c}_i)}{\|\Pi_{\bar{\mathbf{H}}_i^{(j)}}^{\perp}(\boldsymbol{c}_i)\|} \tag{36}$$

where $\boldsymbol{c}_i$ is any non-zero vector chosen beforehand and known at all TXs.

Proposition 3. *The number of DoFs achieved with beacon ZF is equal to*

$$\mathrm{DoF}^{\mathrm{bZF}} = \sum_{k=1}^{K} \min_{\substack{i \in \{1,\ldots,K\}, \\ i \neq k}} \quad \min_{\substack{\ell,j \in \{1,\ldots,K\}, \\ \ell \neq i}} \alpha_\ell^{(j)}. \tag{37}$$

Proof: To derive the number of DoFs at a RX k, we need to compute the scaling of the interference at RX k stemming from the transmission to the $K-1$ other RXs. In the proof of Theorem 4, it is in fact the scaling of the interference resulting from the transmission of one stream which is calculated. To obtain the number of DoFs at one RX, the scaling of the interference resulting from the transmission of each of the $K-1$ interfering streams needs to be computed. This is represented by the first summation over i. Determining the interference leaked by the transmission of symbol s_i using beacon ZF leads to the second minimum in the formula. ∎

We have derived the number of DoFs for beacon ZF, but we will show in the following corollary that beacon ZF is only attractive in terms of number of DoFs in the two-user case.

Corollary 2. *For $K \geq 3$, beacon ZF achieves the same number of DoFs as conventional ZF.*

Proof: The result is easily obtained by studying the effect of the two successive minimums in (37). ∎

C. Active-Passive Zero Forcing

The generalization of AP ZF is intuitive and consists simply, for the computation of each beamforming vector, in letting one TX arbitrarily fix its precoding coefficient while the other TXs adapt to this coefficient. Nevertheless, it requires the introduction of a few more notations.

We define the ordered set $\mathbb{S} \triangleq \{n_1,\ldots,n_K\}$ as the set whose i-th element corresponds to the indice of the TX with fixed coefficient when transmitting the symbol s_i (passive TX for s_i). We then introduce the (column) channel vector from TX ℓ to all the RXs except the i-th RX:

$$\tilde{\boldsymbol{g}}_i^{(j)}(\ell) \triangleq [\{\tilde{\mathbf{H}}^{(j)}\}_{1,\ell},\ldots,\{\tilde{\mathbf{H}}^{(j)}\}_{i-1,\ell},\{\tilde{\mathbf{H}}^{(j)}\}_{i+1,\ell},\ldots,\{\tilde{\mathbf{H}}^{(j)}\}_{K,\ell}]^{\mathrm{T}}. \tag{38}$$

Using the previous definition, we can then define

$$\bar{\mathbf{H}}_i^{(j)}(n_i) \triangleq [\tilde{\boldsymbol{g}}_i^{(j)}(1),\ldots,\tilde{\boldsymbol{g}}_i^{(j)}(n_i-1),\tilde{\boldsymbol{g}}_i^{(j)}(n_i+1),\ldots,\tilde{\boldsymbol{g}}_i^{(j)}(K)] \tag{39}$$

which represents the estimate at TX j of the multi-user channel from all the TXs except TX n_i to all the RXs except RX i.

For a given set $\mathbb{S}$, we write $\mathbf{T}^{\mathrm{APZF}}(\mathbb{S}) \triangleq [\boldsymbol{t}_1^{\mathrm{APZF}}(n_1),\ldots,\boldsymbol{t}_K^{\mathrm{APZF}}(n_K)]$ where the beamformer $\boldsymbol{t}_i^{\mathrm{APZF}}(n_i) \triangleq [\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{t}_i^{\mathrm{APZF}(1)}(n_i),\ldots,\boldsymbol{e}_K^{\mathrm{T}}\boldsymbol{t}_i^{\mathrm{APZF}(K)}(n_i)]^{\mathrm{T}}$

transmits symbol s_i. The beamformer $\boldsymbol{t}_i^{\mathrm{APZF}(j)}(n_i)$ computed at TX j to transmit symbol s_i is given by

$$\boldsymbol{t}_i^{\mathrm{APZF}(j)}(n_i) \triangleq \sqrt{\frac{P}{K \log_2(P)}} \boldsymbol{u}_i^{\mathrm{APZF}(j)}(n_i) \tag{40}$$

where we have defined

$$\boldsymbol{u}_i^{\mathrm{APZF}(j)}(n_i) \triangleq [\, \check{u}_{1i}^{\mathrm{APZF}(j)}(n_i),\ldots,\check{u}_{n_i-1,i}^{\mathrm{APZF}(j)}(n_i), 1 \\ ,\check{u}_{n_i,i}^{\mathrm{APZF}(j)}(n_i),\ldots,\check{u}_{K-1,i}^{\mathrm{APZF}(j)}(n_i)]^{\mathrm{T}} \tag{41}$$

with $\check{\boldsymbol{u}}_i^{\mathrm{APZF}(j)}(n_i) \triangleq \left[\check{u}_{1i}^{\mathrm{APZF}(j)}(n_i),\ldots,\check{u}_{K-1,i}^{\mathrm{APZF}(j)}(n_i)\right]^{\mathrm{T}} \in \mathbb{C}^{K-1}$ and

$$\check{\boldsymbol{u}}_i^{\mathrm{APZF}(j)}(n_i) \triangleq \frac{-\left(\bar{\mathbf{H}}_i^{(j)}(n_i)\right)^{-1}\tilde{\boldsymbol{g}}_i^{(j)}(n_i)}{\sqrt{1 + \|\left(\bar{\mathbf{H}}_i^{(j)}(n_i)\right)^{-1}\tilde{\boldsymbol{g}}_i^{(j)}(n_i)\|^2}}. \tag{42}$$

Even though the notations are quite heavy, the intuition behind the construction of the precoder is exactly the same as for the two-user case. TX n_i is the *passive* TX and transmits with a fixed coefficient $\sqrt{P/K \log_2(P)}$ while the other *active* TXs then choose their coefficients in order to ZF the interference. This is obtained by setting their coefficients so as to fulfill (40). The notational complexity comes only from the fact that we need to introduce a "reduced" channel without the direct channel as well as without the channel from the *passive* TX.

Proposition 4. *Active-Passive ZF with the set $\mathbb{S} = \{n_1,\ldots,n_K\}$ achieves the number of DoFs*

$$\mathrm{DoF}^{\mathrm{APZF}}(\mathbb{S}) = \sum_{k=1}^{K} \min_{\substack{i \in \{1,\ldots,K\}, \\ i \neq k}} \quad \min_{\substack{\ell,j \in \{1,\ldots,K\}, \\ \ell \neq i, j \neq n_i}} \alpha_\ell^{(j)}. \tag{43}$$

Proof: Due to the symmetry between the RXs, we will show the result only for the number of DoFs at RX k. Let assume that AP ZF is used with the set $\mathbb{S}$. To obtain the number of DoFs, we need to derive the scaling of the interference at RX i when all streams are transmitted using AP ZF. The first minimum of the DoFs formula follows from the summation over all the $K-1$ interfering streams. It remains then to determine the scaling of the interference resulting from the transmission of one given data symbol.

TX j computes the beamformer $\boldsymbol{t}_\ell^{\mathrm{APZF}(j)}(n_\ell)$ according to (40). This formula is similar to the one for conventional ZF so that the scaling of the remaining interference power can be derived with a proof very akin to that of Theorem 4 which is omitted to avoid repetitions. Thus, the interference received at RX k due to the transmission of symbol s_i corresponds to the second minimum of the DoFs formula. This expression follows from the fact that the CSI at TX n_ℓ and the CSI on the direct channel $\tilde{\boldsymbol{h}}_\ell$ are not used to design the beamformer transmitting s_ℓ. ∎

The number of DoFs given in Proposition 4 is given by two successive minimizations. This is similar to beacon ZF at the difference that the index of one TX is not taken into account in the second minimization. This leads then to a larger number of DoFs. The formula for the number of DoFs depends on the

set $\mathbb{S}$ but we will show that the optimal set is easily derived when the number of users is larger than 4.

Corollary 3. *For $K \geq 4$ users, it is optimal in terms of number of DoFs to choose all the indices in $\mathbb{S}$ to be equal. Therefore, it is optimal to choose n_i as the indice of the minimum over all the CSI scaling coefficients, and the number of DoFs reads as*

$$\mathrm{DoF}^{\mathrm{APZF}} = K \min_{\substack{i,j \in \{1,\ldots,K\}, \\ j \neq \mathrm{argmin}_k \min_\ell \alpha_\ell^{(k)}}} \alpha_i^{(j)}. \qquad (44)$$

Proof: Similar to the proof of the corollary for Beacon ZF, the proof follows by studying the effect of the two successive minimums and for $K \geq 4$, it has for consequence that it is optimal to choose $\forall i, j, n_i = n_j$. ∎

Exactly as in the two-user case, AP ZF leads to an improvement in number of DoFs but this comes at the cost of an unbounded negative rate offset. To improve on this feature, the percentage of the available power which is consumed by the TXs needs to be increased. The sames solutions as described for the two-user case in Subsection IV-D can be applied, i.e., either a heuristic power control or the transmission of a scalar to control the power. Note that the scalar can be transmitted by any of the other $K - 1$ TXs and that one scalar needs to be transmitted for each stream. We refer to Subsection IV-D for more details.

D. Discussion of the Results

Altogether, we have shown in this section that the results for the two-user case given in Section IV could generalize to an arbitrary number of users. However, the results suggest in all cases a fundamental lack of robustness of the performance as we increase the number of users. Indeed, with conventional ZF, a single inaccurate channel estimate can reduce the number of DoFs of all the users while the novel precoding schemes proposed can only cope with a few channel estimates being of insufficient quality. This shows the need for other methods to make the transmission more robust to imperfect distributed CSI when more than two-user are present.

VI. PRECODING USING HIERARCHICAL QUANTIZATION

In view of the rather pessimistic results in the previous section, we propose now an alternative method to make the transmission more robust to the CSI discrepancies. It consists in modifying the CSI quantization and using a Hierarchical Quantization (HQ) scheme to encode the CSI [34], [41].

A. Hierarchical Quantization

Hierarchical quantization (or multi-resolution quantization) is a quantization scheme in which the information is encoded so that the original message can be decoded up to a number of bits depending on the quality of the feedback channel. The better the channel is, the more bits can be decoded. Thus, if one entity receives a codeword with a higher accuracy than another entity, and has the knowledge of the feedback qualities,

it also knows what has been decoded at the other entity. Conversely, if one entity can detect the feedback information at a given resolution level but knows that another entity can decode the same information at a higher resolution level, it can use its individual decoded codeword to form a limited set of guesses around it as to which higher resolution codeword may have been detected at the other TX.

In our setting, it means that each TX can decode the CSI feedback up to a certain number of bits depending on the quality of the feedback link. If TX j_1 receives a CSI of better quality than another TX j_2, it can decode more bits from the CSI and can get the knowledge of the CSI at TX j_2 with less decoded bits. Note that this implies that two TXs with the same CSI quality have the *same* codebook and thus exactly the same realization for the channel estimation error. This is in contrast to what has been considered in the previous sections.

We wish to continue using the properties of RVQ so that we need to design *hierarchical random codebooks*, i.e., codebooks fulfilling the properties of both kinds of codebooks. Since this is not the main focus of the work, we just briefly describe a possible method to construct such codebooks and the quantization scheme associated.

We start by considering a random codebook of size corresponding to the best accuracy, say $2^{\ell_{\max}}$. This random codebook is then divided into two random codebooks containing each half the elements. This process is then applied on the two smaller codebooks obtained until having $2^{\ell_{\max}}$ codebooks of one element. In each of the sub-codebooks of different sizes created, we pick randomly one elements to be the *representative* of this codebook.

Once the quantized vector maximizing the figure of merit has been chosen among the $2^{\ell_{\max}}$ vectors, the encoding can be easily done. The chosen vector belongs to one set of each size and the encoding bits are used to select among the two possible choices, the set to which the quantized vector belongs.

The decoding step works as follows. The first bit denotes one of the two codebooks of size $2^{\ell_{\max}-1}$, the second bit denotes one of the two codebooks of size $2^{\ell_{\max}-2}$ inside this codebook, and so on, until the last bit is decoded. Once this is done, the codeword decoded is chosen to be the *representative* codeword of the obtained codebook.

It is then easily verified that the proposed quantization scheme has the hierarchical properties desired.

B. Conventional Zero Forcing with Hierarchical Quantization

In the previous sections, we have shown that the quality of the estimation of one channel $\tilde{h}_i$ to one given RX had an impact on the number of DoFs achieved *at all RXs*. This is a surprising property which follows from the particular structure of the DCSI-MIMO channel where the consistency between the transmissions of the different TXs is critical. We will show how the hierarchical quantization described above can be used to avoid this very inefficient property.

In the following, we will consider a particularly simple use of hierarchical quantization consisting in letting all the TXs designing the beamforming vector use only the part of the CSI which is common to all the TXs, and simply "forget"

about the more accurate CSI knowledge. We then obtain a CSI configuration where all the TXs share the same CSI and the number of DoFs can be obtained from Theorem 1.

Theorem 5. *The number of DoFs achieved using Conventional ZF with hierarchical quantization is*

$$\text{DoF}^{\text{cZF}} = \sum_{i=1}^{K} \min_{j \in \{1,\ldots,K\}} \alpha_i^{(j)}. \tag{45}$$

Using HQ as described, i.e., using only the estimate of a channel vector $\tilde{\boldsymbol{h}}_i$ common to all the TXs, follows from the observation that the worst estimation error of $\tilde{\boldsymbol{h}}_i$ limits in any case the number of DoFs at RX i. Thus, using only the common part of the estimate of $\tilde{\boldsymbol{h}}_i$ does not reduce the number of DoFs at RX $\tilde{\boldsymbol{h}}_i$. Yet, it leads to an improved consistency between the beamformers computed at the TXs. This has for consequence that the error in the estimate of the channel $\tilde{\boldsymbol{h}}_i$ only impacts the number of DoFs at RX i and not at the other RXs.

Note that the proposed scheme using HQ is very simple and more gains could certainly be obtained with a more sophisticated use of the additional CSI knowledge available at some TXs.

C. Active-Passive Zero Forcing with Hierarchical Quantization

Hierarchical quantization is used for AP ZF in the same way as for Conventional ZF. This consists in using the CSI which is common to all the active TXs considered in the definition of the beamformer in (40).

Proposition 5. *The number of DoFs achieved using Active-Passive ZF with Hierarchical Quantization and the set $\mathbb{S}$ is*

$$\text{DoF}^{\text{APZF}}(\mathbb{S}) = \sum_{k=1}^{K} \min_{\substack{i \in \{1,\ldots,K\}, \\ i \neq k}} \min_{\substack{j \in \{1,\ldots,K\}, \\ j \neq n_i}} \alpha_k^{(j)}. \tag{46}$$

The two successive minimums come from the fact that it is not the same TX which is *passive* for the different streams. It is clear from (46) that it is optimal to choose all the n_i to be equal for $K \geq 3$. However, the indice of the optimal passive TX, which we denote by n_{HQ}, is now different from the case without HQ. It is easily obtained by looking for the passive TX bringing the largest improvement in number of DoFs:

$$n_{\text{HQ}} \triangleq \underset{n \in \{1,\ldots,K\}}{\arg\max} \sum_{k=1}^{K} \min_{\substack{j \in \{1,\ldots,K\}, \\ j \neq n}} \alpha_k^{(j)}. \tag{47}$$

The maximum number of DoFs using AP-ZF with HQ follows then directly.

Proposition 6. *For $K \geq 3$, it is optimal to choose the passive TX to be TX j with $j = n_{\text{HQ}}$ defined in (47), for all the data streams. The number of DoFs achieved with Active-Passive ZF based on Hierarchical Quantization is then equal to*

$$\text{DoF}^{\text{APZF}} = \sum_{i=1}^{K} \min_{\substack{j \in \{1,\ldots,K\}, \\ j \neq n_{\text{HC}}}} \alpha_i^{(j)}. \tag{48}$$

VII. DoF Optimal Sharing of the Feedback under a Total Feedback Constraint

In this section, we consider the opposite side of the problem which consists in deriving how to distribute a maximum number B of feedback bits across the TXs and the channel vectors so as to maximize the number of DoFs. Since our focus remains on the number of DoFs and considering previous results, it is meaningful to introduce $\gamma \triangleq \lim_{P \to \infty} B / \log_2(P)$ which we call the *total feedback scaling*.

Thus, we consider a constraint on the sum of the scaling coefficients of the total feedback transmitted through the multi-user channel feedback:

$$\sum_{i,j \in \{1,\ldots,K\}} \alpha_i^{(j)} \leq \gamma. \tag{49}$$

We study first conventional ZF before extending the results to Active-Passive ZF. To optimize the CSI allocation efficiently, it becomes necessary to also optimize the number of users being served, which means that time sharing will this time be explicitly considered.

A. Conventional Zero Forcing

Proposition 7. *With conventional ZF (with or without Hierarchical Quantization), it is optimal in terms of number of DoFs to share equally the number of bits across the TXs and across the channels to quantize and to let the number of TX being actually transmitting be equal to n for $\gamma \in [n(n-1)^2, (n+1)n^2]$. It follows that the optimal number of DoFs using Conventional ZF is equal to*

$$\begin{cases} \text{DoF}^{\text{cZF}} = \gamma/(n(n-1)), & \text{if } \gamma \in [n(n-1)^2, n^2(n-1)] \\ \text{DoF}^{\text{cZF}} = n, & \text{if } \gamma \in [n^2(n-1), (n+1)n^2]. \end{cases} \tag{50}$$

Proof: We study first the case without HQ. Since the number of DoFs scales as the worst CSI scaling across the TXs and the channel vectors, it is clearly optimal to have the same CSI accuracy at all the TXs and for all the channel vectors. To achieve a number of DoFs of α at n RXs, the number of bits to quantize a channel vector has to be equal to $\alpha(n-1)\log_2(P)$, where n is the number of transmitting TXs. Hence, the total feedback in the channel is given by $n^2\alpha(n-1)\log_2(P)$ when considering the n estimates needed at the n TXs.

Let us assume that n TXs are serving n RXs with the maximal feedback scaling γ, we obtain that $\alpha = \gamma/(n^2(n-1))$. For $\gamma \leq n^2(n-1)$ the number of DoFs achieved at the RXs is lower or equal to one so that the sum number of DoFs is equal to $n\alpha = \gamma/(n(n-1))$. For $\gamma \geq n^2(n-1)$, the number of DoFs at each RX reaches its maximal value of one and the sum number of DoFs is equal to n. Comparing the sum number of DoFs achieved by two successive configurations, with respectively n and $n+1$ users served, leads to the value of γ given in the proposition as switching point between the configurations.
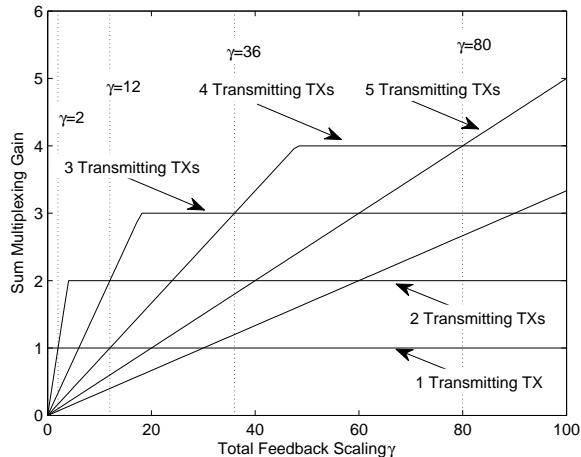
Fig. 2. Degrees of Freedom as a function of the total feedback scaling γ for different number of users.

When HQ is used, the number of DoFs still scales as the minimum over the CSI scaling across the TXs so that it is still optimal to let all the TXs have the same CSI scaling. ∎

Using HQ does not increase the number of DoFs when the CSI configuration can be optimized. However, many more configurations are optimal as the CSI can be allocated indifferently to any channel vector as long as the scaling of the CSI does not exceed one and all the TXs receive the same CSI.

The results from Proposition 7 are very intuitive, yet the formula is not very enlightening and the intuition is better understood in a plot of the number of DoFs with optimal CSI sharing. Thus, we plot in Figure 2 the number of DoFs in terms of the total feedback scaling γ for different numbers of transmitting TXs. The parts with a positive slope correspond to values of α smaller than one while the flat parts correspond to a saturation of the number of DoFs, i.e., $\alpha \geq 1$.

The values of γ corresponding to the saturation of the number of DoFs and to the activation of an additional user, respectively, are given in Appendix X-C. When n TXs are transmitting, the slope of the number of DoFs as a function of γ is known to be equal to $1/(n^2(n-1))$ and we can observe in the figure how the values for γ given in the proposition fit with the observation in terms of saturation and intersection of the curves.

It is possible to observe that the saturated parts are optimal for some values of γ. This follows from the fact that using an additional TX induces an increase of the feedback necessary (lower slope in the figure). Thus, a possibly large increase in γ is necessary before reaching the point where it starts being more interesting to serve the additional RX and use an additional TX.

B. Extension to Active-Passive Zero Forcing

Our analysis for conventional ZF can be extended to Active-Passive ZF without difficulty. The only difference consists in the number of bits necessary to achieve a scaling of α which is then $n(n-1)^2 \alpha \log_2(P)$ instead of $n^2(n-1)\alpha \log_2(P)$ since one TX (passive TX) does not need to be shared any

CSI. Thus, it holds that $\alpha = \gamma/(n(n-1)^2)$ which leads to the following result.

Proposition 8. *When using Active-Passive ZF (with or without HQ), it is optimal to share equally the number of bits across the active TXs and across the channels, and to let the number of transmitting TXs be equal to n for $\gamma \in [n(n-1)^2, (n+1)n^2]$. It follows that the optimal number of DoFs is equal to*

$$
\begin{cases}
\mathrm{DoF}^{\mathrm{APZF}} = \gamma/(n-1)^2, & \text{if } \gamma \in [(n-1)^3, n(n-1)^2] \\
\mathrm{DoF}^{\mathrm{APZF}} = n, & \text{if } \gamma \in [n(n-1)^2, n^3].
\end{cases}
\tag{51}
$$

The proof and the plot of the number of DoFs in terms of the total feedback scaling γ follow both the same pattern as conventional ZF and are omitted to avoid repetition.

The general insight behind those results is that it is better to achieve the maximal number of DoFs at less users instead of serving more users with a lower number of DoFs. This is an intuitive consequence of the very quick increase of the size of the aggregate feedback required in terms of the number of TXs used.

VIII. SIMULATIONS

A. In the Two-User Case

We consider two models for the imperfect channel CSI, a statistical model and RVQ.

In the statistical model, the quantization error is modeled by adding a Gaussian i.i.d. quantization noise to the channel with the covariance matrix at TX j equal to $\mathrm{diag}([P^{-\alpha_1^{(j)}}, P^{-\alpha_2^{(j)}}])$. This corresponds to the scaling in P of the variance provided in Proposition 10 of Appendix X-A. The Gaussian distribution maximizes the entropy for the given variance [42] so that we will obtain a priori a lower bound for the performance. Yet, it is expected that only the scaling of the variance will have an impact so that the statistical model should be accurate. The averaging is then done over 10000 realizations.

In the RVQ, we consider a given number of feedback bits and we average over 100 random codebooks and 1000 channel realizations. In the simulations, we consider the following precoders: ZF with perfect CSI, conventional ZF [cf. (12)], Beacon ZF [cf. (20)], and Active-Passive ZF [cf. (23)] with heuristic power control and with 3-bits power control.

In Fig. 3, we consider the statistical model with the CSI scaling $[\alpha_1^{(1)}, \alpha_1^{(2)}] = [1, 0.5]$ and $[\alpha_2^{(1)}, \alpha_2^{(2)}] = [0, 0.7]$. To emphasize the number of DoFs (i.e., the slope of the curve in the figure), we let the SNR grow large. As expected theoretically, conventional ZF scales with the worst accuracy and saturates at high SNR, while Beacon ZF has a positive slope and Active-Passive ZF performs closer to perfect ZF with a slope only slightly smaller than the optimal one.

In Fig. 4, we plot the sum rate achieved with the CSI feedback $[B_1^{(1)}, B_1^{(2)}] = [6, 3]$ and $[B_2^{(1)}, B_2^{(2)}] = [3, 6]$ using RVQ. From the theoretical analysis, the number of DoFs should be equal to zero for all the precoding schemes since the number of feedback bits used does not increase with the SNR. This is confirmed by the saturation of the sum rate
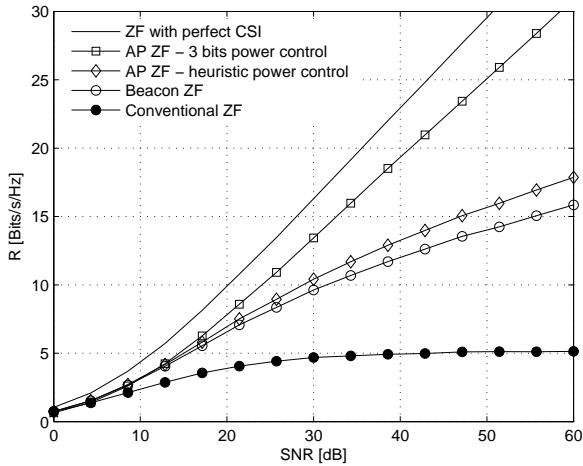
Fig. 3. Sum rate in terms of the SNR with a statistical modeling of the error from RVQ using $[\alpha_1^{(1)}, \alpha_1^{(2)}] = [1, 0.5]$ and $[\alpha_2^{(1)}, \alpha_2^{(2)}] = [0, 0.7]$.

Fig. 5. Sum rate achieved for the arbitrarily chosen CSI scaling configuration α given in Appendix X-D.
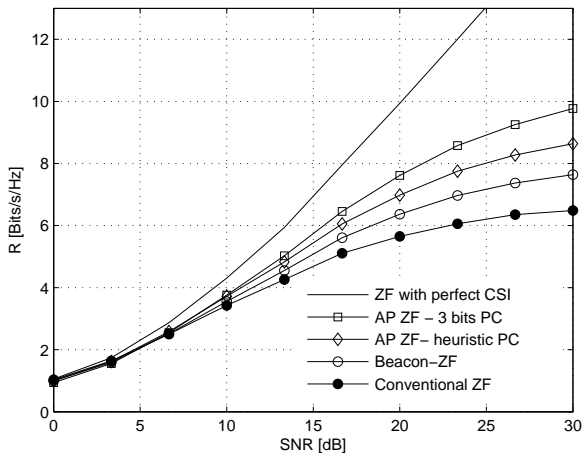

Fig. 4. Sum rate in terms of the SNR with RVQ using $[B_1^{(1)}, B_1^{(2)}] = [6, 3]$ and $[B_2^{(1)}, B_2^{(2)}] = [3, 6]$.

as the SNR increases. Yet, the saturation occurs at a higher SNR for Beacon ZF compared to conventional ZF, and at an even higher SNR for Active-Passive ZF. This translates into an improvement of the sum rate at intermediate SNR.

B. With Arbitrary Number of Users

For the simulations with arbitrary number of users, only the statistical model described in the previous paragraph for the two-user case is considered. To model easily the use of Hierarchical Quantization, we simply consider that a TX has the knowledge of the channel estimate at another TX if this TX receives a feedback concerning this channel vector with a lower CSI scaling coefficient. Since we have derived that Beacon ZF [Cf. (36)] does not bring any improvement in number of DoFs for $K \geq 3$, we will consider in the figures only conventional ZF [Cf. (34)] and Active-Passive ZF [Cf. (40)] where the transmission of 3-bits to the *passive* TX is allowed for every beamforming vector. For both precoding schemes,
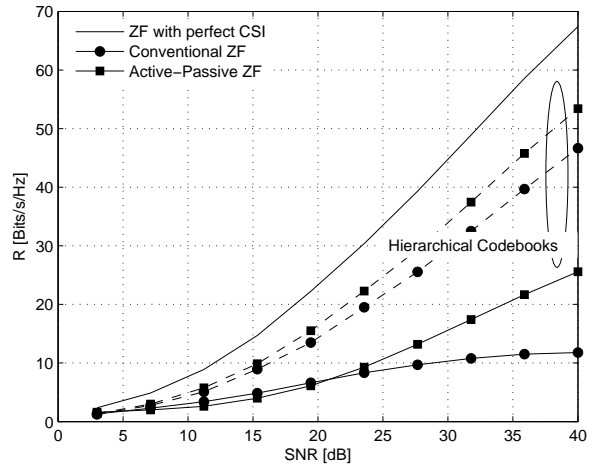
we will furthermore consider both the case of Hierarchical Quantization with random codebooks and conventional RVQ.

We consider the performance achieved with an arbitrary chosen CSI scaling matrix to verify that the precoding schemes behave as expected. Thus, we consider $K = 7$ users and we set all the elements of the CSI scaling matrix α equal to 1 at the exception of two coefficients corresponding to different TXs and RXs set to 0 and 0.3, respectively. The CSI scaling matrix is given explicitly in Appendix X-D as well as the number of DoFs obtained analytically for that setting.

In Fig. 5, we plot the average sum rate achieved for the previous setting in terms of the SNR. We can observe that the schemes using HQ achieve a much larger number of DoFs (i.e., slope in terms of the SNR) which is in agreement with the theoretical results. Furthermore, the increase in number of DoFs translates to better performance at intermediate SNRs.

IX. CONCLUSION

In this work, we have introduced a new model, called distributed CSI-MIMO channel, consisting in a multicell downlink channel where each transmitter has its own local estimate of the whole multi-user channel. We have shown that conventional ZF precoding applied without taking into account the CSI discrepancies achieves far from the maximal number of DoFs and is limited by the worst accuracy of the CSI over the whole multi-user channel. This is particularly striking as the bad estimate of the channel to one particular user at a unique TX reduces the number of DoFs of all the users. This represents a different behavior from the conventional MIMO BC. In the particular case with only two users, we have provided a precoding scheme achieving the number of DoFs corresponding to the most accurate CSI across the TXs. With arbitrary number of users, the number of DoFs achieved by conventional ZF has been derived and precoding schemes to improve over this number of DoFs value have been provided. Particularly, it has been shown how using codebooks with a hierarchical structure to quantize the CSI could lead to a significant number of DoFs improvement. Moreover,

considering the opposite problem of optimizing the sharing of the CSI feedback under a total feedback constraint, we have derived a number of DoFs maximizing CSI configuration when ZF is used. Finally, simulations have confirmed that the novel precoding schemes outperform known linear precoding schemes at intermediate SNRs.

This paper represents the first step on our work on the DCSI-MIMO channel and many problems remain open. Firstly, the DCSI-MIMO channel has been studied asymptotically for analytical tractability and the extension to finite SNR represents a challenging problem. The design of other robust precoders forms also an interesting problem with a strong potential. Finally, there are many other scenarios where distributed TXs want to cooperate but cannot practically share the exact same CSI (Relay channels, interference channels,...). In such settings, similar analysis could be developed to make the transmission more robust to the CSI discrepancies which are likely to exist in practical settings.

X. APPENDIX

A. Some Results on Vector Quantization

We consider the quantization of the unit-norm complex vector $\tilde{\boldsymbol{h}} \in \mathbb{C}^K$ over a codebook $\mathcal{C}$ where both the channel to quantize and the elements of the codebook are multiplied by a unit-norm complex number (i.e., are rotated in the complex space) so as to let the first element of the vector be real valued. The quantized vector $\hat{\boldsymbol{h}}$ is then obtained as

$$\hat{\boldsymbol{h}} = \underset{\boldsymbol{c} \in \mathcal{C}}{\operatorname{argmin}} \|\boldsymbol{c} - \tilde{\boldsymbol{h}}\|. \tag{52}$$

The multiplication by the unit-norm complex number is done in order to optimize the performance of the quantization. Since the norm is conserved when considering the canonical isomorphism from $\mathbb{C}^K$ to $\mathbb{R}^{2K}$, we can consider for the quantization the vectors as elements of $\mathbb{R}^{2K}$ made of the stacked real and imaginary parts of the original vector.

With the first coefficient real valued, it is only necessary to consider $\mathbb{R}^{2K-1}$. Thus, a vector $\boldsymbol{u} = [u_1, u_2, \ldots, u_K]^{\mathrm{T}} \in \mathbb{C}^K$ with its first coefficient real valued is represented in $\mathbb{R}^{2K-1}$ by $\boldsymbol{u}_{\mathbb{R}^{2K-1}}$ defined as

$$\boldsymbol{u}_{\mathbb{R}^{2K-1}} \triangleq \begin{bmatrix} \operatorname{Re}(u_1) & \operatorname{Re}(u_2) & \operatorname{Im}(u_2) & \operatorname{Re}(u_3) & \ldots & \operatorname{Im}(u_K) \end{bmatrix}^{\mathrm{T}}. \tag{53}$$

We can then define the angle between $\boldsymbol{u}_{\mathbb{R}^{2K-1}}$ and $\boldsymbol{v}_{\mathbb{R}^{2K-1}}$ in $\mathbb{R}^{2K-1}$ as

$$\angle(\boldsymbol{u}_{\mathbb{R}^{2K-1}}, \boldsymbol{v}_{\mathbb{R}^{2K-1}}) \triangleq \arccos\left(\frac{|\boldsymbol{u}_{\mathbb{R}^{2K-1}}^{\mathrm{T}} \boldsymbol{v}_{\mathbb{R}^{2K-1}}|}{\|\boldsymbol{u}_{\mathbb{R}^{2K-1}}\| \|\boldsymbol{v}_{\mathbb{R}^{2K-1}}\|}\right). \tag{54}$$

Using the conservation of the norm by the canonical isomorphism, the quantization in (52) is rewritten as

$$\hat{\boldsymbol{h}}_{\mathbb{R}^{2K-1}} = \underset{\boldsymbol{c}_{\mathbb{R}^{2K-1}} \in \mathcal{C}_{\mathbb{R}^{2K-1}}}{\operatorname{argmin}} \|\boldsymbol{c}_{\mathbb{R}^{2K-1}} - \tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}\|^2 \tag{55}$$

$$= \underset{\boldsymbol{c}_{\mathbb{R}^{2K-1}} \in \mathcal{C}_{\mathbb{R}^{2K-1}}}{\operatorname{argmin}} (2 - 2\boldsymbol{c}_{\mathbb{R}^{2K-1}}^{\mathrm{T}} \tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}). \tag{56}$$

We can see from (56) that the quantization scheme aims at maximizing $\boldsymbol{c}_{\mathbb{R}^{2K-1}}^{\mathrm{T}} \tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}$. This figure of merit can be linked

to the commonly used chordal distance $\mathrm{d}(\bullet)$ which is defined for two vectors as [43]

$$\mathrm{d}(\tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}, \boldsymbol{c}_{\mathbb{R}^{2K-1}}) = \sqrt{\sin^2(\angle(\boldsymbol{c}_{\mathbb{R}^{2K-1}}, \tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}))} \tag{57}$$

$$= \sqrt{1 - |\boldsymbol{c}_{\mathbb{R}^{2K-1}}^{\mathrm{T}} \tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}|^2}. \tag{58}$$

Thus, minimizing the chordal distance is equivalent to maximizing $|\boldsymbol{c}_{\mathbb{R}^{2K-1}}^{\mathrm{T}} \tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}|^2$. This is then equivalent to the quantization scheme (54) if the half-space where $\tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}$ belongs is known. This requires solely one additional bit. Since we are interested in the scaling of the number of bits, this will not make any difference. Consequently, we will study in the following the quantization scheme based on the minimization of the chordal distance

$$\hat{\boldsymbol{h}}_{\mathbb{R}^{2K-1}} = \underset{\boldsymbol{c}_{\mathbb{R}^{2K-1}} \in \mathcal{C}_{\mathbb{R}^{2K-1}}}{\operatorname{argmin}} \sqrt{\sin^2(\angle(\boldsymbol{c}_{\mathbb{R}^{2K-1}}, \tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}))}. \tag{59}$$

On that account, we now study the quantization scheme given by (59) over the Grassmannian manifold of dimensions $(1, 2K-1)$ in the field $\mathbb{R}$ (i.e., on the unitary ball in $\mathbb{R}^{2K-1}$). This quantization scheme is studied (in a much more general form) in [43] and we start by recalling some results adapted to our notations. We then derive some new properties which will be needed in the derivations. [3]

Let us first denote by $\mathrm{F}(x) \triangleq \Pr\{\sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{c})) \leq x\}$ the cumulative distribution function (CDF) of $\mathrm{d}^2(\tilde{\boldsymbol{h}}, \boldsymbol{c}) = \sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{c}))$ where $\boldsymbol{c} \in \mathbb{R}^{2K-1}$ is an element of a random codebook.

Proposition 9 ([43], Corollary 2). *The CDF* $\mathrm{F}(x)$ *verifies that for all* $x \leq 1$

$$c_{2K-1} x^{K-1} \leq \mathrm{F}(x) \leq c_{2K-1} x^{K-1} (1-x)^{\frac{-1}{2}} \tag{60}$$

where $c_{2K-1} \triangleq \Gamma(K - 1/2)/(\Gamma(K)\Gamma(1/2))$.

Proposition 10 ([43], Theorem 2). *When the size* $L = 2^B$ *of the random codebook is sufficiently large* $(c_{2K-1}^{-1/(K-1)} 2^{-B/(K-1)} \leq 1$ *is necessary), then it holds that*

$$\frac{2K-1}{2K+1} c_{2K-1}^{-1/(K-1)} 2^{-B/(K-1)} \lesssim$$

$$\mathrm{E}_{\mathcal{C}, \tilde{\boldsymbol{h}}}[\min_{\boldsymbol{c} \in \mathcal{C}} \sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{c}))] \lesssim \frac{\Gamma(\frac{1}{K-1})}{K-1} c_{2K-1}^{-1/(K-1)} 2^{-B/(K-1)}. \tag{61}$$

Proposition 11. *When the size* $L = 2^B$ *of the random codebook is sufficiently large, the expectation of the logarithm of the quantization error is bounded as*

$$\frac{B + \log_2(c_{2K-1})}{(K-1)} \lesssim$$

$$\mathrm{E}_{\mathcal{C}, \tilde{\boldsymbol{h}}}\left[-\log_2\left(\min_{\boldsymbol{c} \in \mathcal{C}} \sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{c}))\right)\right] \lesssim \frac{B + \log_2(c_{2K-1}) + \log_2(e)}{(K-1)}. \tag{62}$$

---

[3]We will do the abuse of notation consisting in removing the subscript $\mathbb{R}^{2K-1}$ in the derivations but it will be clear that any mention of an angle will refer to the angle defined in $\mathbb{R}^{2K-1}$.

Proof: **Upper Bound:** The derivation of an upper bound follows the same idea as the proof in Appendix B of [43] which derives an upper bound for the same expectation as in this proof, only without the logarithm. We start by recalling a Lemma from [43] which follows easily from the definition but is helpful.

Lemma 1 ([43], Lemma 3). *The empirical distribution function minimizing the distorsion for a given $L = 2^B$ is*

$$
\mathrm{F}_{\mathcal{C}^*}^*(x) = \begin{cases} 0 & \text{if } x < 0 \\ L\mathrm{F}(x) & \text{if } 0 \le x \le x^* \\ 1 & \text{if } x > x^* \end{cases} \tag{63}
$$

where x^* satisfies $L\mathrm{F}(x^*) = 1$ and $\mathrm{F}(x) \triangleq \Pr\{\sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{c}))| \le x\}$.

Note that Lemma 1 corresponds to the optimal codebook minimizing the average distance and leads thusly to a lower bound for the distorsion. We can then write

$$
\mathrm{E}_{\mathcal{C},\tilde{\boldsymbol{h}}}\left[-\log\left(\min_{\boldsymbol{c} \in \mathcal{C}} \sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{c}))\right)\right]
$$
$$
= \int_0^\infty \Pr\{-\log\left(\min_{\boldsymbol{c} \in \mathcal{C}} \sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{c}))\right) \ge z\}\mathrm{d}z \tag{64}
$$
$$
= \int_0^\infty \Pr\{\min_{\boldsymbol{c} \in \mathcal{C}} \sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{c})) \le e^{-z}\}\mathrm{d}z \tag{65}
$$
$$
\le \int_0^{-\log(x^*)} \mathrm{d}z + \int_{-\log(x^*)}^{-\infty} L\Pr\{\sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{c})) \le e^{-z}\}\mathrm{d}z \tag{66}
$$

where (64) is obtained by exploiting the fact that the term in the expectation is a positive random variable and (66) follows from the previous lemma since the CDF obtained with the optimal codebook dominates the CDF obtained with any other codebook of the same size.

Following the same approach as the proof in Appendix B of [43], we define $\mathrm{F}_0(x) \triangleq c_{2K-1}x^{K-1}$ and x_0 so that $L\mathrm{F}_0(x_0) = 1$. Let also define $\mathrm{F}_{\mathrm{ub}}(x) \triangleq c_{2K-1}x^{K-1}(1-x)^{-1/2}$ and x_{ub} so that $L\mathrm{F}_{\mathrm{ub}}(x_{\mathrm{ub}}) = 1$. Finally, we define $\mathrm{F}_{\mathrm{ubub}}(x) \triangleq c_{2K-1}x^{K-1}(1-x_0)^{-1/2}$ and x_{ubub} so that $L\mathrm{F}_{\mathrm{ubub}}(x_{\mathrm{ubub}}) = 1$.

It holds by construction that $x_{\mathrm{ub}} \le x^* \le x_0$ since we know from Proposition 9 that $\mathrm{F}_0(x) \le \mathrm{F}(x) \le \mathrm{F}_{\mathrm{ub}}(x)$. Clearly $(1-x)^{-1/2} \le (1-x_0)^{-1/2}$ for $x \in [0, x_0]$ so that $\mathrm{F}_{\mathrm{ub}}(x) \le \mathrm{F}_{\mathrm{ubub}}(x)$ for $x \in [0, x_0]$, which finally implies $x_{\mathrm{ubub}} \le x_{\mathrm{ub}}$. We can then use these relations to derive an upper bound for (66).

$$
\mathrm{E}_{\mathcal{C},\tilde{\boldsymbol{h}}}\left[-\log\left(\min_{\boldsymbol{c} \in \mathcal{C}} \sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{c}))\right)\right]
$$
$$
\le \int_0^{-\log(x^*)} \mathrm{d}z + \int_{-\log(x^*)}^{-\infty} L\mathrm{F}(e^{-z})\mathrm{d}z \tag{67}
$$
$$
\le \int_0^{-\log(x_{\mathrm{ubub}})} \mathrm{d}z + \int_{-\log(x_0)}^{\infty} L\mathrm{F}(e^{-z})\mathrm{d}z \tag{68}
$$
$$
\le \int_0^{-\log(x_{\mathrm{ubub}})} \mathrm{d}z + \int_{-\log(x_0)}^{\infty} L\mathrm{F}_{\mathrm{ubub}}(e^{-z})\mathrm{d}z. \tag{69}
$$

Equation (68) follows from $x_{\mathrm{ubub}} \le x^* \le x_0$ and (69) follows from the fact that $\mathrm{F}_{\mathrm{ub}}(x) \le \mathrm{F}_{\mathrm{ubub}}(x)$ for $x \in [0, x_0]$. We

now replace $\mathrm{F}_{\mathrm{ubub}}(x)$, x_{ubub}, and x_0 by their expressions to evaluate the integral.

$$
\mathrm{E}_{\mathcal{C},\tilde{\boldsymbol{h}}}\left[-\log\left(\min_{\boldsymbol{c} \in \mathcal{C}} \sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{c}))\right)\right]
$$
$$
\le -\frac{1}{K-1}\log\left(\frac{(1-x_0)^{1/2}}{Lc_{2K-1}}\right) + \frac{Lc_{2K-1}}{(1-x_0)^{1/2}}\int_{-\log(x_0)}^{\infty} e^{-z(K-1)}\mathrm{d}z \tag{70}
$$
$$
= -\frac{1}{K-1}\log\left(\frac{(1-x_0)^{1/2}}{Lc_{2K-1}}\right) + \frac{1}{(1-x_0)^{1/2}(K-1)} \tag{71}
$$
$$
= \frac{1}{K-1}\left(\log\left(Lc_{2K-1}\right) + 1\right) + o(1) \tag{72}
$$

as L increases. Dividing by $\log(2)$ yields the final upper bound.

Lower Bound: We start from the lower bound for the CDF given in Proposition 9. It has a form very similar to the CDF for the quantization of a complex vector in the unit-ball in $\mathbb{C}^K$ which is usually used for multiple-antenna BC. Hence, we adapt the approach of the proof of Lemma 3 in [1] to the current setting.

From the lower bound in Proposition 9, we write

$$
\Pr\{\min_{\boldsymbol{c} \in \mathcal{C}}\left(\sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{c}))\right) \le z\} \ge 1 - (1 - c_{2K-1}x^{(K-1)})^L. \tag{73}
$$

A lower bound for the expectation of the logarithm can then be calculated as follows.

$$
\mathrm{E}_{\mathcal{C},\tilde{\boldsymbol{h}}}\left[-\log\left(\min_{\boldsymbol{c} \in \mathcal{C}} \sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{c}))\right)\right] \tag{74}
$$
$$
= \int_0^\infty \Pr\{\min_{\boldsymbol{c} \in \mathcal{C}} \sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{c})) \le e^{-z}\}\mathrm{d}z \tag{75}
$$
$$
\ge \int_0^\infty 1 - (1 - c_{2K-1}e^{-z(K-1)})^L\mathrm{d}z \tag{76}
$$
$$
= \int_0^\infty 1 - \sum_{k=0}^L \binom{L}{k}(-1)^k c_{2K-1}^k e^{-z(K-1)k}\mathrm{d}z \tag{77}
$$
$$
= \frac{1}{K-1}\sum_{k=1}^L \binom{L}{k}(-1)^{k+1}\frac{c_{2K-1}^k}{k} \tag{78}
$$
$$
= \frac{1}{K-1}f(L) \tag{79}
$$

where we have defined $f(p) \triangleq \sum_{k=1}^p \binom{p}{k}(-1)^{k+1}\frac{c_{2K-1}^k}{k}$ for $p \in \mathbb{N}$. To compute the value of $f(L)$, we will use the following relation given in [44, Sec. 0.155].

$$
\sum_{k=0}^n \binom{n}{k}\frac{\alpha^{k+1}}{k+1} = \frac{(\alpha+1)^{n+1} - 1}{n+1}. \tag{80}
$$

We now rewrite $f(L)$ in order to be able to apply (80)

$$f(L) \triangleq \sum_{k=1}^{L} \binom{L}{k} (-1)^{k+1} \frac{c_{2K-1}^k}{k} \tag{81}$$

$$= \sum_{k=1}^{L} \left[\binom{L-1}{k-1} + \binom{L-1}{k} \right] (-1)^{k+1} \frac{c_{2K-1}^k}{k} \tag{82}$$

$$= \sum_{k'=0}^{L-1} \binom{L-1}{k'} (-1)^{k'+2} \frac{c_{2K-1}^{k'+1}}{k'+1} + f(L-1) \tag{83}$$

$$= -\frac{(-c_{2K-1}+1)^L - 1}{L} + f(L-1) \tag{84}$$

$$= \sum_{p=2}^{L} \frac{1 - (-c_{2K-1}+1)^p}{p} + f(1) \tag{85}$$

$$= \sum_{p=1}^{L} \frac{1}{p} - \sum_{p=1}^{L} \frac{1 - (-c_{2K-1}+1)^p}{p}. \tag{86}$$

where we have used (80) to obtain (83) and we have iteratively expressed $f(x)$ in terms of $f(x-1)$ to write (85). Furthermore we have the two following relations:

$$\log(L) \le \sum_{p=1}^{L} \frac{1}{p} \le \log(L) + 1, \tag{87}$$

$$\log(1-x) = -\sum_{n=1}^{\infty} \frac{x^n}{n} \text{ , for } -1 < x < 1. \tag{88}$$

Inserting the expression derived for $f(x)$ inside (79) and using the two bounds provided above, we can obtain the final lower bound as

$$\mathrm{E}_{\mathcal{C}, \tilde{\boldsymbol{h}}} \left[-\log_2 \left(\min_{\boldsymbol{c} \in \mathcal{C}} \sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{c})) \right) \right]$$

$$\ge \frac{1}{(K-1)\log(2)} \sum_{p=1}^{L} \frac{1}{p} - \frac{1}{(K-1)\log(2)} \sum_{p=1}^{L} \frac{(1-c_{2K-1})^p}{p} \tag{89}$$

$$\ge \frac{\log_2(L)}{(K-1)} - \frac{1}{(K-1)\log(2)} \sum_{p=1}^{\infty} \frac{(1-c_{2K-1})^p}{p} \tag{90}$$

$$= \frac{\log_2(L) + \log_2(c_{2K-1})}{(K-1)} \tag{91}$$

where we have used that the constant c_{2K-1} is smaller than one to apply (88) and obtain the term $\log_2(c_{2K-1})$. ∎

B. Proof of Theorem 4

The proof generalizes to the distributed CSI configuration the proof of Theorem 4 in Appendix IV of [1], which derives the number of DoFs for the multiple-antenna BC with finite rate feedback. The generalization is non-trivial due to the fact that in the DCSI-MIMO channel it is not only the inner product between the beamformer and the channel $\tilde{\boldsymbol{h}}_k^{\mathrm{H}} \boldsymbol{t}_i^{(j)}$ which matters, but also the coherency between the coefficients used at the different TXs. Following this difference, we do not use the conventional Grassmannian quantization scheme but we use instead the quantization scheme described in Subsection II-B. In a word, it consists in exploiting the fact that the norm is conserved by the canonical isomorphism between $\mathbb{C}^K$ and $\mathbb{R}^{2K}$, to use the Grassmannian quantization in the

real subspace $\mathbb{R}^{2K-1}$. The reduction to $2K-1$ real dimensions comes from the multiplication by a unit-norm complex number to let the first coefficient be real valued. We then define the angles between vectors in that real linear space. We refer to Appendix X-A for more detail.

The estimation error made at TX j about the channel vector $\tilde{\boldsymbol{h}}_i$ is denoted by $\boldsymbol{\delta}_i^{(j)}$ such that $\boldsymbol{\delta}_i^{(j)} \triangleq \tilde{\boldsymbol{h}}_i - \tilde{\boldsymbol{h}}_i^{(j)}$. The estimation error vectors made at TX j are stacked in the estimation error matrix $\boldsymbol{\Delta}^{(j)}$ defined as

$$\boldsymbol{\Delta}^{(j)} \triangleq \begin{bmatrix} (\boldsymbol{\delta}_1^{(j)})^{\mathrm{H}} \\ (\boldsymbol{\delta}_2^{(j)})^{\mathrm{H}} \\ \vdots \\ (\boldsymbol{\delta}_K^{(j)})^{\mathrm{H}} \end{bmatrix}. \tag{92}$$

We also denote by $\boldsymbol{u}_i^{(j)} \triangleq \boldsymbol{t}_i^{(j)} / \|\boldsymbol{t}_i^{(j)}\|$ the conventional ZF unit-norm beamformer computed at TX j and by $\boldsymbol{u}_i^* \triangleq \boldsymbol{t}_i^* / \|\boldsymbol{t}_i^*\|$ the same beamformer based on perfect CSI. We omit in this proof the superscript cZF for clarity.

Furthermore, we consider in the following that the accuracy of the channel estimates increases with the SNR, i.e., the CSI scaling coefficients $\alpha_i^{(j)}$ are all positive. If there is one pair of indices (i, j) for which $\alpha_i^{(j)} = 0$, then the Euclidean distance between $\boldsymbol{u}_k^{(j)}$ and $\boldsymbol{u}_k^*$ does not decrease with P for all k such that the number of DoFs at all RXs vanishes. When this is not the case, the norm of the channel estimation errors can be approximated as

$$\|\boldsymbol{\delta}_i^{(j)}\|^2 = \|\tilde{\boldsymbol{h}}_i^{(j)} - \tilde{\boldsymbol{h}}_i\|^2 \tag{93}$$

$$= 2 - 2(\tilde{\boldsymbol{h}}_i^{(j)})^{\mathrm{H}} \tilde{\boldsymbol{h}}_i \tag{94}$$

$$= 2 - 2|(\tilde{\boldsymbol{h}}_i^{(j)})^{\mathrm{H}} \tilde{\boldsymbol{h}}_i| \tag{95}$$

$$= 2 - 2\sqrt{1 - \sin^2(\angle(\tilde{\boldsymbol{h}}_i^{(j)}, \tilde{\boldsymbol{h}}_i))} \tag{96}$$

$$= \sin^2(\angle(\tilde{\boldsymbol{h}}_i^{(j)}, \tilde{\boldsymbol{h}}_i)) + o(\sin^2(\angle(\tilde{\boldsymbol{h}}_i^{(j)}, \tilde{\boldsymbol{h}}_i))) \tag{97}$$

where (95) is verified when the channel estimate belongs to the same half-space as the true channel vector. This holds true in this work for the reason explained in Appendix X-A. Equality (96) follows from the definition of the angle between two vectors and (97) is obtained via a Taylor expansion on the first order in the estimation error.

From (92), we conclude that the square norm of the estimation error $\|\boldsymbol{\delta}_i^{(j)}\|^2$ is asymptotically equal to the chordal distance between the channel estimate and the true channel $\sin^2(\angle(\boldsymbol{h}_i^{(j)}, \boldsymbol{h}_i))$ when the SNR increases. The chordal distance corresponds to the distance minimized by the Grassmannian quantization so that this will allow us to apply the theoretical results provided in Appendix X-A. As a preliminary step, we will now evaluate the impact of the estimation error into the computation of the beamformers, i.e., evaluate the norm of the vector $\boldsymbol{u}_i^{(j)} - \boldsymbol{u}_i^*$ for all j.

Lemma 2. *Let us assume that $\forall i, \alpha_i^{(j)} > 0$, then it holds asymptotically as P increases*

$$\mathrm{E}\left[\log_2 \left\| \boldsymbol{u}_i^{(j)} - \boldsymbol{u}_i^* \right\|^2\right] = \mathrm{E}\left[\log_2 \left(\max_{i=1,\dots,K} \left(\sin^2(\angle(\tilde{\boldsymbol{h}}_i^{(j)}, \tilde{\boldsymbol{h}}_i)) \right) \right) \right] + O(1). \tag{98}$$

Proof: We consider w.l.o.g. the precoding at TX j. Since $\forall i, \alpha_i^{(j)} > 0$, the estimation error is infinitely small as P increases and we can do a first order approximation of the channel inverse and write

$$\mathbf{H}^{-1} - (\tilde{\mathbf{H}}^{(j)})^{-1} = -\mathbf{H}^{-1}\boldsymbol{\Delta}^{(j)}\mathbf{H}^{-1} + o(\|\boldsymbol{\Delta}^{(j)}\|_{\mathrm{F}}). \quad (99)$$

Derivation of the Upper Bound: After multiplying by $\boldsymbol{e}_j$ to obtain the j-th beamformer, the Right Hand-Side (RHS) of (99) can then be upper bounded as follows

$$\|(\mathbf{H}^{-1}-(\tilde{\mathbf{H}}^{(j)})^{-1})\boldsymbol{e}_i\|^2 \leq \|\mathbf{H}^{-1}\boldsymbol{\Delta}^{(j)}\mathbf{H}^{-1}\|_{\mathrm{F}}^2 + o(\|\boldsymbol{\Delta}^{(j)}\|_{\mathrm{F}}^2) \tag{100}$$

$$\leq \|\mathbf{H}^{-1}\|_{\mathrm{F}}^4 \|\boldsymbol{\Delta}^{(j)}\|_{\mathrm{F}}^2 + o(\|\boldsymbol{\Delta}^{(j)}\|_{\mathrm{F}}^2) \tag{101}$$

$$\leq K^2 \lambda_{\min}^2(\mathbf{H}) \sum_{k=1}^{K} \|\boldsymbol{\delta}_k^{(j)}\|^2 + o(\|\boldsymbol{\Delta}^{(j)}\|_{\mathrm{F}}^2) \tag{102}$$

with $\lambda_{\min}^2(\mathbf{H})$ denoting the smallest eigenvalue of the channel matrix $\mathbf{H}$. We then take the expectation of the logarithm of this term according to both the channel estimation error and the channel distribution. The term $\log(\lambda_{\min}^2(\mathbf{H}))$ is shown to be integrable and its expectation is given in [45]. The result follows by upper-bounding each of the estimation errors $\|\boldsymbol{\delta}_k^{(j)}\|^2$ by the error which is asymptotically the largest, i.e., the one corresponding to the smallest $\alpha_i^{(j)}$.

Derivation of the Lower Bound: we start by factorizing the estimation error matrix as follows

$$\boldsymbol{\Delta}^{(j)} = \bar{\boldsymbol{\Delta}}^{(j)} \operatorname{diag}([\|\boldsymbol{\delta}_1^{(j)}\|, \|\boldsymbol{\delta}_2^{(j)}\|, \ldots, \|\boldsymbol{\delta}_K^{(j)}\|]) \tag{103}$$

with the columns of $\bar{\boldsymbol{\Delta}}^{(j)}$ consequently normalized to be unit-norm. We then assume w.l.o.g. that the asymptotic largest estimation error corresponds to the channel $\tilde{\boldsymbol{h}}_1$ (i.e., the smallest CSI scaling coefficient is $\alpha_1^{(j)}$). Furthermore, we consider for the sake of exposition that no other channel has the same CSI scaling coefficient. The proof holds similarly if this condition does not hold.

Inserting this definition in (99) and taking the logarithm followed by the expectation over the channel and the estimation errors, we can write equation (104). The absolute value $|\log(\|\mathbf{H}^{-1}\|_{\mathrm{F}}^2)|$ can be upper bounded as in (102) by $|\log(K\lambda_{\min}(\mathbf{H}))|$ whose expectation is shown to exist in [45], thus its expectation also exists. Similarly, the absolute value of the last term of the RHS in (104) can be upper-bounded by an integrable function such that it is also integrable and its expectation is then a $O(1)$, i.e., it remains bounded as the SNR P increases. This concludes the proof. ∎

Proof of Theorem 4: We will now use Lemma 2 to prove the theorem. We consider for simplicity that the CSI scaling coefficients are all different. The proof easily extends to the configurations with some coefficients equal and this is done solely to simplify the exposition. We assume w.l.o.g. that the TX with the smallest CSI scaling coefficient is TX 1.

DoF Lower Bound : We denote by $\boldsymbol{u}_i \in \mathbb{C}^{K \times 1}$ the beamforming vector[4] such that

$$\forall j \in \{1, \ldots, K\}, \quad \{\boldsymbol{u}_i\}_j = \{\boldsymbol{u}_i^{(j)}\}_j = \frac{\{\boldsymbol{t}_i^{(j)}\}_j}{P/K}. \tag{105}$$

We start from the number of DoFs expression in (6) that we rewrite as

$$\mathrm{DoF}_i^{\mathrm{cZF}} = 1 - \lim_{P\to\infty} \mathrm{E}_{\mathbf{H},\{\mathcal{W}_{i,j}\}} \left[\frac{\log_2\left(1 + \frac{P}{K}\sum_{k\neq i}\|\boldsymbol{h}_i\|^2 |\tilde{\boldsymbol{h}}_i^{\mathrm{H}}\boldsymbol{u}_k|^2\right)}{\log_2(P)} \right] \tag{106}$$

$$= -\lim_{P\to\infty} \mathrm{E}_{\mathbf{H},\{\mathcal{W}_{i,j}\}} \left[\frac{\log_2\left(\sum_{k\neq i}|\tilde{\boldsymbol{h}}_i^{\mathrm{H}}\boldsymbol{u}_k|^2\right)}{\log_2(P)} \right]. \tag{107}$$

To obtain a lower bound for the number of DoFs, we need to derive an upper bound for the leaked interference in (107). We start by defining the selection matrices $\mathbf{E}_i = \operatorname{diag}(\boldsymbol{e}_i)$. We can then use this notation to write

$$\left|\tilde{\boldsymbol{h}}_i^{\mathrm{H}}\boldsymbol{u}_k\right|^2 = \left|\tilde{\boldsymbol{h}}_i^{\mathrm{H}}(\boldsymbol{u}_k^* + \sum_{j=1}^{K} \mathbf{E}_j(\boldsymbol{u}_k^{(j)} - \boldsymbol{u}_k^*))\right|^2 \tag{108}$$

$$\leq \|\sum_{j=1}^{K} \mathbf{E}_j(\boldsymbol{u}_k^{(j)} - \boldsymbol{u}_k^*)\|^2 \tag{109}$$

$$= \sum_{j=1}^{K} \left|\boldsymbol{e}_j^{\mathrm{T}}(\boldsymbol{u}_k^{(j)} - \boldsymbol{u}_k^*)\right|^2 \tag{110}$$

$$\leq \sum_{j=1}^{K} \left\|\boldsymbol{u}_k^{(j)} - \boldsymbol{u}_k^*\right\|^2 \tag{111}$$

which we insert in (107) to obtain

$$\mathrm{DoF}_i^{\mathrm{cZF}} \geq -\lim_{P\to\infty} \mathrm{E}_{\mathbf{H},\{\mathcal{W}_{i,j}\}} \left[\frac{\log_2\left(\sum_{k\neq i}\sum_{j=1}^{K}\left\|\boldsymbol{u}_k^{(j)} - \boldsymbol{u}_k^*\right\|^2\right)}{\log_2(P)} \right] \tag{112}$$

$$\geq -\lim_{P\to\infty} \mathrm{E}_{\mathbf{H},\{\mathcal{W}_{i,j}\}} \left[\frac{\log_2\left(\sum_{k\neq i} K \max_j(\|\boldsymbol{u}_k^{(j)} - \boldsymbol{u}_k^*\|^2)\right)}{\log_2(P)} \right]. \tag{113}$$

From (99), the TX whose computed beamformer exhibits the largest mean square error $\|\boldsymbol{u}_k^{(j)} - \boldsymbol{u}_k^*\|^2$ at arbitrarily large SNR P is the TX to whom the lowest CSI scaling coefficient

---

[4]The vector $\boldsymbol{u}_i$ corresponds to the normalized version of $\boldsymbol{t}_i$. Yet, it is exactly unit-norm only when all the TXs share the *same* channel estimate. It is otherwise impossible for the TXs to jointly normalize the beamformer based on different channel estimates. This does not represent a problem in practice because the power constraint is exactly fulfilled in average over the channel estimation errors.

$$\mathrm{E}[\log(\|(\mathbf{H}^{-1} - (\tilde{\mathbf{H}}^{(j)})^{-1})\boldsymbol{e}_i\|^2)] = \mathrm{E}[\log(\|\boldsymbol{\delta}_1^{(j)}\|^2)] + 2\mathrm{E}[\log(\|\mathbf{H}^{-1}\|_{\mathrm{F}}^2)]$$
$$+ \mathrm{E}[\log(\|(\frac{\mathbf{H}^{-1}}{\|\mathbf{H}^{-1}\|_{\mathrm{F}}}\bar{\boldsymbol{\Delta}}^{(j)}\operatorname{diag}([1, \|\boldsymbol{\delta}_2^{(j)}\|/\|\boldsymbol{\delta}_1^{(j)}\|, \ldots, \|\boldsymbol{\delta}_K^{(j)}\|/\|\boldsymbol{\delta}_1^{(j)}\|])\frac{\mathbf{H}^{-1}}{\|\mathbf{H}^{-1}\|_{\mathrm{F}}}\boldsymbol{e}_i)\|^2)] + o(\mathrm{E}[\log(\|\boldsymbol{\Delta}^{(j)}\|_{\mathrm{F}}^2)]) \tag{104}$$

belongs, which is by assumption TX 1. We can then write

$$\mathrm{DoF}_i^{\mathrm{cZF}} \geq \lim_{P \to \infty} \frac{\mathrm{E}_{\mathbf{H}, \{\mathcal{W}_{i,j}\}}\left[-\log_2\left(\sum_{k \neq i} \|\boldsymbol{u}_k^{(1)} - \boldsymbol{u}_k^*\|^2\right)\right]}{\log_2(P)} \tag{114}$$

$$\geq \lim_{P \to \infty} \frac{\mathrm{E}_{\mathbf{H}, \{\mathcal{W}_{i,j}\}}\left[-\log_2\left(\max_i\left(\sin^2(\angle(\tilde{\boldsymbol{h}}_i^{(1)}, \tilde{\boldsymbol{h}}_i))\right)\right)\right]}{\log_2(P)} \tag{115}$$

$$\geq \lim_{P \to \infty} \frac{\min_i B_i^{(1)} + \log_2(c_{2K-1}) + \log_2(e)}{(K-1)\log_2(P)} \tag{116}$$

$$= \min_{i=1,\ldots,K} \alpha_i^{(1)} \tag{117}$$

where (114) is obtained by permuting the expectation and the limit, (115) follows from Lemma 2 and we have used Proposition 11 to obtain inequality (116). The last equation (117) corresponds to the smallest CSI scaling coefficient and provides the lower bound.

DoF Upper Bound: We now derive an upper bound for the number of DoFs, which means a lower bound for the interference. We proceed similarly to (108) but this time to obtain a lower bound for the interference remaining after precoding:

$$\left|\tilde{\boldsymbol{h}}_i^{\mathrm{H}}\boldsymbol{u}_k\right| = \left|\tilde{\boldsymbol{h}}_i^{\mathrm{H}}\boldsymbol{a}_k\right|\|\sum_{j=1}^K \mathbf{E}_j(\boldsymbol{u}_k^{(j)} - \boldsymbol{u}_k^*)\| \tag{118}$$

$$\geq \left|\tilde{\boldsymbol{h}}_i^{\mathrm{H}}\boldsymbol{a}_k\right|\|\mathbf{E}_1(\boldsymbol{u}_k^{(1)} - \boldsymbol{u}_k^*)\| \tag{119}$$

$$= \left|\tilde{\boldsymbol{h}}_i^{\mathrm{H}}\boldsymbol{a}_k\right|\,|\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{b}_k^{(1)}|\|\boldsymbol{u}_k^{(1)} - \boldsymbol{u}_k^*\| \tag{120}$$

where we have defined

$$\boldsymbol{a}_k \triangleq \frac{\boldsymbol{u}_k - \boldsymbol{u}_k^*}{\|\boldsymbol{u}_k - \boldsymbol{u}_k^*\|}, \quad \boldsymbol{b}_k^{(1)} \triangleq \frac{(\boldsymbol{u}_k^{(1)} - \boldsymbol{u}_k^*)}{\|\boldsymbol{u}_k^{(1)} - \boldsymbol{u}_k^*\|}. \tag{121}$$

From (99), we can see that $\boldsymbol{a}_k$ can in fact be written as $\boldsymbol{a}_k = \mathbf{H}^{-1}\tilde{\boldsymbol{a}}_k + o(\max_i \|\boldsymbol{\Delta}^{(j)}\|_{\mathrm{F}})$ with

$$\forall j \in \{1, \ldots, K\}, \quad \boldsymbol{e}_j^{\mathrm{T}}\tilde{\boldsymbol{a}}_k \triangleq -\boldsymbol{e}_j^{\mathrm{T}}\boldsymbol{\Delta}^{(j)}\mathbf{H}^{-1}. \tag{122}$$

Thus, we can write

$$\left|\tilde{\boldsymbol{h}}_i^{\mathrm{H}}\boldsymbol{u}_k\right| \geq |\boldsymbol{e}_i^{\mathrm{T}}\tilde{\boldsymbol{a}}_k|\,|\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{b}_k^{(1)}|\|\boldsymbol{u}_k^{(1)} - \boldsymbol{u}_k^*\|. \tag{123}$$

The two vectors $\tilde{\boldsymbol{a}}_k$ and $\boldsymbol{b}_k^{(1)}$ are isotropically distributed so that the expectation of their logarithm can be easily calculated and is finite. Inserting (122) inside the number of DoFs

formula in (107), we can write the lower bound

$$\mathrm{DoF}_i^{\mathrm{cZF}} = \frac{\lim_{P \to \infty} \mathrm{E}_{\mathbf{H}, \{\mathcal{W}_{i,j}\}}\left[-\log_2\left(\sum_{k \neq i} |\tilde{\boldsymbol{h}}_i^{\mathrm{H}}\boldsymbol{u}_k|^2\right)\right]}{\log_2(P)} \tag{124}$$

$$\geq \frac{\lim_{P \to \infty} \mathrm{E}_{\mathbf{H}, \{\mathcal{W}_{i,j}\}}\left[-\log_2\left(\|\boldsymbol{u}_1^{(1)} - \boldsymbol{u}_1^*\|^2\right)\right]}{\log_2(P)} \tag{125}$$

$$\geq \frac{\lim_{P \to \infty} \mathrm{E}\left[-\log_2\left(\max_i\left(\sin^2(\angle(\tilde{\boldsymbol{h}}_i^{(1)}, \tilde{\boldsymbol{h}}_i))\right)\right)\right]}{\log_2(P)} \tag{126}$$

with inequality (126) obtained from Lemma 2. The proof concludes in the same way as the proof of the upper bound after using Proposition 11. ∎

C. Numerical Values for the Total Feedback Scaling γ

The particular values for the optimization of the feedback sharing can be found in Table I.

D. CSI Scaling Matrix Used in the Simulations

For Fig. 5, the CSI scaling matrix arbitrarily chosen is

$$\boldsymbol{\alpha} = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0.3 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{127}$$

The number of DoFs achieved with the different precoding schemes read as follows:

Precoding Scheme	Number of DoFs
Conventional ZF	0
Active-Passive ZF	2.1
Conventional ZF with HQ	5.3
Active-Passive ZF with HQ	6.3

REFERENCES

[1] N. Jindal, "MIMO Broadcast Channels With Finite-Rate Feedback," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5045–5060, November 2006.

[2] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network Coordination for Spectrally Efficient Communications in Cellular Systems," *IEEE Wireless Communications*, vol. 13, no. 4, pp. 56–61, August 2006.

[3] O. Somekh, O. Simeone, Y. Bar-Ness, and A. M. Haimovich, "Distributed Multi-Cell Zero-Forcing Beamforming in Cellular Downlink Channels," in *Proc. GLOBECOM 2006*.

[4] D. Gesbert, S. Hanly, H. Huang, S. Shamai (Shitz), O. Simeone, and W. Yu, "Multi-Cell MIMO Cooperative Networks: A New Look at Interference," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380–1408, December 2010.

TABLE I
NUMERICAL VALUES FOR THE OPTIMIZATION OF THE FEEDBACK SHARING

Number of transmitting TXs: n	Saturation of the number of DoFs: $n^2(n-1)$	Activation of the $(n+1)$-th TX: $n^2(n+1)$
1	0	2
2	4	12
3	18	36
4	48	80
5	100	150

[5] *Artist 4G deliverable: Feedback from RAN constraints*. Available: https://ict-artist4g.eu/projet/deliverables, [Accessed: 16/05/2012], October 2011.

[6] M. A. Maddah-Ali and D. N. C. Tse, "Completely Stale Transmitter Channel State Information is Still Very Useful," in *Proc. Allerton 2010*.

[7] P. Marsch and G. Fettweis, "On Multicell Cooperative Transmission in Backhaul-Constrained Cellular Systems," *Annals of Telecommunications*, vol. 63, no. 5, 2008.

[8] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai (Shitz), "Downlink Multicell Processing with Limited-Backhaul Capacity," *EURASIP Journal on Advances in Signal Processing*, may 2009.

[9] S. Shamai (Shitz) and M. Wigger, "Rate-Limited Transmitter-Cooperation in Wyners Asymmetric Interference Network," in *Proc. ISIT 2011*.

[10] R. Zakhour and D. Gesbert, "Optimized Data Sharing in Multicell MIMO With Finite Backhaul Capacity," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 6102–6111, december 2011.

[11] G. Caire, N. Jindal, and S. Shamai (Shitz), "On the Required Accuracy of Transmitter Channel State Information in Multiple Antenna Broadcast Channels," in *Proc. Asilomar Conference 2007*.

[12] C. K. Au-Yeung and D. J. Love, "On the Performance of Random Vector Quantization Limited Feedback Beamforming in a MISO System," *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 458–462, february 2007.

[13] D. Samardzija and N. Mandayam, "Unquantized and Uncoded Channel State Information Feedback in Multiple-antenna Multiuser Systems," *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1335–1345, July 2006.

[14] P. Ding, D. J. Love, and M. D. Zoltowski, "Multiple Antenna Broadcast Channels With Shape Feedback and Limited Feedback," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3417–3428, July 2007.

[15] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-Antenna Downlink Channels with Limited Feedback and User Selection," *IEEE Journal on Selected Areas in Communication*, vol. 25, no. 7, pp. 1478–1491, September 2007.

[16] B. Song, F. Roemer, and M. Haardt, "Efficient Channel Quantization Scheme for Multi-user MIMO Broadcast Channels with RBD Precoding," in *Proc. ICASSP 2008*.

[17] M. B. Shenouda and T. N. Davidson, "Robust Linear Precoding for Uncertain MISO Broadcast Channels," in *Proc. ICASSP 2006*.

[18] N. Vucic, H. Boche, and S. Shi, "Robust Transceiver Optimization in Downlink Multiuser MIMO Systems," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3576–3587, September 2009.

[19] E. Björnson, R. Zakhour, D. Gesbert, and B. Ottersten, "Cooperative Multicell Precoding: Rate Region Characterization and Distributed Strategies With Instantaneous and Statistical CSI," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4298–4310, august 2010.

[20] M. Kobayashi, M. Debbah, and J.-C. Belfiore, "Outage Efficient Strategies for Network MIMO with Partial CSIT," in *Proc. ISIT 2009*.

[21] A. Tajer, N. Prasad, and X. Wang, "Robust Linear Precoder Design for Multi-Cell Downlink Transmission," *IEEE Transactions on Signal Processing*, vol. 59, no. 1, pp. 235–251, january 2011.

[22] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO With Linear Zero-Forcing Beamforming: Large System Analysis, Impact of Channel Estimation, and Reduced-Complexity Scheduling," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.

[23] I. Sohn, C. S. Park, and K. B. Lee, "Downlink Multiuser MIMO Systems With Adaptive Feedback Rate," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 3, pp. 1445–1451, March 2012.

[24] J. Zhang and J. G. Andrews, "Adaptive Spatial Intercell Interference Cancellation in Multicell Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1455–1468, December 2010.

[25] B. Özbek and D. Le Ruyet, "Adaptive Limited Feedback for Intercell Interference Cancelation in Cooperative Downlink Multicell Networks," in *Proc. ISWCS 2010*.

[26] W. W. L. Ho, T. Q. S. Quek, S. Sun, and R. W. Heath, "Decentralized Precoding for Multicell MIMO Downlink," *IEEE Transactions on Wireless Communications*, vol. 10, no. 6, pp. 1798–1809, June 2011.

[27] R. Bhagavatula and R. W. Heath, "Adaptive Limited Feedback for Sum-Rate Maximizing Beamforming in Cooperative Multicell Systems," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 800–811, february 2011.

[28] H. A. A. Saleh and S. D. Blostein, "Single-Cell vs. Multicell MIMO Downlink Signalling Strategies with Imperfect CSI," in *Proc. GLOBECOM 2010*.

[29] A. Tajer and X. Wang, "Information Exchange Limits in Cooperative MIMO Networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2927–2942, June 2011.

[30] B. Khoshnevis, W. Yu, and Y. Lostanlen, "Two-Stage Channel Feedback for Beamforming and Scheduling in Network MIMO Systems," in *Proc. ICC 2012 (to appear)*.

[31] N. Lee and W. Shin, "Adaptive Feedback Scheme on K-Cell MISO Interfering Broadcast Channel with Limited Feedback," *IEEE Transactions on Wireless Communications*, vol. 10, no. 2, pp. 401–406, february 2011.

[32] R. Bhagavatula and R. W. Heath, "Adaptive Bit Partitioning for Multicell Intercell Interference Nulling with Delayed Limited Feedback," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3824–3836, August 2011.

[33] J. Marschak and R. Radner, *Economic Theory of Teams*. Yale University Press, New Haven and London, february 1972.

[34] R. Zakhour and D. Gesbert, "Team Decision for the Cooperative MIMO Channel with Imperfect CSIT Sharing," in *Proc. ITA 2010*.

[35] V. R. Cadambe and S. A. Jafar, "Interference Alignment and Degrees of Freedom of the K-User Interference Channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3425–3441, August 2008.

[36] R. Etkin, D. Tse, and H. Wang, "Gaussian Interference Channel Capacity to Within One Bit," *IEEE Trans. on Information Theory*, vol. 54, no. 12, dec. 2008.

[37] A. Lozano, J. G. Andrews, and R. W. Heath, "On The Limitations of Cooperation in Wireless Networks," in *Proc. ITA 2012*.

[38] P. de Kerret and D. Gesbert, "The Multiplexing Gain of a Two-cell MIMO Channel with Unequal CSI," in *Proc. ISIT 2011*.

[39] W. Santipach and M. L. Honig, "Capacity of a Multiple-Antenna Fading Channel With a Quantized Precoding Matrix," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 1218–1234, March 2009.

[40] C. S. Vaze and M. K. Varanasi, "CSI feedback scaling rate vs multiplexing gain tradeoff for DPC-based transmission in the Gaussian MIMO broadcast channell," in *Proc. ISIT 2010*.

[41] F. Boccardi, H. Huang, and A. Alexiou, "Hierarchical Quantization and its Application to Multiuser Eigenmode Transmissions for MIMO Broadcast Channels with Limited Feedback," in *Proc. PIMRC 2007*.

[42] T. Cover and A. Thomas, *Elements of Information Theory*. Wiley-Interscience, July 2006.

[43] W. Dai, Y. Liu, and B. Rider, "Quantization Bounds on Grassmann Manifolds and Applications to MIMO Communications," *IEEE Transactions on Information Theory*, vol. 54, no. 3, pp. 1108–1123, March 2008.

[44] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. Elsevier/Academic Press, Amsterdam, 2007.

[45] A. Edelman, "Eigenvalues and Condition Numbers of Random Matrices," Ph.D. dissertation, MIT, 1989.

PLACE
PHOTO
HERE

David Gesbert (IEEE Fellow) is Professor and Head of the Mobile Communications Department, EURE-COM, France, where he also heads the Communications Theory Group. He obtained the Ph.D degree from Ecole Nationale Superieure des Telecommunications, France, in 1997. From 1997 to 1999 he has been with the Information Systems Laboratory, Stanford University. In 1999, he was a founding engineer of Iospan Wireless Inc, San Jose, Ca.,a startup company pioneering MIMO-OFDM (now Intel). Between 2001 and 2003 he has been with the Department of Informatics, University of Oslo as an adjunct professor. D. Gesbert has published about 170 papers and several patents all in the area of signal processing, communications, and wireless networks.

D. Gesbert was a co-editor of several special issues on wireless networks and communications theory, for JSAC (2003, 2007, 2009), EURASIP Journal on Applied Signal Processing (2004, 2007), Wireless Communications Magazine (2006). He served on the IEEE Signal Processing for Communications Technical Committee, 2003-2008. He's an associate editor for IEEE Transactions on Wireless Communications and the EURASIP Journal on Wireless Communications and Networking. He authored or co-authored papers winning the 2004 IEEE Best Tutorial Paper Award (Communications Society) for a 2003 JSAC paper on MIMO systems, 2005 Best Paper (Young Author) Award for Signal Proc. Society journals, and the Best Paper Award for the 2004 ACM MSWiM workshop. He co-authored the book Space time wireless communications: From parameter estimation to MIMO systems, Cambridge Press, 2006.

PLACE
PHOTO
HERE

Paul de Kerret (IEEE Student Member) was born in 1987 in Paris, France. In 2009, he graduated from Ecole Nationale Superieure des Telecommunications de Bretagne, France and obtained a diploma degree in electrical engineering from Munich University of Technology (TUM), Germany. He also earned a four year degree in mathematics at the Universite de Bretagne Occidentale, France in 2008. From january 2010 to september 2010, he has been a research assistant at the Institute for Theoretical Information Technology, RWTH Aachen University, Germany. Since October 2010, he is working toward a Ph.D. degree in the Mobile Communications Department at EURECOM, France.

The main focus of his current work is on the design and the analysis of distributed cooperation schemes between transmitters in wireless networks. His research interests include signal processing in wireless networks, multi-user MIMO systems and multi-user information theory.