Spatial CSIT Allocation Policies for Network MIMO Channels

Paul de Kerret and David Gesbert

Mobile Communications Department, Eurecom

Campus SophiaTech, 450 Route des Chappes, 06410 Biot, France

{dekerret@eurecom.fr,gesbert@eurecom.fr}

Abstract

In this work[1], we study the problem of the optimal dissemination of channel state information (CSI) among K spatially distributed transmitters (TXs) jointly cooperating to serve K receivers (RXs). One of the particularities of this work lies in the fact that the CSI is *distributed* in the sense that each TX obtains its *own* estimate of the global multi-user MIMO channel with no further exchange of information being allowed between the TXs. Although this is well suited to model the cooperation between non-colocated TXs, e.g., in cellular Coordinated Multipoint (CoMP) schemes, this type of setting has received little attention so far in the information theoretic society. We study in this work what are the CSI requirements at every TX, as a function of the network geometry, to ensure that the maximal number of degrees-of-freedom (DoF) is achieved, i.e., the same DoF as obtained under perfect CSI at all TXs. We advocate the use of the generalized DoF to take into account the geometry of the network in the analysis. Consistent with the intuition, the derived DoF maximizing CSI allocation policy suggests that TX cooperation should be limited to a specific finite neighborhood around each TX. This is in sharp contrast with the conventional (uniform) CSI dissemination policy which induces CSI requirements that grow unbounded with the network size. The proposed CSI allocation policy suggests an alternative to clustering which overcomes fundamental limitations such as (i) edge interference and (ii) unbounded increase of the CSIT requirements with the cluster size. Finally, we show how finite neighborhood CSIT exchange translates into finite neighborhood message exchange so that finally global interference management is possible with only local cooperation.

[1]This work has been performed under the Celtic-Plus project SHARING. Preliminary results have been published in [1].

I. INTRODUCTION

Network (or Multicell) MIMO methods, whereby multiple interfering transmitters (TXs) share user messages and allow for joint precoding, are currently considered for next generation wireless networks [2]. 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. However, the sharing of the user's data symbols and the CSI to all cooperating TXs imposes huge requirements on the backhaul architecture, particularly as the number of cooperating TXs increases.

Consequently, the cooperation is usually limited to small *cooperation clusters* inside which the TXs cooperate. The optimal way of forming these clusters has recently become an active research topic [3]–[7]. Still, clustering leads to some fundamental limitations. Firstly, there is inevitably inter-cluster interference on the boundaries of the cluster and secondly, it requires the obtaining at all the TXs inside the cluster of the CSI relative to the entire cluster which means that the amount of CSI feedback required quickly increases with the number of TXs inside the cluster. Several works have focused on determining the optimal size of the clusters when taking into account the cost of estimating the channel elements, e.g., [8], [9]. They suggest that TX cooperation cannot efficiently manage interference, even if the backhaul links are strong enough to form large clusters. The main message behind [9] is that pilot-based channel estimates incurs a substantial loss when trying to learn the channel from a large number of users within a finite coherence time interval, causing the DoF to saturate. We do not focus in this work on the estimation of the channel but only on the problem of uplink feedback and of CSI sharing between the TXs so that our results do not directly challenge the conclusions of [9] but are in fact complementary. In particular a new perspective arises from the accounting of path loss modeling (and network geometry) in the feedback requirement analysis.

One other important element is that we do not restrict to clustering which has the aforementioned limitations but instead we allow each TX to obtain the CSI relative to any other TX or RX, with the only constraint being on the total amount of information exchanged. Note that the optimization of the feedback allocation in the case where the user's data symbols are *not* shared between the TXs (i.e., coordinated beamforming) yields a completely different problem setting [10]–[16].

The question that we state is whether it is possible to overcome the fundamental limitations of clustering by optimizing directly the spatial allocation of CSIT. Hence, we study the minimization of the CSI shared across the TXs subject to a given required performance. To tackle this intricate question, we consider the high SNR regime and we study the number of degrees-of-freedom (DoF) achieved. We consider also that the pathloss of the interfering links is parameterized as some function of the SNR. This parameterization allows to model the network geometry and leads to analyze the *generalized DoF* as in [17]–[23]. This modeling of the pathloss as a function of the SNR is essential to model the effect of the network geometry in a DoF analysis where the SNR is assumed to become infinitely large. Indeed, omitting to use such a parameterization and letting the SNR become large makes the pathloss differences (i.e., the network geometry) negligible.

In previous works [24], [25], the DoF and the rate offset have been considered in a distributed CSI setting, where all the wireless links between a TX and a RX have the same pathloss. The focus of [24] is on the derivation of robust precoders and the approach is completely different due to the restrictive geometry with homogeneous pathloss only.

In this work, we provide a CSIT dissemination policy, denoted as *distanced-based* allowing to achieve the same generalized DoF as that of a cooperative network with perfect CSI at every TX. This CSIT dissemination policy requires only the sharing of the user's data symbols and of the CSI to within a neighborhood which does not increase with the size of the network. Hence, we show that the pathloss attenuation effectively limits the impact of interference to a local neighborhood around each TX and allows for global interference management with only local cooperation.

The analysis is carried out in a 2-dimensional network where the *interference level* of a wireless channel, defined as in [17], increases regularly as the distance between the nodes increases.

Notations: We denote by $\boldsymbol{e}_i$ the i-th column of the $K \times K$ identity matrix, by $\bullet^{\mathrm{H}}$ the Hermitian transpose, and by δ_{ij} the Kronecker symbol which is equal to 1 if $j = i$ and to zero otherwise. The operator $[\bullet]^+$ takes the maximum between the real argument and 0, and $\lceil \bullet \rceil$ denotes the ceiling operator. $|\mathcal{A}|$ is used to denote the cardinality of the finite set $\mathcal{A}$. The complex circularly symmetric Gaussian distribution with zero mean and variance σ^2 is represented by $\mathcal{N}(0, \sigma^2)$. The (ij)-th elements of a matrix $\mathbf{A}$ is denoted equivalently as $\{\mathbf{A}\}_{ij}$ or as A_{ij}. Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{C}^n$ and $x \in \mathbb{R}$, we write $\boldsymbol{a} = \boldsymbol{b} + o(x)$ if $\exists \boldsymbol{\tau} \in \mathbb{C}^n, \boldsymbol{a} = \boldsymbol{b} + \boldsymbol{\tau}$ with $\lim_{x \to 0} x \|\boldsymbol{\tau}\| = 0$. We use the

exponential equality $f(x) \doteq x^b$ as in [26] to denote that $\lim_{x\to\infty} \frac{\log(f(x))}{\log(x)} = b$.

II. System Model

A. Network MIMO

We consider a network MIMO setting in which K non-colocated transmitters (TXs) transmit *jointly* via linear precoding to K receivers (RXs) equipped with a single antenna and applying single user decoding. The extension to multiple-antennas TXs will be discussed later in this work. If the RXs are equipped with multiple-antennas but apply a matched-filter, the analysis directly holds for the virtual MISO channel obtained. Taking into account the zero forcing (ZF) capabilities of the RX is however out of the scope of this work.

Each TX initially has the knowledge of the K data symbols to transmit to the K RXs (owing to TX cooperation friendly routing protocol for user-plane data). Note that this assumption will be challenged in Section IV. The transmission is then described as

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

where $\boldsymbol{y} \triangleq [y_1, \ldots, y_K]^{\mathrm{H}} \in \mathbb{C}^{K\times 1}$ contains the received signals at the K RXs, $\boldsymbol{h}_i^{\mathrm{H}} \in \mathbb{C}^{1\times K}$ is the channel to the i-th RX, $\boldsymbol{\eta} \triangleq [\eta_1, \ldots, \eta_K]^{\mathrm{H}} \in \mathbb{C}^{K\times 1}$ is the i.i.d. $\mathcal{N}(0,1)$ normalized noise at the RXs, and $\mathbf{x} \triangleq [x_1, \ldots, x_K]^{\mathrm{H}} \in \mathbb{C}^{K\times 1}$ represents the transmit signals at the K TXs.

We also define the multi-user channel $\mathbf{H} \triangleq [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K]^{\mathrm{H}}$. H_{ki} designates the fading coefficient between TX i and RX k. We consider a Rayleigh fast fading channel such that $H_{ki} = \sigma_{ki}\tilde{H}_{ki}$ where $\tilde{H}_{ki} \sim \mathcal{N}(0,1)$ is a Gaussian random variable and the value of σ_{ki} will reflect the geometry (topology) of the network. We consider in the following for the sake of clarity that $\forall i, \sigma_{ii}^2 = 1$.

The transmit signal $\mathbf{x}$ is obtained from the user's data symbols $\boldsymbol{s} \triangleq [s_1, \ldots, s_K]^{\mathrm{H}} \in \mathbb{C}^{K\times 1}$ (i.i.d. $\mathcal{N}(0,1)$) as

$$\mathbf{x} = \begin{bmatrix} \boldsymbol{t}_1 & \ldots & \boldsymbol{t}_K \end{bmatrix} \cdot \begin{bmatrix} s_1 \\ \vdots \\ s_K \end{bmatrix}. \tag{2}$$

Hence, the vector $\boldsymbol{t}_i \in \mathbb{C}^{K\times 1}$ represents the beamforming vector used to transmit s_i to RX i and we define as $\mathbf{T} \triangleq [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_K] \in \mathbb{C}^{K\times K}$ the multi-user joint precoder. We consider a sum power

constraint and an equal power allocation to the users, both for clarity and because it does not impact the DoF. Note that because of the normalization of the noise and of the direct channels, P denotes also the average per-stream SNR. The ergodic rate of user i is then written as

$$R_i \triangleq \mathrm{E}\left[\log_2\left(1+\frac{|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_i|^2}{1+\sum_{\ell\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_\ell|^2}\right)\right]. \tag{3}$$

The *DoF* at RX i is then defined as

$$\mathrm{DoF}_i \triangleq \lim_{P\to\infty}\frac{R_i}{\log_2(P)}. \tag{4}$$

We use the notion of *generalized DoF* [17]–[23] to represent the attenuation of the interference in relation to the power. Hence, the generalized DoF at RX i is defined by

$$\mathrm{DoF}_i(\{\mathbf{B}^{(j)}\}_{j=1}^K,\mathbf{\Gamma}) \triangleq \lim_{P\to\infty}\frac{R_i}{\log_2(P)} \quad \text{, subject to } \sigma_{ki}^2 = P^{\{\mathbf{\Gamma}\}_{ki}-1}, \qquad \forall k,i \tag{5}$$

where the CSIT allocation $\{\mathbf{B}^{(j)}\}_{j=1}^K$ and the precoding used will be described in the following. The matrix $\mathbf{\Gamma}\in[-\infty,1]^{K\times K}$ is called the *interference level matrix* and can be written as

$$\{\mathbf{\Gamma}\}_{ki} = \frac{\log(\sigma_{ki}^2 P)}{\log(P)}, \qquad \forall k,i. \tag{6}$$

Remark 1. As already discussed in the introduction, the goal of the generalized DoF is not to model an unrealistic channel where the pathloss increases with the SNR, in the same way that a DoF analysis does not really apply for transmission with infinite amount of power. It consists simply, starting from a practical setting, in letting both the SNR and the pathloss increase at the same time, instead of letting simply the SNR increase, as in a conventional DoF analysis. This ensures that the differences of power between the wireless links do not become negligible when considering the high SNR regime and hence allows us to take into account the geometry of the network in our analysis.

Let for example consider a transmission at finite SNR P where TX i interferes to RX k through a channel with a pathloss $\sigma_{ki}^2 = 0.01$. In terms of DoF, the pathloss is a constant and the DoF would be the same if $\sigma_{ki}^2 = 1$. In contrast, we would consider in a generalized DoF that the interference level matrix verifies $\{\mathbf{\Gamma}\}_{ki} = \log(\sigma_{ki}^2 P)/\log(P)$, thus letting the pathloss difference impact our analysis. □

B. Distributed CSI at the TXs

The joint precoder is implemented distributively at the TXs with each TX relying solely on its own estimate of the channel matrix in order to compute its transmit coefficient, without any exchange of information with the other TXs [24], [27]. To model the imperfect CSI at the TX (CSIT), the channel estimate at each TX is assumed to be obtained from a limited rate digital feedback scheme. Consequently, we introduce the following definitions.

Definition 1 (Distributed Finite-Rate CSIT)**.** *We represent a* CSIT allocation *by the collections of matrices* $\{\mathbf{B}^{(j)}\}_{j=1}^{K}$ *where* $\mathbf{B}^{(j)} \in \mathbb{R}_{+}^{K\times K}$ *denotes the CSIT allocation at TX* j*. Hence, TX* j *receives the multi-user channel estimate* $\mathbf{H}^{(j)}$ *defined from*

$$H_{ki}^{(j)} = \sigma_{ki}\tilde{H}_{ki} + \sigma_{ki}\sqrt{2^{-B_{ki}^{(j)}}}\Delta\tilde{H}_{ki}^{(j)}, \qquad \forall i,k \tag{7}$$

where $\Delta\tilde{H}_{ki}^{(j)} \sim \mathcal{N}(0,1)$ *and the* $\Delta\tilde{H}_{ki}^{(j)}$ *are mutually independent and independent of the channel.*

Remark 2. The reasons for modeling the imperfect CSIT via (7) are as follows. First, it is well known from rate-distortion theory that the minimal distortion when quantizing a standard Gaussian source using B bits is equal to 2^{-B} [28, Theorem 13.3.3] while this distortion value is also achieved up to a multiplicative constant using for example the Lloyd algorithm [29] or even scalar quantization. Thus, the decay in 2^{-B} as the number of quantizations bits increases, represents a reasonable model.

Furthermore, only the asymptotic behavior exponentially in the SNR is of interest in this work such that the distribution of the CSIT error does not matter here. We have chosen a Gaussian model for simplicity but other distributions fulfilling some regularity constraints could be chosen. □

It is a well known result that the number of CSI feedback bits should scale with the SNR in order to achieve a positive DoF in MISO BCs [24], [30], [31]. Hence, the prelog factor represents an appropriate measure at high SNR of the amount of CSIT required. Thus, we define the *size* of a CSIT allocation as follows.

Definition 2 (Size of a CSIT allocation)**.** *The size* $\mathrm{s}(\bullet)$ *of a CSIT allocation* $\mathbf{B}^{(j)}$ *at TX* j *is defined as*

$$\mathrm{s}(\mathbf{B}^{(j)}) \triangleq \lim_{P\to\infty} \frac{\sum_{i,k} B_{ki}^{(j)}}{\log_2(P)} \tag{8}$$

such that the total size of a CSIT allocation $\{\mathbf{B}^{(j)}\}_{j=1}^{K}$ *is*

$$\mathrm{s}(\{\mathbf{B}^{(j)}\}_{j=1}^{K}) \triangleq \sum_{j=1}^{K} \mathrm{s}(\mathbf{B}^{(j)}) \tag{9}$$

$$= \lim_{P\to\infty} \frac{\sum_{i,j,k} B_{ki}^{(j)}}{\log_2(P)}. \tag{10}$$

Remark 3. We consider here a digital quantization of the channel vectors but the results can be easily translated to a setting where analog feedback is used since digital quantization is simply used as a way to quantify the variance of the CSIT errors [31], [32]. Furthermore, only CSI requirements at the TXs are investigated, and different scenarios can be envisaged for the sharing of the channel estimates (e.g., direct broadcasting from the RXs to all the TXs, sharing through a backhaul, ...) [33]. □

C. Distributed precoding

Based on its individual CSIT, each TX designs its transmit coefficients. We focus here on the CSI dissemination problem under a conventional precoding framework. Hence, we assume that the sub-optimal zero forcing (ZF) precoder is used. Based on its own channel estimate $\mathbf{H}^{(j)}$, TX j computes then the ZF beamforming vector $\boldsymbol{t}_i^{(j)}$ to transmit symbol s_i such that

$$\boldsymbol{t}_i^{(j)} \triangleq \sqrt{P}\frac{\left(\mathbf{H}^{(j)}\right)^{-1}\boldsymbol{e}_i}{\|\left(\mathbf{H}^{(j)}\right)^{-1}\boldsymbol{e}_i\|}, \qquad \forall i \in \{1,\ldots,K\}. \tag{11}$$

Remark 4. ZF represents a priori a sub-optimal precoding scheme. It is however well known to achieve the maximal DoF in the MIMO BC with perfect CSIT [30], [31]. Furthermore, considering limited feedback in the compound MIMO BC, it is revealed in [34] that no other precoding scheme can achieve the maximal DoF with a lower feedback scaling. This confirms the efficiency of ZF in terms of DoF, even when confronted with imperfect CSI. ZF represents also the most widely used scheme to manage interference at high SNR.

An exciting yet challenging question is whether there exist strictly better schemes (from a DoF point of view) dealing specifically with the distributed CSI case. This question is however beyond the scope of our work here. □

Although a given TX j may compute the whole precoding matrix $\mathbf{T}^{(j)}$, only the j-th row is of practical interest. Indeed, TX j transmits solely $x_j = \boldsymbol{e}_j^{\mathrm{H}}\mathbf{T}^{(j)}\boldsymbol{s}$. The effective multi-user precoder $\mathbf{T}$ verifies then

$$\mathbf{x} = \mathbf{T}\boldsymbol{s} = \begin{bmatrix} \boldsymbol{e}_1^{\mathrm{H}}\mathbf{T}^{(1)} \\ \boldsymbol{e}_2^{\mathrm{H}}\mathbf{T}^{(2)} \\ \vdots \\ \boldsymbol{e}_K^{\mathrm{H}}\mathbf{T}^{(K)} \end{bmatrix} \boldsymbol{s}. \tag{12}$$

Remark 5. Each TX independently proceeds with the normalization of the beamformer independently and based on a-priori different channel estimates. Hence, the power constraint is only approximately fulfilled. Yet, the power constraint is asymptotically fulfilled for all the DoF achieving CSIT allocations that we will consider in the following. □

Finally, we denote by $\mathbf{T}^\star = [\boldsymbol{t}_1^\star, \ldots, \boldsymbol{t}_K^\star]$ the precoder obtained with perfect CSI at all TXs. It then verifies

$$\boldsymbol{t}_i^\star \triangleq \sqrt{P}\frac{(\mathbf{H})^{-1}\boldsymbol{e}_i}{\|(\mathbf{H})^{-1}\boldsymbol{e}_i\|}, \qquad \forall i \in \{1, \ldots, K\}. \tag{13}$$

D. Optimization of the CSIT allocation

Optimizing directly the allocation of the number of bits at finite SNR represents a challenging problem which gives little hope for analytical results. Instead, we will try to identify one CSIT allocation solution achieving the same DoF as under the fully shared CSIT setting.

Definition 3. *We define the set of DoF-achieving CSIT allocations* $\mathbb{B}_{\mathrm{DoF}}(\mathbf{\Gamma})$ *as*

$$\mathbb{B}_{\mathrm{DoF}}(\mathbf{\Gamma}) \triangleq \{\{\mathbf{B}^{(j)}\}_{j=1}^K | \forall i, \mathrm{DoF}_i(\{\mathbf{B}^{(j)}\}_{j=1}^K, \mathbf{\Gamma}) = 1\}. \tag{14}$$

Hence, an interesting problem consists in finding the minimal CSIT allocation (where minimality refers to the size in Definition 2) which achieves the maximal generalized DoF at every user:

$$\text{minimize}\, \mathrm{s}\left(\{\mathbf{B}^{(j)}\}_{j=1}^K\right), \text{ subject to } \{\mathbf{B}^{(j)}\}_{j=1}^K \in \mathbb{B}_{\mathrm{DoF}}(\mathbf{\Gamma}). \tag{15}$$

In this paper, we focus on an "achievability" result, by exhibiting a CSIT allocation that achieves the maximal DoF while having a much lower size than the conventional (uniform) CSIT allocation. Furthermore, the proposed "achievable scheme" will prove to have particularly interesting properties, which distinguish it from other solutions in the literature (e.g., clustering). The

problem of finding a minimal-size allocation policy while guaranteeing full DoF (i.e. DoF equal to the perfect CSIT case) is an interesting problem, but an extreme challenging one, which, to our best knowledge, remains open.

III. Preliminary Results

As a preliminary step, we derive a simple sufficient criterion on the precoder for achieving the maximal DoF.

Proposition 1. *The maximal DoF is achieved by using the precoder* $\mathbf{T}$ *if the CSIT allocation* $\{\mathbf{B}^{(j)}\}_{j=1}^{K}$ *is such that*

$$\mathrm{E}\left[\|\mathbf{T}-\mathbf{T}^{\star}\|_{\mathrm{F}}^{2}\right] \doteq P^{0} \tag{16}$$

where the equivalence sign $f(P) \doteq P^{b}$ *denotes the exponential equality* $\lim_{P\to\infty} \frac{\log(f(P))}{\log(P)} = b$ *[26] and* $\mathbf{T}^{\star}$ *has been defined previously as the precoder based on perfect CSIT.*

Proof: A detailed proof is provided in Appendix I. ■

The condition obtained above is very intuitive and will be used in the remaining of this work. However, Proposition 1 does not solve the main question, which is to determine what kind of CSIT allocation allows to achieve (16). This question is central to this work and will be tackled in Section IV.

A. The conventional CSIT allocation is DoF achieving

The term "conventional" hereby corresponds to conveying to each TX the CSI relative to the full multi-user channel, enabling all the TXs to do the same processing and compute a common precoder $\mathbf{T}^{(j)} = \hat{\mathbf{T}}$. Hence, the condition of Proposition 1 can be rewritten as

$$\mathrm{E}\left[\left\|\mathbf{T}^{(j)}-\mathbf{T}^{\star}\right\|_{\mathrm{F}}^{2}\right] \doteq P^{0}, \qquad \forall j \in \{1,\ldots,K\}. \tag{17}$$

Based on this, the following result is obtained.

Proposition 2. *The following "conventional" CSIT allocation* $\{\mathbf{B}^{\mathrm{conv},(j)}\}_{j=1}^{K}$ *such that*

$$\{\mathbf{B}^{\mathrm{conv},(j)}\}_{ki} = [\lceil \log_2(P\sigma_{ki}^2) \rceil]^{+}, \qquad \forall k,i,j \tag{18}$$

$$= [\lceil \{\mathbf{\Gamma}\}_{ki} \log_2(P) \rceil]^{+} \tag{19}$$

is DoF achieving, i.e., $\{\mathbf{B}^{\mathrm{conv},(j)}\}_{j=1}^{K} \in \mathbb{B}_{\mathrm{DoF}}$.

Proof: A detailed proof is provided in Appendix II. ■

This CSIT allocation provides to each TX the K channel vectors relative to the K RXs. This means that each TX requires a number of channel estimates growing unbounded with K. This represents a serious issue in large/dense networks which prompts designers, in practice, to restrict cooperation to small cooperations clusters.

B. CSIT allocation with distributed precoding

We now turn our attention to the derivation of a more efficient CSIT allocation strategy. A crucial observation is that each TX does not need to compute accurately the *full* precoder. Indeed, the sufficient criterion (16) can be written in the distributed CSI setting as

$$\mathrm{E}\left[\left\|\boldsymbol{e}_j^{\mathrm{H}}(\mathbf{T}^{(j)}-\mathbf{T}^{\star})\right\|^2\right] \doteq P^0, \qquad \forall j \in \{1,\ldots,K\}. \tag{20}$$

Intuition has it that a channel coefficient relative to a TX/RX pair which interferes little with TX/RX j has little impact on the jth precoding row and hence does not need to be known accurately at TX j. What follows is a quantitative assessment of this intuition.

IV. Distance-Based CSIT Allocation

We consider in this section a two dimensional network consisting of K TX/RX pairs. The geometry of the network is abstracted as follows. Let us assume that there is a given map from the user's indices $i \in \{1,\ldots,K\}$ to some coordinates $(x_i, y_i) \in \mathbb{R}^2$. TX/RX pair i is then assumed to be located at the position (x_i, y_i). We denote by $\mathrm{d}(i,k)$ the Euclidian distance over $\mathbb{R}^2$. The pathloss between TX i and RX k is then given by

$$\sigma_{k,i}^2 = (\mu^2)^{\mathrm{d}(i,k)}, \qquad \forall i,k \in \{1,\ldots,K\} \tag{21}$$

for a given $\mu^2 < 1$. Equation (6) is then rewritten as

$$\{\mathbf{\Gamma}\}_{ki} = \frac{\log((\mu^2)^{\mathrm{d}(i,k)}P)}{\log(P)}, \qquad \forall k,i \tag{22}$$

$$= 1 + (\gamma - 1)\,\mathrm{d}(k,i) \tag{23}$$

where we have introduced

$$\gamma = \frac{\log(\mu^2 P)}{\log(P)}. \tag{24}$$

A. Distance-based CSIT allocation

With the notations set above, the conventional CSIT allocation given in (19) can be rewritten as

$$\{\mathbf{B}^{\mathrm{conv},(j)}\}_{ki} = \lceil [1 + (\gamma - 1)\,\mathrm{d}(k,i)]^{+} \log_2(P) \rceil, \qquad \forall k, i, j. \tag{25}$$

We can now state our first main result.

Theorem 1. *Let us define the CSIT allocation* $\{\mathbf{B}^{\mathrm{dist},(j)}\}_{j=1}^{K}$ *such that*

$$\{\mathbf{B}^{\mathrm{dist},(j)}\}_{ki} = \lceil [1 + (\gamma - 1)(\mathrm{d}(j,k) + \mathrm{d}(k,i))]^{+} \log_2(P) \rceil, \qquad \forall k, i, j. \tag{26}$$

Then $\mathbf{B}^{\mathrm{dist}} \in \mathbb{B}_{\mathrm{DoF}}$.

Proof: A detailed proof is provided in Appendix III. ■

Letting γ tend to one inside (26), the network geometry becomes homogeneous with all the links having the same variance, asymptotically in the SNR. There is then no attenuation of the interference due to the pathloss and the distance-based CSIT allocation converges as expected to the conventional CSIT allocation given in (25). More generally, the distance-based CSIT allocation exploits the fact that two users being further away that a given distance do not impact the design of precoding coefficients for each other. This can be related to the the space of infinite (either polynomially or exponentially) decaying matrices being closed under inversion [35], [36].

Remark 6. The result easily extends to the case of multiple TX/RX pairs located at the same position. This models then a TX with multiple-antennas serving multiple single-antenna RXs. □

B. Scaling properties of the Distance-based CSIT allocation

We consider in this section the scaling behaviour as the number of TX/RX pairs increases. It is then differenciated in the literature between so-called *dense* networks and *extended* networks [37], [38]. In the first model, the size of the network remains constant and the density (number of TX/RX pairs/m^2) increases, while in the second the density of the network remains constant as the number of TX/RX pairs increases. Our analysis being on large networks, we consider the extended model and we assume that the density of TX/RX pairs remains constant, i.e., that the size of the network increases with the number of TX/RX pairs.

The critical question that we want to tackle here is to determine how the size of the CSIT allocation increases with the number of TX/RX pairs. Indeed, this scaling represents an important figure-of-merit to evaluate the feasibility of large cooperation areas.

Corollary 1. *In the distance-based CSIT allocation, TX j does not need to receive any CSIT relative to the ith RX if*

$$\mathrm{d}(i,j) > d_0 \tag{27}$$

with d_0 defined as

$$d_0 \triangleq \frac{1}{1-\gamma}. \tag{28}$$

In particular, the size of the distance-based CSIT allocation at any TX remains bounded in the extended model as the number of TX/RX pairs K increases:

$$\mathrm{s}(\mathbf{B}^{\mathrm{dist},(j)}) = O(1) \text{ as } K \text{ grows}, \qquad \forall j \in \{1,\ldots,K\}. \tag{29}$$

Proof: This result follows directly by observing that $\{\mathbf{B}^{\mathrm{dist},(j)}\}_{ki}$ is equal to zero if $\mathrm{d}(k,j) > d_0$. From the asumption of finite density, there are only a finite number of RXs fulfilling this condition such that the size of the CSIT allocation at TX j remains bounded as K increases. ■

This result is in stark contrast with the conventional CSIT allocation where the size $\mathrm{s}(\mathbf{B}^{\mathrm{conv},(j)})$ scales linearly with K. This corollary confirms the intuition that a CSIT-exchange restricted to a finite neighborhood is sufficient to achieve global coordination.

We have considered so far a scenario with global sharing of the user's data symbols. The result above leads to ask ourselves whether the sharing of the user's data could also be reduced to a local neighborhood without reduction of the DoF achieved.

Corollary 2. *Let us denote by $\mathcal{K}_j$ the set of user's data symbols being shared to TX j. It is sufficient for achieving the maximal DoF that $s_i \in \mathcal{K}_j$ if*

$$\mathrm{d}(i,j) < d_0. \tag{30}$$

In the extended model, it follows that

$$\forall j,\ |\mathcal{K}_j| = O(1), \text{ as } K \text{ grows}. \tag{31}$$

Proof: From the proof of Theorem 1 in Appendix III, it holds that

$$\mathrm{E}[|\boldsymbol{e}_j^{\mathrm{H}}\mathbf{H}^{-1}\boldsymbol{e}_i|^2] \dot{\leq} (\mu^2)^{\mathrm{d}(i,j)}. \tag{32}$$

Hence, if (27) is fulfilled, $(\mu^2)^{\mathrm{d}(i,j)}P = P^{1+(\gamma-1)\,\mathrm{d}(i,j)} \dot{\leq} 1$ and setting $\{\mathbf{T}\}_{ji} = 0$ does not impact the DoF. Thus, TX j does not need to participate to the transmission of the ith stream and does not need to receive data symbol s_i. ■

The operational meaning of the above result is that no exchange of information (CSI or data symbol) is necessary between two TXs i and k if they verify that $\mathrm{d}(i,k) > d_0$. Intuitively, d_0 is the size of the neighborhood inside which the cooperation should occur. Altogether, the distance-based CSIT allocation along with the matching users data sharing provides an attractive alternative to clustering. The difference being that the hard-boundaries of the cluster are replaced by a smooth decrease of the level of cooperation.

V. SIMULATIONS

We verify now by simulations that the maximal DoF per user is achieved by the distance based CSIT allocation. At the same time, we compare the distance based CSIT allocation to the CSI disseminations commonly used, i.e., uniform CSIT allocation and clustering.

We consider a channel model with $\gamma = 0.6$ and we use Monte-Carlo averaging over 1000 channel realizations. We study first a network with a regular geometry where $K = 36$ TX/RX pairs are placed at the integer values inside a square of dimensions 6×6. We show in Fig. 1 the average rate achieved with different CSIT allocation policies. Specifically, the distance-based CSIT allocation in (26) is compared to two alternative CSIT allocations, being the uniform CSIT allocation $\{\mathbf{B}^{\mathrm{unif},(j)}\}_{j=1}^K$ where the bits are allocated *uniformly* to the TXs, and the clustering one $\{\mathbf{B}^{\mathrm{cluster},(j)}\}_{j=1}^K$ in which (non-overlapping) *regular clustering of size* 4 is used. Both CSIT allocations are chosen to have the same size as the distance-based one:

$$\mathrm{s}\left(\{\mathbf{B}^{\mathrm{unif},(j)}\}_{j=1}^K\right) = \mathrm{s}\left(\{\mathbf{B}^{\mathrm{cluster},(j)}\}_{j=1}^K\right) = \mathrm{s}\left(\{\mathbf{B}^{(\mathrm{dist},(j)}\}_{j=1}^K\right). \tag{33}$$

With these parameters, the size of the distance based CSIT allocation is only equal to 6.5% of the size of the conventional CSIT allocation. Nevertheless, it can be observed to achieve the maximal generalized DoF while the clustering solution has a smaller slope, yet larger than the uniform CSIT allocation. The distance-based CSIT allocation suffers from a strong negative rate

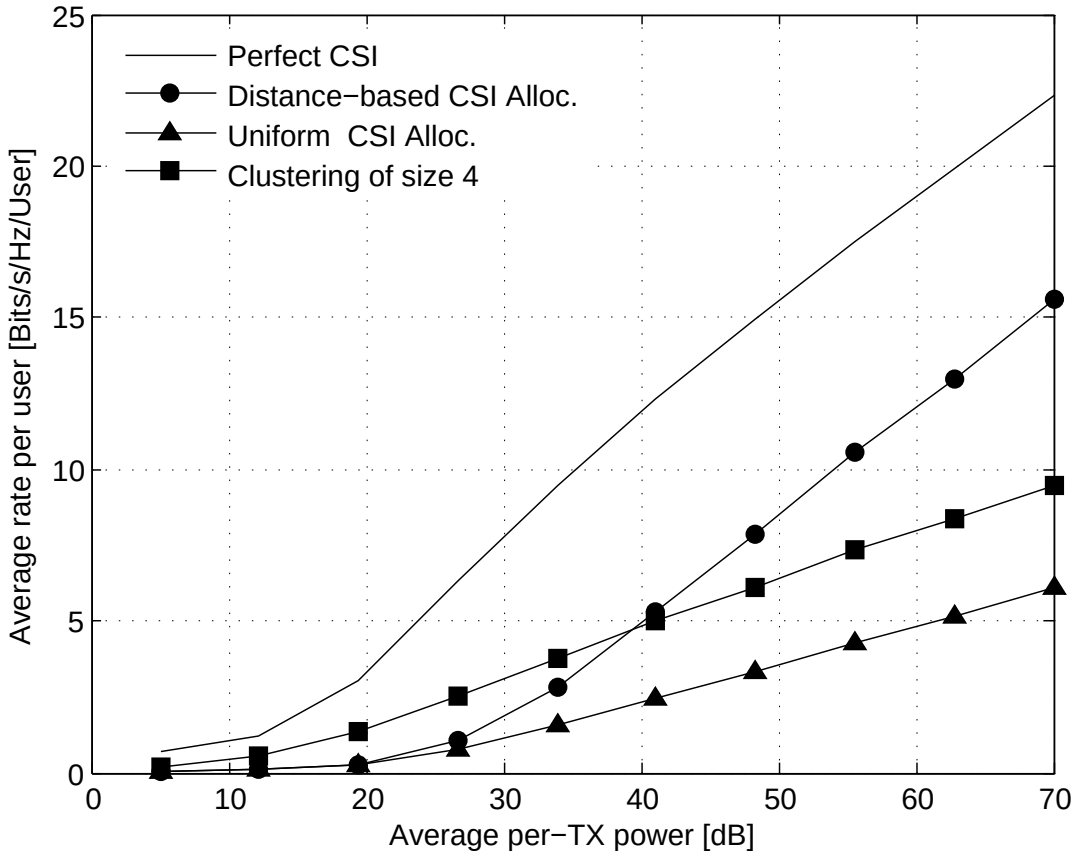

Fig. 1. Average rate per user as a function of the SNR P for $K = 36$ and $\gamma = 0.6$. The TX/RX pairs are positioned at the integers values inside a square of dimensions 6×6. The 3 limited feedback CSIT allocations used have the same size which is equal to 6.5% of the size of the conventional CSIT allocation in (25).

offset. This offset is a consequence of our analysis being limited to the high SNR regime. Indeed, using ZF with many users is very inefficient at intermediate SNR, particularly in a network with strong pathloss. Furthermore, the number of TX/RX pairs K which is here relatively large, has not been taken into account. Hence, this strong negative rate offset can be easily reduced by optimizing the precoding scheme and the CSIT allocation at finite SNR. The key element being that the distance-based CSIT allocation does not present the usual limitations of clustering, i.e., edge-interference and bad scaling properties as the size of the cluster increases.

Finally, we show in Fig. 2 the average rate per user in a network made of $K = 15$ TX/RX pairs being located uniformly at random over a square of dimensions 6×6 for $\gamma = 0.7$. We compare

the average rate achieved with the distance-based CSIT allocation to the average rate obtained if we use the distance-based allocation given in Theorem 1, but with $1+\alpha(\gamma-1)(\mathrm{d}(j,k)+\mathrm{d}(k,i))$ for a given parameter $\alpha > 0$ instead of $1 + (\gamma - 1)(\mathrm{d}(j,k) + \mathrm{d}(k,i))$. This allows to observe the impact of reducing ($\alpha > 1$) or increasing ($\alpha < 1$) the CSIT compared to the distance-based CSIT allocation.

We can observe that reducing the CSIT allocation leads to reducing the slope, i.e., the DoF, while using more feedback bits leads to a vanishing rate offset. This is in agreement with our theoretical result that the distance-based CSIT allocation leads to a finite rate offset.

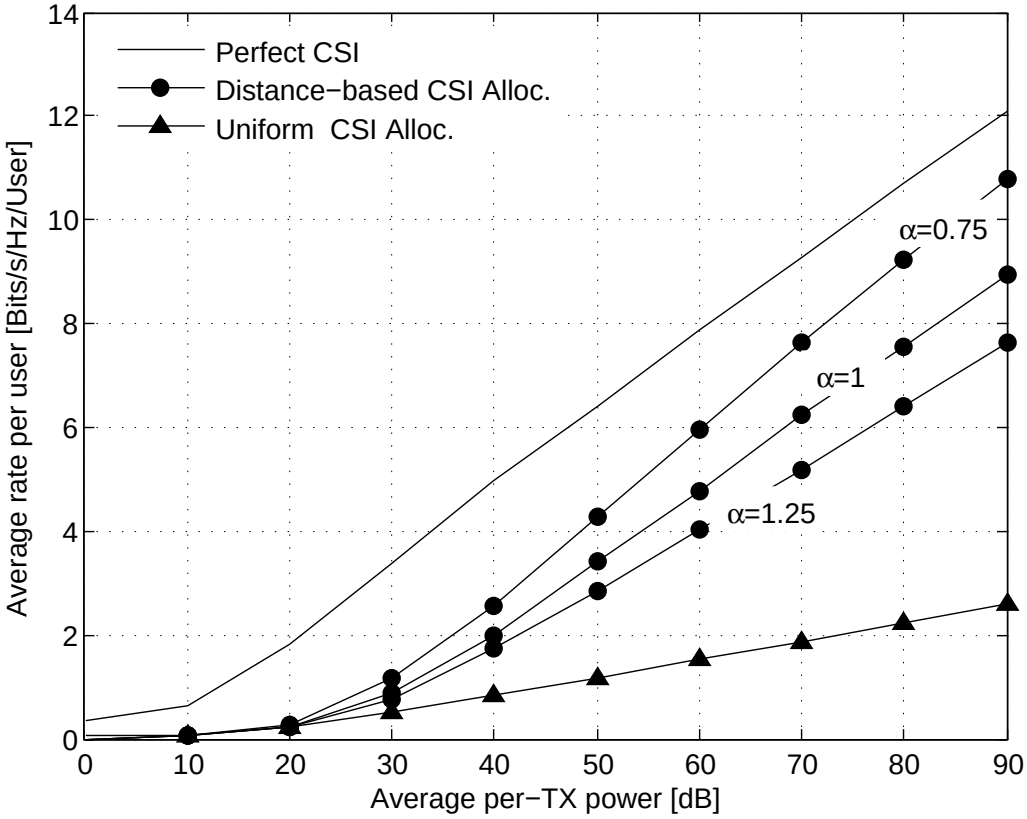

Fig. 2. Average rate per user as a function of the SNR P for $K = 15$ and $\gamma = 0.7$. The TX/RX pairs are placed uniformly at random over a square of dimensions 6×6. The CSIT allocation with $\alpha = 1.25$, $\alpha = 1$, and $\alpha = 1.25$, use respectively 10%, 17%, and 30% of the number of bits relative to the conventional CSIT allocation.

VI. Conclusion

We have discussed the problem of optimizing the CSIT dissemination in a network MIMO scenario. In particular, following a generalized DoF analysis, we have exhibited a CSIT allocation which allows to achieve the optimal generalized DoF while restricting the cooperation to a local scale. This behavior is critical for the cooperation of a large number of TXs to be practical. Hence, the proposed CSIT allocation appears as an alternative to clustering where the hard boundaries of the cluster are replaced by a smooth decrease of the cooperation strength. Our focus has been on the high SNR performance, and the distance-based CSIT allocation should be further optimized to lead to gain in realistic transmissions. Yet, it appears to have a strong potential as an alternative to clustering. We have also considered only the CSIT requirements in order to achieve global interference management. The design of feedback schemes and backhaul links allowing to achieve these requirements represents another very interesting research area.

This work shows that the CSIT requirements do not have to scale unbounded with the size of the network, which differs from the conclusion of several works from the literature. This is a consequence from letting the pathloss increase with the SNR, which makes the pathloss non-negligible at high SNR. We believe that this is the proper modelization of the pathloss in order to keep the impact of the network geometry, which, in contrast, becomes negligible in a DoF analysis with fixed pathloss.

Appendix I
Proof of Proposition 1

Proof: We start by defining the rate difference $\Delta_{\mathrm{R},i}$ between the rate of user i based on perfect CSI and the rate achieved with limited feedback. As in [30], [31], we can then write

$$\Delta_{\mathrm{R},i} \triangleq \mathrm{E}\left[\log_2(1+|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_i^{\star}|^2)\right] - \mathrm{E}\left[\log_2\left(1+\frac{|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_i|^2}{1+\sum_{j\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_j|^2}\right)\right] \tag{34}$$

$$= \mathrm{E}\left[\log_2\left(\frac{1+|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_i^{\star}|^2}{1+\sum_{j=1}^{K}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_j|^2}\right)\right] + \mathrm{E}\left[\log_2\left(1+\sum_{j\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_j|^2\right)\right] \tag{35}$$

$$\doteq \mathrm{E}\left[\log_2\left(1+\sum_{j\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_j|^2\right)\right] \tag{36}$$

where we have denoted by $\boldsymbol{t}_i^\star$ the ith ZF beamformer based on perfect CSIT. We further obtain

$$\Delta_{\mathrm{R},i} \dot{\leq} \mathrm{E}\left[\log_2\left(1+\sum_{j\neq i}|\boldsymbol{h}_i^{\mathrm{H}}(\boldsymbol{t}_j^\star+(\boldsymbol{t}_j-\boldsymbol{t}_j^\star))|^2\right)\right] \tag{37}$$

$$= \mathrm{E}\left[\log_2\left(1+\sum_{j\neq i}(|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_j^\star|^2+|\boldsymbol{h}_i^{\mathrm{H}}(\boldsymbol{t}_j-\boldsymbol{t}_j^\star)|^2)\right)\right] \tag{38}$$

$$= \mathrm{E}\left[\log_2\left(1+\sum_{j\neq i}|\boldsymbol{h}_i^{\mathrm{H}}(\boldsymbol{t}_j-\boldsymbol{t}_j^\star)|^2)\right)\right]. \tag{39}$$

We can then easily upper-bound (39) to write

$$\Delta_{\mathrm{R},i} \dot{\leq} \mathrm{E}\left[\log_2\left(1+\|\boldsymbol{h}_i\|^2\sum_{j\neq i}\|\boldsymbol{t}_j-\boldsymbol{t}_j^\star\|^2\right)\right] \tag{40}$$

$$= \mathrm{E}\left[\log_2\left(1+\|\mathbf{T}-\mathbf{T}^\star\|_{\mathrm{F}}^2\right)\right]+\mathrm{E}\left[\log_2\left(1+\|\boldsymbol{h}_i\|^2\right)\right] \tag{41}$$

$$\dot{\leq} \log_2\left(\mathrm{E}\left[\|\mathbf{T}-\mathbf{T}^\star\|_{\mathrm{F}}^2\right]\right). \tag{42}$$

The maximal DoF is achieved if the rate difference $\Delta_{\mathrm{R},i}/\log_2(P)$ tends to zero as the SNR increases, which is the result of the proposition. ■

APPENDIX II
PROOF OF PROPOSITION 2

Proof: We consider without loss of generality the precoding at TX j. Using the CSIT allocation in (19), it holds that

$$\sigma_{ki}\sqrt{2^{-B_{ki}^{(j)}}} = \sqrt{\frac{1}{P}} \tag{43}$$

such that $\mathbf{H}^{(j)} = \mathbf{H} + \sqrt{\frac{1}{P}}\boldsymbol{\Delta}\mathbf{H}^{(j)}$. We start by recalling the well known resolvent equality.

Proposition 3 (Resolvent equality)**.** *Let* $\mathbf{A}\in\mathbb{C}^{n\times n}$ *and* $\mathbf{B}\in\mathbb{C}^{n\times n}$ *be two invertible matrices, it then holds that*

$$\mathbf{A}^{-1}-\mathbf{B}^{-1} = \mathbf{B}^{-1}(\mathbf{B}-\mathbf{A})\mathbf{A}^{-1}. \tag{44}$$

Using the resolvent equality two times successively, we can then write

$$\left(\mathbf{H}^{(j)}\right)^{-1} - \mathbf{H}^{-1} = \left(\mathbf{H} + \sqrt{P^{-1}}\mathbf{\Delta H}^{(j)}\right)^{-1} - \mathbf{H}^{-1} \tag{45}$$

$$= \mathbf{H}^{-1}(-\sqrt{P^{-1}}\mathbf{\Delta H}^{(j)}) \left(\mathbf{H} + \sqrt{P^{-1}}\mathbf{\Delta H}^{(j)}\right)^{-1} \tag{46}$$

$$= \mathbf{H}^{-1}(-\sqrt{P^{-1}}\mathbf{\Delta H}^{(j)})\mathbf{H}^{-1} + \mathbf{H}^{-1}(-\sqrt{P^{-1}}\mathbf{\Delta H}^{(j)}) \left((\mathbf{H}^{(j)})^{-1} - \mathbf{H}^{-1}\right) \tag{47}$$

$$= -\sqrt{P^{-1}}\mathbf{H}^{-1}\mathbf{\Delta H}^{(j)}\mathbf{H}^{-1} + P^{-1}\mathbf{H}^{-1}\mathbf{\Delta H}^{(j)}\mathbf{H}^{-1}\mathbf{\Delta H}^{(j)}(\mathbf{H}^{(j)})^{-1}. \tag{48}$$

We can then use the triangular inequality to obtain the upperbound

$$\| \left(\mathbf{H}^{(j)}\right)^{-1} - \mathbf{H}^{-1}\|_{\mathrm{F}} \leq \sqrt{P^{-1}}\|\mathbf{H}^{-1}\|_{\mathrm{F}}^2\|\mathbf{\Delta H}^{(j)}\|_{\mathrm{F}} + P^{-1}\|(\mathbf{H}^{(j)})^{-1}\|_{\mathrm{F}}\|\mathbf{H}^{-1}\|_{\mathrm{F}}^2\|\mathbf{\Delta H}^{(j)}\|_{\mathrm{F}}^2 \tag{49}$$

$$\doteq \sqrt{P^{-1}}\|\mathbf{H}^{-1}\|_{\mathrm{F}}^2\|\mathbf{\Delta H}^{(j)}\|_{\mathrm{F}}. \tag{50}$$

Because of our assumption of outer-diagonal decay, the channel is well conditioned and all the expectations in (50) are finite. Taking the square and the expectation, we obtain then

$$\mathrm{E}[\| \left(\mathbf{H}^{(j)}\right)^{-1} - \mathbf{H}^{-1}\|_{\mathrm{F}}^2] \dot{\leq} P^{-1}. \tag{51}$$

After normalization, it follows easily from (51) that the sufficient condition in Proposition 1 is fulfilled, which concludes the proof. ■

APPENDIX III
PROOF OF THEOREM 1

Proof: Let us focus without loss of generality on the CSIT allocation at TX j. Following the sufficient condition in Proposition 1, we want to find a CSIT allocation such that

$$\mathrm{E}[|\boldsymbol{e}_j^{\mathrm{H}}\mathbf{H}^{-1}\boldsymbol{e}_i - \boldsymbol{e}_j^{\mathrm{H}} \left(\mathbf{H}^{(j)}\right)^{-1} \boldsymbol{e}_i|^2] \dot{\leq} P^{-1}, \qquad \forall i \in \{1, \ldots, K\}. \tag{52}$$

Indeed, once (52) is fulfilled, the the sufficient condition in Proposition 1 follows easily.

Let us first define the matrix $\mathbf{D} \triangleq \mathrm{diag}(\mathbf{H})$. It holds because of the outer-diagonal attenuation by $\mu^{\min_{i,j} \mathrm{d}(i,j)}$ that

$$\lim_{n\to\infty} (\mathbf{I}_K - \mathbf{D}^{-1}\mathbf{H})^n = \mathbf{0}_K. \tag{53}$$

Hence, we can proceed by writing the channel inverse using the series von Neumann as

$$\boldsymbol{e}_j^{\mathrm{H}}\mathbf{H}^{-1}\boldsymbol{e}_i = \boldsymbol{e}_j^{\mathrm{H}} \sum_{n=0}^{\infty} (\mathbf{D}^{-1}(\mathbf{D} - \mathbf{H}))^n \mathbf{D}^{-1}\boldsymbol{e}_i, \qquad \forall i, j \tag{54}$$

$$= \sum_{n=0}^{\infty} C_n^{ji} \tag{55}$$

where we have defined

$$C_n^{ji} \triangleq \boldsymbol{e}_j^{\mathrm{H}}(\mathbf{D}^{-1}(\mathbf{D}-\mathbf{H}))^n\mathbf{D}^{-1}\boldsymbol{e}_i, \qquad \forall i,j,n. \tag{56}$$

It can be seen from (56) that C_n^{ji} verifies

$$\mathrm{E}[|C_n^{ji}|^2] \dot{\leq} (\mu^{2\min_{i,j}\mathrm{d}(i,j)})^n \tag{57}$$

$$= P^{n\left(\frac{\log(\mu^{2\min_{i,j}\mathrm{d}(i,j)}P)-\log(P)}{\log(P)}\right)} \tag{58}$$

$$= P^{(\gamma_{\min}-1)n} \tag{59}$$

where we have defined

$$\gamma_{\min} \triangleq \frac{\log(\mu^{2\min_{i,j}\mathrm{d}(i,j)}P)}{\log(P)}. \tag{60}$$

Hence, the infinite summation can be truncated to a finite summation up to $n_0 \triangleq \lceil 1/(1-\gamma_{\min})\rceil$ without impacting the DoF. We further introduce

$$\mathbf{D}^{(j)} \triangleq \mathrm{diag}(\mathbf{H}^{(j)}), \qquad \forall j, \tag{61}$$

$$C_n^{ji,(j)} \triangleq \boldsymbol{e}_j^{\mathrm{H}}((\mathbf{D}^{(j)})^{-1}(\mathbf{D}^{(j)}-\mathbf{H}^{(j)}))^n(\mathbf{D}^{(j)})^{-1}\boldsymbol{e}_i, \qquad \forall i,j,n. \tag{62}$$

We can then write

$$\mathrm{E}[|\boldsymbol{e}_j^{\mathrm{H}}\mathbf{H}^{-1}\boldsymbol{e}_i - \boldsymbol{e}_j^{\mathrm{H}}\left(\mathbf{H}^{(j)}\right)^{-1}\boldsymbol{e}_i|^2] \doteq \mathrm{E}\left[\left|\sum_{n=1}^{n_0} C_n^{ji} - C_n^{ji,(j)}\right|^2\right] \tag{63}$$

$$\leq \mathrm{E}\left[\left(\sum_{n=1}^{n_0} |C_n^{ji} - C_n^{ji,(j)}|\right)^2\right] \tag{64}$$

$$\dot{\leq} \sum_{n=1}^{n_0} \mathrm{E}[|C_n^{ji} - C_n^{ji,(j)}|^2] \tag{65}$$

where we have used iteratively that $(a+b)^2 \leq 2(a^2+b^2), \forall a,b \in \mathbb{R}^2$ to obtain the last inequality (and the multiplicative constants could be removed because of the exponential inequality). We now look for a CSIT allocation $\mathbf{B}^{(j)}$ ensuring that

$$\mathrm{E}[|C_n^{ji} - C_n^{ji,(j)}|^2] \dot{\leq} P^{-1}, \qquad \forall i,j,n. \tag{66}$$

In particular, let us write the first coefficients

$$\mathrm{E}[|C_0^{ji} - C_0^{ji,(j)}|^2] = \mathrm{E}\left[\left|\frac{\boldsymbol{e}_j^{\mathrm{H}}\boldsymbol{e}_i}{H_{ii}} - \frac{\boldsymbol{e}_j^{\mathrm{H}}\boldsymbol{e}_i}{H_{ii}^{(j)}}\right|^2\right] \tag{67}$$

$$= \mathrm{E}\left[\left|\frac{\sigma_{ii}^{(j)}\Delta H_{ii}^{(j)}}{H_{ii}H_{ii}^{(j)}}\right|^2\right]\delta_{ji} \tag{68}$$

$$\doteq (\sigma_{ii}^{(j)})^2\delta_{ji}. \tag{69}$$

Thus, we set

$$B_{jj}^{(j)} = \lceil \log_2(P) \rceil. \tag{70}$$

This ensure to fulfill (66) for $n = 0$. The error done over H_{jj} becomes then negligible in terms of DoF (i.e., in terms of exponential equality). For $n \geq 1$, it holds that $C_n^{jj} = 0$ such that we assume that $i \neq j$ in the following,

$$\mathrm{E}[|C_1^{ji} - C_1^{ji,(j)}|^2] = \mathrm{E}\left[\left|\frac{\{\mathbf{D} - \mathbf{H}\}_{j,i}}{H_{jj}H_{ii}} - \frac{\{\mathbf{D}^{(j)} - \mathbf{H}^{(j)}\}_{j,i}}{H_{jj}^{(j)}H_{ii}^{(j)}}\right|^2\right] \tag{71}$$

$$\doteq \mathrm{E}\left[\left|\frac{H_{i,i}^{(j)}\tilde{H}_{j,i} - H_{i,i}\tilde{H}_{j,i}^{(j)}}{H_{jj}H_{ii}H_{ii}^{(j)}}\right|^2\right](\mu^2)^{\mathrm{d}(i,j)} \tag{72}$$

$$\doteq ((\sigma_{ii}^{(j)})^2 + (\sigma_{ji}^{(j)})^2)P^{(\gamma-1)\,\mathrm{d}(i,j)} \tag{73}$$

$$\doteq (2^{-B_{ii}^{(j)}} + 2^{-B_{ji}^{(j)}})P^{(\gamma-1)\,\mathrm{d}(i,j)} \tag{74}$$

Setting

$$B_{ii}^{(j)} = \lceil [1 + (\gamma - 1)\,\mathrm{d}(i,j)]^+ \log_2(P) \rceil, \qquad \forall i \neq j \tag{75}$$

$$B_{ji}^{(j)} = \lceil [1 + (\gamma - 1)\,\mathrm{d}(i,j)]^+ \log_2(P) \rceil, \qquad \forall i \neq j \tag{76}$$

ensures to fulfill (66) for all streams i for $n=1$. Going further, we consider then C_2^{ji},

$$\mathrm{E}[|C_2^{ji} - C_2^{ji,(j)}|^2] \tag{77}$$

$$= \mathrm{E}[|\boldsymbol{e}_j^{\mathrm{H}}(\mathbf{D}^{-1}(\mathbf{D}-\mathbf{H}))^2\mathbf{D}^{-1}\boldsymbol{e}_i - \boldsymbol{e}_j^{\mathrm{H}}((\mathbf{D}^{(j)})^{-1}(\mathbf{D}^{(j)}-\mathbf{H}^{(j)}))^2(\mathbf{D}^{(j)})^{-1}\boldsymbol{e}_i|^2] \tag{78}$$

$$= \mathrm{E}[|\sum_{k=1}^{K} \boldsymbol{e}_j^{\mathrm{H}}\mathbf{D}^{-2}(\mathbf{D}-\mathbf{H})\boldsymbol{e}_k\boldsymbol{e}_k^{\mathrm{H}}(\mathbf{D}-\mathbf{H})\mathbf{D}^{-1}\boldsymbol{e}_i - \boldsymbol{e}_j^{\mathrm{H}}(\mathbf{D}^{(j)})^{-2}(\mathbf{D}^{(j)}-\mathbf{H}^{(j)})\boldsymbol{e}_k$$

$$\cdot\, \boldsymbol{e}_k^{\mathrm{H}}(\mathbf{D}^{(j)}-\mathbf{H}^{(j)})(\mathbf{D}^{(j)})^{-1}\boldsymbol{e}_i|^2] \tag{79}$$

$$= \mathrm{E}[|\sum_{k=1,k\neq i,k\neq j}^{K} \frac{1}{H_{jj}^2 H_{ii}}\tilde{H}_{jk}\tilde{H}_{ki} - \frac{1}{(H_{jj}^{(j)})^2 H_{ii}^{(j)}}\tilde{H}_{jk}^{(j)}\tilde{H}_{ki}^{(j)}|^2](\mu^2)^{\mathrm{d}(j,k)+\mathrm{d}(k,i)} \tag{80}$$

$$\dot{\leq} \mathrm{E}[|\sum_{k=1,k\neq i,k\neq j}^{K} \tilde{H}_{jk}\tilde{H}_{ki} - \tilde{H}_{jk}^{(j)}\tilde{H}_{ki}^{(j)}|^2]P^{(\gamma-1)(\mathrm{d}(j,k)+\mathrm{d}(k,i))} \tag{81}$$

$$\doteq \sum_{k=1,k\neq i,k\neq j}^{K} ((\sigma_{jk}^{(j)})^2 + (\sigma_{ki}^{(j)})^2)P^{(\gamma-1)(\mathrm{d}(j,k)+\mathrm{d}(k,i))} \tag{82}$$

$$\doteq \sum_{k=1,k\neq i,k\neq j}^{K} (\sigma_{ki}^{(j)})^2 P^{(\gamma-1)(\mathrm{d}(j,k)+\mathrm{d}(k,i))} \tag{83}$$

where we could remove $\sigma_{jk}^{(j)}$ because of (76). Setting

$$B_{ki}^{(j)} = \lceil [1+(\gamma-1)(\mathrm{d}(j,k)+\mathrm{d}(k,i))]^+ \log_2(P) \rceil, \qquad \forall k\neq i, k\neq j \tag{84}$$

allows to fulfill (66) for $n=2$. Going to arbitrary value of n, we write

$$\mathrm{E}[|C_n^{ji} - C_n^{ji,(j)}|^2] = \mathrm{E}\left[\left|\sum_{k_1\neq j}^{K}\sum_{k_2\neq k_1}^{K}\cdots\sum_{k_{n-1}\neq k_{n-2},k_{n-1}\neq i}^{K}\left(\frac{\tilde{H}_{j,k_1}\tilde{H}_{k_1,k_2}\ldots,\tilde{H}_{k_{n-1},i}}{H_{jj}^n H_{ii}}\right.\right.\right.$$

$$\left.\left.\left. - \frac{\tilde{H}_{j,k_1}^{(j)}\tilde{H}_{k_1,k_2}^{(j)}\ldots,\tilde{H}_{k_{n-1},i}^{(j)}}{(H_{jj}^{(j)})^n H_{ii}^{(j)}}\right)\right|^2\right](\mu^2)^{\mathrm{d}(j,k_1)+\mathrm{d}(k_1,k_2)+\ldots+\mathrm{d}(k_{n-1},i)}. \tag{85}$$

Yet, it follows from the distance properties of d (triangular inequality, positivity) that the exponents in μ will always be larger than the ones obtained for $C_0^{ji}, C_1^{ji}, C_2^{ji}$. Hence, setting the $B_{ki}^{(j)}$ as in (70), (75),(76) and (84) ensures that (66) is fulfilled for all n and for all i. Inserting this result in (65) concludes the proof. ■

REFERENCES

[1] P. de Kerret and D. Gesbert, "CSI feedback allocation in multicell MIMO channels," in *Proc. International Conference on Communications (ICC)*, 2012.

[2] D. Gesbert, S. Hanly, H. Huang, S. Shamai (Shitz), O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: a new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[3] A. Papadogiannis, D. Gesbert, and E. Hardouin, "A dynamic clustering approach in wireless networks with multi-cell cooperative processing," in *Proc. International Conference on Communications (ICC)*, 2008.

[4] A. Papadogiannis and G. C. Alexandropoulos, "The value of dynamic clustering of base stations for future wireless networks," in *Proc. IEEE International Conference on Fuzzy Systems (FUZZ)*, 2010.

[5] J. Gong, S. Zhou, Z. Niu, L. Geng, and M. Zheng, "Joint scheduling and dynamic clustering in downlink cellular networks," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2011.

[6] A. Giovanidis, J. Krolikowski, and S. Brueck, "A 0-1 program to form minimum cost clusters in the downlink of cooperating base stations," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, 2012.

[7] I. Bergel, D. Yellin, and S. Shamai (Shitz), "Linear precoding bounds for Wyner-type cellular networks with limited base-station cooperation and dynamic clustering," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3714–3725, Jul. 2012.

[8] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.

[9] A. Lozano, R. W. Heath, and J. G. Andrews, "Fundamental limits of cooperation," *Information Theory, IEEE Transactions on*, vol. 59, no. 9, pp. 5213–5226, 2013.

[10] J. Zhang and J. G. Andrews, "Adaptive spatial intercell interference cancellation in multicell wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1455–1468, Dec. 2010.

[11] B. Özbek and D. Le Ruyet, "Adaptive limited feedback for intercell interference cancelation in cooperative downlink multicell networks," in *Proc. IEEE International Symposium on Wireless Communication Systems (ISWCS)*, 2010.

[12] H. A. A. Saleh and S. D. Blostein, "Single-cell vs. multicell MIMO downlink signalling strategies with imperfect CSI," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2010.

[13] W. W. L. Ho, T. Q. S. Quek, S. Sun, and R. W. Heath, "Decentralized precoding for multicell MIMO downlink," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1798–1809, Jun. 2011.

[14] R. Bhagavatula and R. W. Heath, "Adaptive limited feedback for sum-rate maximizing beamforming in cooperative multicell systems," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 800–811, Feb. 2011.

[15] A. Tajer and X. Wang, "Information exchange limits in cooperative MIMO networks," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2927–2942, Jun. 2011.

[16] B. Khoshnevis, W. Yu, and Y. Lostanlen, "Two-stage channel feedback for beamforming and scheduling in network MIMO systems," in *Proc. International Conference on Communications (ICC)*, 2012.

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

[18] D. Tuninetti, "On interFerence channel with generalized feedback (IFC-GF)," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2007.

[19] W. Wu, S. Vishwanath, and A. Arapostathis, "Capacity of a class of cognitive radio channels: Interference channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, Nov. 2007.

[20] S. A. Jafar and S. Vishwanath, "Generalized degrees of freedom of the symmetric Gaussian K user interference channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3297–3303, Jul. 2010.

[21] P. Mohapatra and C. Murthy, "Inner bound on the GDOF of the K-user MIMO Gaussian symmetric interference channel," *IEEE Trans. Commun.*, vol. PP, no. 99, pp. 1–10, 2012.

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

[23] A. Chaaban and A. Sezgin, "On the generalized degrees of freedom of the Gaussian interference relay channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4432–4461, Jul. 2012.

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

[25] P. de Kerret, J. Hoydis, and D. Gesbert, "Rate loss analysis of transmitter cooperation with distributed CSIT," in *Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2013.

[26] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theo.*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[27] R. Zakhour and D. Gesbert, "Team decision for the cooperative MIMO channel with imperfect CSIT sharing," in *Proc. Information Theory and Applications Workshop (ITA)*, 2010.

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

[29] B. Girod. (2013, Aug.) Lecture on image and video compression (EE398A). [Online]. Available: http://www.stanford.edu/class/ee398a/handouts/lectures/05-Quantization.pdf

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

[31] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.

[32] D. Samardzija and N. Mandayam, "Unquantized and uncoded channel state information feedback in multiple-antenna multiuser systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1335–1345, Jul. 2006.

[33] P. de Kerret and D. Gesbert, "CSI sharing strategies for transmitter cooperation in wireless networks," *IEEE Wireless Commun. Mag.*, vol. 20, no. 1, pp. 43–49, Feb. 2013.

[34] G. Caire, N. Jindal, and S. Shamai (Shitz), "On the required accuracy of transmitter channel state information in multiple antenna broadcast channels," in *Proc. IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC)*, 2007.

[35] S. Jaffard, "Decay rates for inverses of band matrices," *Gauthiers-Villars, Annales de l'I.H.P., section C*, vol. 7, no. 5, 1990.

[36] K. Groechenig and M. Leinert, "Symmetry and inverse-closedness of matrix algebras and functional calculus for infinite matrices," *Trans. of the American Mathematical Society*, Jan 2006.

[37] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communication: scaling laws and optimal operation," *IEEE Trans. Inf. Theo.*, vol. 50, no. 5, pp. 748–767, May 2004.

[38] A. Ozgur, O. Leveque, and D. N. C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in Ad Hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3549–3572, Oct. 2007.