

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 work¹, 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 generalized 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 at finite SNR with only local cooperation.

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.

As a consequence, there has been a large literature dealing with the joint precoding across TXs based on limited backhaul (See [3]–[7] and references therein). In [5], distributed schemes based on iterative updates of the transmit coefficients

were designed to avoid the requirements of explicit CSI at the TXs. However, this approach cannot be applied in many delay-limited scenarios as the iterations introduce significant delay. In [4], [7], considering centralized precoding, the impact of limited backhaul links between the central node and the TXs is discussed. In [3], [6], the problem of the sharing of the users data symbols is studied with the assumption of perfect CSI at all TXs.

However, when considering the imperfect sharing of CSI in distributed precoding, it is always assumed that all the TXs designing jointly the precoder have the *same* CSI. This means that practical schemes to reduce the sharing of the CSI with distributed precoding comes down to clustering with the CSI exchange being limited to small *cooperation clusters* inside which the TXs cooperate. The optimal way of forming these clusters has recently become an active research topic [8]–[12]. 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., [13], [14]. 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 [14] 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 [14] 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

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

[15]–[21].

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 [22]–[28]. 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 [29], [30], 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 [29] 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 *distance-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.

Notations: We denote by e_i the i -th column of the $K \times K$ identity matrix, by \bullet^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 f and g be two functions taking their value in \mathbb{R} with the function g taking only non-zero values. We write $f(x) = o(g(x))$ to denote that $\lim_{x \rightarrow \infty} \frac{|f(x)|}{|g(x)|} = 0$. We also use the exponential equality $f(x) \doteq x^b$ as in [31] to denote that $\lim_{x \rightarrow \infty} \frac{\log(f(x))}{\log(x)} = b$.

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. 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

$$\begin{bmatrix} y_1 \\ \vdots \\ y_K \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1^H \\ \vdots \\ \mathbf{h}_K^H \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_K \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_K \end{bmatrix} \quad (1)$$

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

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

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

$$\mathbf{x} = [\mathbf{t}_1 \quad \dots \quad \mathbf{t}_K] \cdot \begin{bmatrix} s_1 \\ \vdots \\ s_K \end{bmatrix}. \quad (2)$$

Hence, the vector $\mathbf{t}_i \in \mathbb{C}^{K \times 1}$ represents the beamforming vector used to transmit s_i to RX i and we define as $\mathbf{T} \triangleq [\mathbf{t}_1, \dots, \mathbf{t}_K] \in \mathbb{C}^{K \times K}$ 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 written as

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

The *DoF* at RX i is defined as commonly used in the literature as [32]

$$\text{DoF}_i \triangleq \lim_{P \rightarrow \infty} \frac{R_i}{\log_2(P)}. \quad (4)$$

Note that the channel elements are all non-zero with probability one. Consider only these channel realizations does not reduce the DoF achieved but has for consequence that all the expectations can be shown to exist. For the sake of clarity, we consider then only these channel realizations in the following. A more detailed explanation can be found in Appendix I.

B. Generalized DoF Analysis

However, we aim in this work at studying the effect of the network topology in the high SNR performance. In the conventional DoF analysis, any finite pathloss difference is neglected which can lead in some cases to a significant gap between the predicted and the true performance at finite SNR. As a consequence, we use the notion of *generalized DoF* [22]–[28] as a way to obtain a more accurate modeling of the performance. In the generalized DoF approach, the attenuation

of the interference is represented as an exponential function of the transmit power so as to preserve the impact of the network geometry in the high SNR analysis. Hence, the generalized DoF at RX i is defined as

$$\text{DoF}_i(\{\mathbf{B}^{(j)}\}_{j=1}^K, \mathbf{\Gamma}) \triangleq \lim_{P \rightarrow \infty} \frac{R_i}{\log_2(P)} \quad (5)$$

subject to $\sigma_{k,i}^2 = P^{-\{\mathbf{\Gamma}\}_{k,i}}, \forall k, i$

where the CSIT allocation $\{\mathbf{B}^{(j)}\}_{j=1}^K$ and the precoding used will be described in the following and the matrix $\mathbf{\Gamma} \in [0, \infty]^{K \times K}$ is called the *interference level matrix* and is given as a function of the parameters of the practical network being studied. Its (k, i) -th element is then denoted by $\Gamma_{k,i}$. More specifically, our goal is to model the transmission in a network with finite pathloss and finite transmit SNR in the most accurate possible way, and the interference level matrix makes the link between the practical network and the model obtained. Denoting by P_0 the finite power used in practice and by $\sigma_{k,i,0}$ the variance of the channel coefficient between TX i and RX k in the practical setting, the interference level matrix is then defined as

$$\Gamma_{k,i} \triangleq -\frac{\log(\sigma_{k,i,0}^2)}{\log(P_0)}, \quad \forall k, i. \quad (6)$$

Note that we assume that all the diagonal coefficients of the interference-level matrix are equal to zero, $\Gamma_{i,i} = 0, \forall i$.

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.

Remark 1. When P_0 – the transmit power used in the practical setting – tends to infinity, the prelog factor converges to the (conventional) DoF. The prelog factor converges to the generalized DoF in a different limiting regime where both the pathloss and the SNR increase at the same time. This could for example be the case if the transmit power is made dependent of the distance between the TXs and the RXs. Both approaches however can be used to approximate the performance in practical settings at finite (high) SNR. When there are significant pathloss differences, the generalized DoF will be more accurate. \square

Example 1. Let us consider as toy example a Gaussian IC with two TX/RX pairs and every node having a single-antenna. They interfere to each other via a channel of variance $\sigma^2 \in (0, 1)$ while the direct links have unit variance. The DoF is well known to be equal to 0.5 independently of the value of σ^2 [32]. However, if the TXs interfere with very low power, e.g., $\sigma^2 = 10^{-12}$, then the interference will be negligible for any realistic range of power used for the

transmission. Considering a transmission at SNR $P_0 = 30$ dB, the interfering coefficient as defined in [22] would be $\alpha = \max(\log(P_0\sigma^2)/\log(P_0), 0) = 0$ and the generalized DoF would then be equal to $1 - \alpha = 1$. Hence, the generalized DoF analysis models more accurately the transmission in that setting. \square

C. 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 [29], [33]. 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 collection 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

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

where $\Delta \tilde{H}_{k,i}^{(j)} \sim \mathcal{N}(0, 1)$ and the $\Delta \tilde{H}_{k,i}^{(j)}$ 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} [34, Theorem 13.3.3] while this distortion value is also achieved up to a multiplicative constant using for example the Lloyd algorithm [35] or even scalar quantization. Thus, the decay in 2^{-B} as the number of quantization 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 mild regularity constraints could be chosen. \square

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 [29], [36], [37]. 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 $s(\bullet)$ of a CSIT allocation $\mathbf{B}^{(j)}$ at TX j is defined as

$$s(\mathbf{B}^{(j)}) \triangleq \lim_{P \rightarrow \infty} \frac{\sum_{i,k} B_{k,i}^{(j)}}{\log_2(P)} \quad (8)$$

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

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

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

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. Indeed, digital quantization is simply used as a way to quantify the variance of the CSIT errors [37], [38]. 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, ...) [39]. \square

D. 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 $\mathbf{t}_i^{(j)}$ to transmit symbol s_i such that

$$\mathbf{t}_i^{(j)} \triangleq \sqrt{P} \frac{(\mathbf{H}^{(j)})^{-1} \mathbf{e}_i}{\|(\mathbf{H}^{(j)})^{-1} \mathbf{e}_i\|}, \quad \forall i \in \{1, \dots, K\}. \quad (11)$$

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 [36], [37]. Furthermore, considering limited feedback in the compound MIMO BC, it is revealed in [40] 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. \square

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 $x_j = \mathbf{e}_j^H \mathbf{T}^{(j)} \mathbf{s}$. The effective multi-user precoder \mathbf{T} verifies then

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

Remark 5. Each TX independently proceeds with the normalization of the beamformer 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. \square

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

$$\mathbf{t}_i^* \triangleq \sqrt{P} \frac{(\mathbf{H})^{-1} \mathbf{e}_i}{\|(\mathbf{H})^{-1} \mathbf{e}_i\|}, \quad \forall i \in \{1, \dots, K\}. \quad (13)$$

E. 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 $\mathbb{B}_{\text{DoF}}(\mathbf{\Gamma})$ as

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

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:

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

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

$$\mathbb{E} \left[\|\mathbf{T} - \mathbf{T}^*\|_{\text{F}}^2 \right] \doteq P^0 \quad (16)$$

where the equivalence $\text{sign } f(P) \doteq P^b$ denotes the exponential equality $\lim_{P \rightarrow \infty} \frac{\log(f(P))}{\log(P)} = b$ [31] and \mathbf{T}^* has been defined previously as the precoder based on perfect CSIT.

Proof. A detailed proof is provided in Appendix II. \square

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

$$\mathbb{E} \left[\left\| \mathbf{T}^{(j)} - \mathbf{T}^* \right\|_{\mathbb{F}}^2 \right] \doteq P^0, \quad \forall j \in \{1, \dots, K\}. \quad (17)$$

Based on this, the following result is obtained.

Proposition 2. *Considering the generalized DoF model where $\sigma_{ki}^2 = P^{-\Gamma_{k,i}}, \forall k, i$, the following ‘‘conventional’’ CSIT allocation $\{\mathbf{B}^{\text{conv},(j)}\}_{j=1}^K$ such that*

$$\{\mathbf{B}^{\text{conv},(j)}\}_{k,i} = \left[\left[\log_2(P\sigma_{k,i}^2) \right]^+ \right], \quad \forall k, i, j \quad (18)$$

$$= \left[\left[1 - \Gamma_{k,i} \right]^+ \log_2(P) \right] \quad (19)$$

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

Proof. A detailed proof is provided in Appendix III. \square

This CSIT allocation provides to each TX the K channel vectors relative to the K RXs. The term $\Gamma_{k,i}$ corresponds to the variance of the channel element $H_{k,i}$ which is equal to $P^{-\Gamma_{k,i}}$, and follows from well known results of rate-distorsion theory. 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

$$\mathbb{E} \left[\left\| e_j^{\text{H}} (\mathbf{T}^{(j)} - \mathbf{T}^*) \right\|^2 \right] \doteq P^0, \quad \forall j \in \{1, \dots, K\}. \quad (20)$$

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 j th 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

A. Distance-based CSIT allocation

Before stating our main result, we first define the notion of shortest path which will be needed for the theorem.

Definition 4. *We define a path from TX j to RX k as the tuple (a_1, \dots, a_n) with $a_i \in \{1, \dots, K\}$ being the index of a TX/RX pair and $a_1 = j$ and $a_n = k$. Given the interference level matrix $\mathbf{\Gamma}$, the length $L(a_1, \dots, a_n)$ of a path is then given by*

$$L(a_1, \dots, a_n) = \sum_{i=1}^{n-1} \Gamma_{a_{i+1}, a_i}. \quad (21)$$

We can then define the shortest path from TX j to RX k , which we denote by $\Gamma_{j \rightarrow k}$, in the sense that

$$\Gamma_{j \rightarrow k} \triangleq \min_{(a_2, \dots, a_{n-1})} L(j, a_2, \dots, a_{n-1}, k). \quad (22)$$

Example 2. *Let us consider the shortest path $\Gamma_{1 \rightarrow 3}$ in a network with 3 TX/RX pairs. Keeping in mind that $\Gamma_{i,i} = 0, \forall i$, it is then simply equal to*

$$\Gamma_{1 \rightarrow 3} = \min(\Gamma_{3,2} + \Gamma_{2,1}, \Gamma_{3,1}). \quad (23)$$

\square

We can now state our first main result.

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

$$\{\mathbf{B}^{\text{dist},(j)}\}_{k,i} \triangleq \left[\left[1 - \Gamma_{k,i} - \gamma_{k,i}^{(j)} \right]^+ \log_2(P) \right], \quad \forall k, i, j \quad (24)$$

with

$$\gamma_{k,i}^{(j)} \triangleq \min \left(\Gamma_{k \rightarrow j}, \min_{\ell} \Gamma_{\ell \rightarrow i} + \Gamma_{j,\ell} \right) \quad (25)$$

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

Proof. A detailed proof is provided in Appendix IV. \square

If the interference level matrix $\mathbf{\Gamma}$ tends to the zero matrix $\mathbf{0}_K$, 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 (19). More generally, the distance-based CSIT allocation exploits the fact that if two TX/RX pairs interfere only through weak channels, they need to exchange a small amount of CSI.

Remark 6. If the interference level matrix is a symmetric matrix, $\gamma_{k,i}^{(j)}$ is then equal to

$$\gamma_{k,i}^{(j)} = \min(\Gamma_{k \rightarrow j}, \Gamma_{i \rightarrow j}) \quad (26)$$

Another interesting case arises if the interference level matrix satisfies that

$$\Gamma_{k,j} \leq \Gamma_{k,i} + \Gamma_{i,j}, \quad \forall i, j, k \quad (27)$$

It then holds that

$$\Gamma_{k \rightarrow j} = \Gamma_{j,k}. \quad (28)$$

In particular, the relation (27) is satisfied when long term attenuation introduces a notion of distance between the TX/RX pairs. \square

Building upon the proof of Theorem 1, it is also possible to comment equation (24) to obtain interesting insights:

- The term $\Gamma_{k,i}$ follows from the variance of the element to quantize and is also present in the conventional CSIT allocation. Hence, the CSIT reduction comes from the parameter $\gamma_{k,i}^{(j)}$.
- The first term $\Gamma_{k \rightarrow j}$ corresponds to a sufficient CSIT allocation such that $e_j^{\text{H}} \mathbf{H}^{-1} e_i$ is known at TX j with a sufficient accuracy for every i .
- With the second term $\min_{\ell} \Gamma_{\ell \rightarrow i} + \Gamma_{j,\ell}$, the CSIT allocation obtained ensures that the norm $\|\mathbf{H}^{-1} e_i\|$ is known for every i with a sufficient accuracy at TX j .

Taken together (which explains the min in (25)), these two requirements allow to compute $e_j^H \mathbf{H}^{-1} e_i / \|\mathbf{H}^{-1} e_i\| = e_j^H \mathbf{t}_i$ at TX j for every i with an accuracy sufficient for achieving the maximal DoF at each RX.

B. Scaling properties of the Distance-based CSIT allocation

For clarity, we consider in the following that the interference level matrix is symmetric such that $\Gamma_{k,i} = \Gamma_{i,k}, \forall k, i$. It also corresponds to most of the practically relevant scenarios. Similar conclusions also hold in the asymmetric scenarios.

An important property of a CSIT allocation is its scaling behaviour as the number of TX/RX pairs increases. It is clear that this property depends on the network geometry. For example, if all the TX/RX pairs are collocated, full CSIT sharing is required while it may not be the case in typical network deployment scenarios, as shown previously. We focus now on the setting where each RX is “interfered” by only a finite number of TXs with an average power larger than $1/P$, who represent the significant interferers in terms of DoF. That is to say

$$\lim_{K \rightarrow \infty} |\{i | \Gamma_{i \rightarrow j} < 1\}| < \infty, \quad \forall j. \quad (29)$$

We will show in the following subsection that this condition is fulfilled in the networks practically encountered in wireless communications.

Corollary 1. *Let us consider a symmetric network where condition (29) is satisfied. It then holds that*

$$\lim_{K \rightarrow \infty} s(\mathbf{B}^{\text{dist},(j)}) < \infty, \quad \forall j. \quad (30)$$

Proof. This result follows directly from observing that $\Gamma_{k,i} \geq \Gamma_{i \rightarrow k}$ in the distance-based CSIT allocation in (24). It follows then trivially from (29) that there are only a finite number of nonzero $\{\mathbf{B}^{\text{dist},(j)}\}_{k,i}$ at every TX j . \square

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

Yet, we have considered so far a scenario with global sharing of the user’s data symbols. The result above leads to ask ourselves whether this assumption is necessary or if it is possible to reduce the sharing of the user’s data symbols without impacting the DoF achieved.

Corollary 2. *Let us consider a symmetric network where condition (29) is satisfied. Let us further denote by \mathcal{K}_j the set containing the user’s data symbols which have to be shared to TX j in order to achieve the maximal DoF at every RX. It then holds that*

$$\lim_{K \rightarrow \infty} |\mathcal{K}_j| < \infty, \quad \forall j \in \{1, \dots, K\}. \quad (31)$$

Proof. It can be seen from the expression of the precoder coefficients as an infinite summation in the proof of Theorem 1 in Appendix IV that

$$\mathbb{E}[|e_j^H \mathbf{H}^{-1} e_i|^2] \stackrel{\leq}{\sim} P^{-\Gamma_{j,i}}, \quad \forall i, j. \quad (32)$$

Setting to 0 all the coefficients in the precoder with an exponent of P smaller than -1 leads to additional interferences which tend to zero as the SNR increases and are hence negligible in terms of DoF. If the coefficient (j, i) is set to 0 at TX j , this means that TX j does not need to receive the user’s data symbol s_i . From the assumption (29), there are then only a finite number of user’s data symbols which need to be known at a given TX. \square

The operational meaning of the above result is that because of the pathloss attenuation it is possible to achieve DoF-perfect coordination on the TX side with each TX exchanging information (CSI or data symbol) only to a local neighborhood.

C. Scaling Behaviour in Wireless Networks

We have studied above the properties of the transmission for a given interference level matrix Γ . We will now show how these results can be used to model the transmission in realistic settings. The first step is to discuss how the interference-level matrix is obtained from the network configuration. Let us consider a network with a polynomial pathloss with exponent $\varepsilon > 0$, which corresponds to the conventional model for wireless networks [32]. This means that the long term attenuation between TX j and RX i is equal to $d_{i,j}^{-\varepsilon}$ where $d_{i,j}$ is the distance separating TX j and RX i . The operational SNR is set to be P_0 such that we obtain that

$$\Gamma_{i,j} = -\frac{\log(d_{i,j}^{-\varepsilon})}{\log(P_0)}, \quad \forall i, j. \quad (33)$$

Upon defining

$$d_0 \triangleq P_0^{\frac{1}{\varepsilon}}, \quad (34)$$

we can see that

$$\Gamma_{i,j} > 1, \quad \text{if } d_{i,j} > d_0. \quad (35)$$

We are particularly interested in this section in the scaling behaviour as the number of TX/RX pairs increases. This requires defining more precisely the spatial distribution of the TX/RX pairs. It is differentiated in the literature between so-called *dense* networks and *extended* networks [41], [42]. 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.

It follows from (35) and from the extended model that the necessary condition (29) is fulfilled in this network configuration. Hence, the corollaries provided above can be applied. Practically, this means that it is possible to achieve the performance of global cooperation with cooperation of the TXs restricted to a local scale. Altogether, the distance-based CSIT allocation along with the matching limited users data sharing provides an attractive alternative to clustering. The difference being that the hard-boundaries of the clusters 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 wireless model with polynomial attenuation as described in Subsection IV-C. We choose $\varepsilon = 2$, $P_0 = 30$ dB and the interference-level matrix Γ is obtained from (33). Note that we consider only $d_{i,j} > 1$ to ensure that the interfering links are weaker than the direct links. We use Monte-Carlo averaging over 1000 channel realizations.

In a first step, we study 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 (27) is compared to two alternative CSIT allocations, being the uniform CSIT allocation $\{\mathbf{B}^{\text{unif},(j)}\}_{j=1}^K$ where the bits are allocated *uniformly* to the TXs, and the clustering one $\{\mathbf{B}^{\text{cluster},(j)}\}_{j=1}^K$ 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:

$$s(\{\mathbf{B}^{\text{unif},(j)}\}_{j=1}^K) = s(\{\mathbf{B}^{\text{cluster},(j)}\}_{j=1}^K) = s(\{\mathbf{B}^{\text{dist},(j)}\}_{j=1}^K). \quad (36)$$

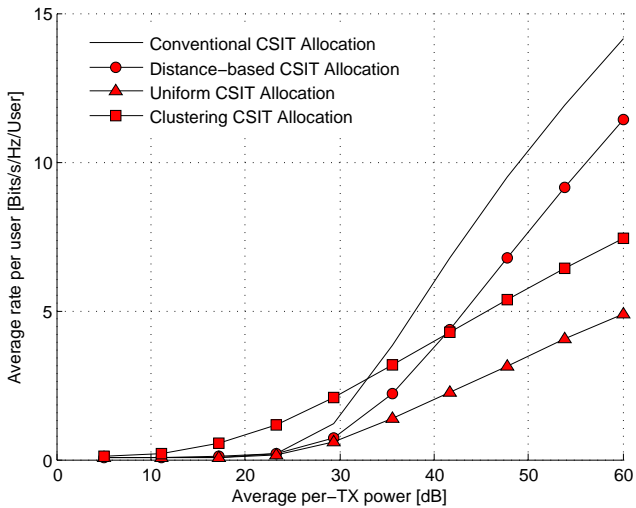


Fig. 1. Average rate per user as a function of the SNR P for $K = 36$ with the polynomial model described in Subsection IV-C. The TX/RX pairs are positioned at the integers values inside a square of dimensions 6×6 . The 3 limited feedback CSIT allocations used have the same size which is equal to 9% of the size of the conventional CSIT allocation in (19).

With these parameters, the size of the distance based CSIT allocation is only equal to 9% 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. The distance-based CSIT allocation suffers from a strong negative rate offset. However, this offset is a consequence of our analysis being limited to the high SNR regime and can be also observed in the fact that clustering outperforms ZF based on the conventional CSIT-allocation,

which represents in fact the true reference for our scheme. 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 the same square of dimensions 6×6 . To verify the impact of allocating more –or less– CSIT, we compare the average rate achieved with the distance-based CSIT allocation to the average rate obtained if we use the following variation of the CSIT allocation:

$$\{\mathbf{B}^{\text{dist},(j)}\}_{k,i}(\alpha) = \lceil [1 - \Gamma_{j,k} - \alpha \gamma_{k,i}^{(j)}]^+ \log_2(P) \rceil, \forall k, i, j. \quad (37)$$

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 (bounded) rate offset.

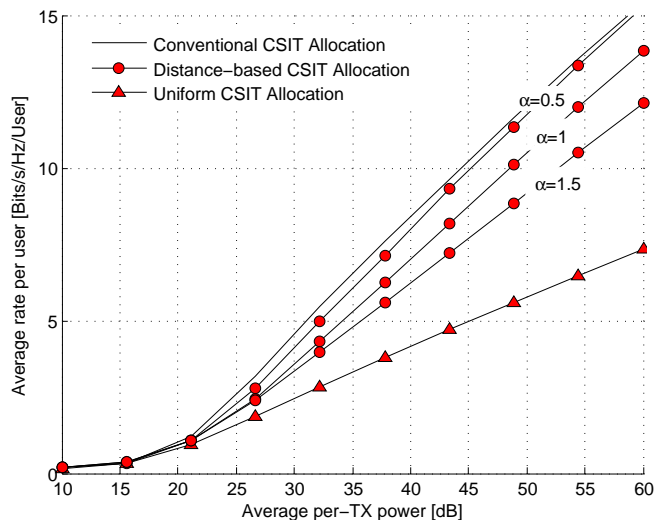


Fig. 2. Average rate per user as a function of the SNR P for $K = 15$ for the channel model described in Subsection IV-C. The TX/RX pairs are placed uniformly at random over a square of dimensions 6×6 . The CSIT allocation with $\alpha = 0.5$, $\alpha = 1$, and $\alpha = 1.5$, use respectively 39%, 17%, and 13% of the number of bits relative to the conventional CSIT allocation in (19).

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. In addition, we have 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 conclusions 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 PRELIMINARY REMARK

We start by a remark which will allow us to consider all the expectations of channel elements as finite. For a given $\varepsilon > 0$, let us consider the channels satisfying

$$\varepsilon < |H_{i,k}|^2 < \frac{1}{\varepsilon}, \quad \forall k, i \quad (38)$$

such that the expectations of fractions of channel coefficients will all be finite. The channel elements being Chi-2 distributed, their are distributed such that

$$\Pr\{|H_{i,k}|^2 < x\} = 1 - \exp\left(-\frac{x}{2}\right). \quad (39)$$

Hence, considering only the channels verifying (38) and assuming the DoF loss to be maximal for the other channel realizations (i.e., equal to 1 for each user) can be easily shown to lead to a DoF loss in the order of $(O(\varepsilon))$. Letting ε tend to zero, the $(O(\varepsilon))$ term vanishes. Hence, considering only the channels verifying the above condition does not modify the DoF and we will in the following consider only such channel realizations.

APPENDIX II PROOF OF PROPOSITION 1

Proof. We start by defining the rate difference $\Delta_{R,i}$ between the rate of user i based on perfect CSI and the rate achieved

with limited feedback. As in [36], [37], we can then write

$$\begin{aligned} \Delta_{R,i} &\triangleq \mathbb{E} \left[\log_2(1 + |\mathbf{h}_i^H \mathbf{t}_i^*|^2) \right] - \mathbb{E} \left[\log_2 \left(1 + \frac{|\mathbf{h}_i^H \mathbf{t}_i|^2}{1 + \sum_{j \neq i} |\mathbf{h}_i^H \mathbf{t}_j|^2} \right) \right] \\ &= \mathbb{E} \left[\log_2 \left(\frac{1 + |\mathbf{h}_i^H \mathbf{t}_i^*|^2}{1 + \sum_{j=1}^K |\mathbf{h}_i^H \mathbf{t}_j|^2} \right) \right] + \mathbb{E} \left[\log_2 \left(1 + \sum_{j \neq i} |\mathbf{h}_i^H \mathbf{t}_j|^2 \right) \right] \end{aligned} \quad (40)$$

$$= \mathbb{E} \left[\log_2 \left(1 + \sum_{j \neq i} |\mathbf{h}_i^H \mathbf{t}_j|^2 \right) \right] + o(\log(P)) \quad (41)$$

$$= \mathbb{E} \left[\log_2 \left(1 + \sum_{j \neq i} |\mathbf{h}_i^H \mathbf{t}_j|^2 \right) \right] + o(\log(P)) \quad (42)$$

where we have denoted by \mathbf{t}_i^* the i th ZF beamformer based on perfect CSIT. We further obtain

$$\Delta_{R,i} = \mathbb{E} \left[\log_2 \left(1 + \sum_{j \neq i} |\mathbf{h}_i^H (\mathbf{t}_j^* + (\mathbf{t}_j - \mathbf{t}_j^*))|^2 \right) \right] + o(\log(P)) \quad (43)$$

$$= \mathbb{E} \left[\log_2 \left(1 + \sum_{j \neq i} |\mathbf{h}_i^H (\mathbf{t}_j - \mathbf{t}_j^*)|^2 \right) \right] + o(\log(P)). \quad (44)$$

We can then easily upper-bound (44) to write

$$\Delta_{R,i} \leq \mathbb{E} \left[\log_2 \left(1 + \|\mathbf{h}_i\|^2 \sum_{j \neq i} \|\mathbf{t}_j - \mathbf{t}_j^*\|^2 \right) \right] + o(\log(P)) \quad (45)$$

$$\stackrel{(a)}{\leq} \mathbb{E} \left[\log_2 \left(1 + \|\mathbf{T} - \mathbf{T}^*\|_F^2 \right) \right] + \mathbb{E} \left[\log_2 \left(1 + \|\mathbf{h}_i\|^2 \right) \right] + o(\log(P)) \quad (46)$$

$$\stackrel{(b)}{\leq} \log_2 \left(\mathbb{E} \left[\|\mathbf{T} - \mathbf{T}^*\|_F^2 \right] \right) + o(\log(P)) \quad (47)$$

$$\leq o(\log(P)) \quad (48)$$

where inequality (a) follows from the property that for $a, b \geq 0$, then $\log(1+ab) \leq \log(1+a) + \log(1+b)$ and inequality (b) from the assumption that $\mathbb{E} \left[\|\mathbf{T} - \mathbf{T}^*\|_F^2 \right] \doteq P^0$. The maximal DoF is achieved if the rate difference $\Delta_{R,i}/\log_2(P)$ tends to zero as the SNR increases, which is exactly what has been demonstrated above. \square

APPENDIX III 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

$$\sigma_{k,i} \sqrt{2^{-B_{k,i}^{(j)}}} = \sqrt{\frac{1}{P}} \quad (49)$$

such that $\mathbf{H}^{(j)} = \mathbf{H} + \sqrt{\frac{1}{P}} \Delta \mathbf{H}^{(j)}$. 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*

$$\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{B}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{A}^{-1}. \quad (50)$$

Using the resolvent equality two times successively, we can write

$$\begin{aligned} & \left(\mathbf{H}^{(j)} \right)^{-1} - \mathbf{H}^{-1} \\ &= \left(\mathbf{H} + \sqrt{P^{-1}} \Delta \mathbf{H}^{(j)} \right)^{-1} - \mathbf{H}^{-1} \end{aligned} \quad (51)$$

$$= \mathbf{H}^{-1} \left(-\sqrt{P^{-1}} \Delta \mathbf{H}^{(j)} \right) \left(\mathbf{H}^{(j)} \right)^{-1} \quad (52)$$

$$\begin{aligned} &= \mathbf{H}^{-1} \left(-\sqrt{P^{-1}} \Delta \mathbf{H}^{(j)} \right) \mathbf{H}^{-1} \\ &\quad + \mathbf{H}^{-1} \left(-\sqrt{P^{-1}} \Delta \mathbf{H}^{(j)} \right) \left(\left(\mathbf{H}^{(j)} \right)^{-1} - \mathbf{H}^{-1} \right) \end{aligned} \quad (53)$$

$$\begin{aligned} &= -\sqrt{P^{-1}} \mathbf{H}^{-1} \Delta \mathbf{H}^{(j)} \mathbf{H}^{-1} \\ &\quad + P^{-1} \mathbf{H}^{-1} \Delta \mathbf{H}^{(j)} \mathbf{H}^{-1} \Delta \mathbf{H}^{(j)} \left(\mathbf{H}^{(j)} \right)^{-1}. \end{aligned} \quad (54)$$

We can then use the properties of the norm (matrix norm inequality and triangular inequality) to obtain the upperbound

$$\begin{aligned} \left\| \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i - \mathbf{H}^{-1} \mathbf{e}_i \right\| &\leq \sqrt{P^{-1}} \left\| \mathbf{H}^{-1} \right\|_{\mathbb{F}}^2 \left\| \Delta \mathbf{H}^{(j)} \right\|_{\mathbb{F}} \\ &\quad + P^{-1} \left\| \left(\mathbf{H}^{(j)} \right)^{-1} \right\|_{\mathbb{F}} \left\| \mathbf{H}^{-1} \right\|_{\mathbb{F}}^2 \left\| \Delta \mathbf{H}^{(j)} \right\|_{\mathbb{F}}^2 \end{aligned} \quad (55)$$

It follows then directly from the norm properties that

$$\left| \left\| \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i \right\| - \left\| \mathbf{H}^{-1} \mathbf{e}_i \right\| \right| \leq \left\| \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i - \mathbf{H}^{-1} \mathbf{e}_i \right\| \quad (56)$$

We can then use this result to write

$$\left\| \frac{\left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i}{\left\| \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i \right\|} - \frac{\mathbf{H}^{-1} \mathbf{e}_i}{\left\| \mathbf{H}^{-1} \mathbf{e}_i \right\|} \right\| \quad (57)$$

$$= \left\| \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i \right\| \left\| \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i - \frac{\left\| \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i \right\|}{\left\| \mathbf{H}^{-1} \mathbf{e}_i \right\|} \mathbf{H}^{-1} \mathbf{e}_i \right\| \quad (58)$$

$$\begin{aligned} &\leq \left\| \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i \right\| \left\| \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i - \mathbf{H}^{-1} \mathbf{e}_i \right\| \\ &\quad + \left\| \mathbf{H}^{-1} \mathbf{e}_i \right\| - \left\| \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i \right\| \end{aligned} \quad (59)$$

$$\leq \left(\left\| \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i \right\| + 1 \right) \left\| \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i - \mathbf{H}^{-1} \mathbf{e}_i \right\|. \quad (60)$$

Using (55), inside (60) yields

$$\begin{aligned} &\left\| \frac{\left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i}{\left\| \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i \right\|} - \frac{\mathbf{H}^{-1} \mathbf{e}_i}{\left\| \mathbf{H}^{-1} \mathbf{e}_i \right\|} \right\| \\ &\leq \sqrt{P^{-1}} \left\| \mathbf{H}^{-1} \right\|_{\mathbb{F}}^2 \left(\left\| \mathbf{H}^{-1} \right\|_{\mathbb{F}}^2 + 1 \right) \left\| \Delta \mathbf{H}^{(j)} \right\|_{\mathbb{F}} \\ &\quad + P^{-1} \left\| \left(\mathbf{H}^{(j)} \right)^{-1} \right\|_{\mathbb{F}} \left\| \mathbf{H}^{-1} \right\|_{\mathbb{F}}^2 \left(\left\| \mathbf{H}^{-1} \right\|_{\mathbb{F}}^2 + 1 \right) \left\| \Delta \mathbf{H}^{(j)} \right\|_{\mathbb{F}}^2. \end{aligned} \quad (61)$$

Because of our preliminary remark in Appendix I, we consider that all the expectations exist and are finite. Taking the square and the expectation, we obtain then

$$\mathbb{E} \left[\left\| \sqrt{P} \frac{\left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i}{\left\| \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i \right\|} - \sqrt{P} \frac{\mathbf{H}^{-1} \mathbf{e}_i}{\left\| \mathbf{H}^{-1} \mathbf{e}_i \right\|} \right\|^2 \right] \leq P^0. \quad (62)$$

Applying Proposition 1 concludes the proof. \square

APPENDIX IV PROOF OF THEOREM 1

Proof. Let us focus now without loss of generality on the CSIT allocation at TX j . Following the sufficient condition in Proposition 1, the maximal DoF is achieved at every RX if

$$\mathbb{E} \left[\left| \frac{\mathbf{e}_j^H \mathbf{H}^{-1} \mathbf{e}_i}{\left\| \mathbf{H}^{-1} \mathbf{e}_i \right\|} - \frac{\mathbf{e}_j^H \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i}{\left\| \left(\mathbf{H}^{(j)} \right)^{-1} \mathbf{e}_i \right\|} \right|^2 \right] \leq P^{-1}, \quad \forall i. \quad (63)$$

Hence, we will in the following show that the CSIT allocation given in Theorem 1 ensures that (63) is satisfied.

Following a similar calculation as in (60) in Appendix III, we obtain the relation given in (64) at the top of next page. We will bound the first term of (64) in Subsection IV-A and the second term in Subsection IV-B.

Our calculations rely on the following series expansion for the channel inverse. It can be seen that the outer-diagonal of \mathbf{H} is at least equal to $P^{-\min_{i \neq j} \Gamma_{i,j}}$. Upon defining the matrix $\mathbf{D} \triangleq \text{diag}(\mathbf{H})$, this means that we have for P large enough

$$\left\| \mathbf{I}_K - \mathbf{D}^{-1} \mathbf{H} \right\|_{\mathbb{F}} < 1. \quad (65)$$

It follows that we can define the geometric sum of the matrices as

$$\mathbf{S} \triangleq \sum_{n=0}^{\infty} \left(\mathbf{I}_K - \mathbf{D}^{-1} \mathbf{H} \right)^n. \quad (66)$$

Studying the partial sums of \mathbf{S} , the following well known result can be shown

$$\left(\mathbf{D}^{-1} \mathbf{H} \right)^{-1} = \mathbf{S}. \quad (67)$$

This means that the elements of the channel inverse can be written as

$$\mathbf{e}_j^H \mathbf{H}^{-1} \mathbf{e}_i = \mathbf{e}_j^H \sum_{n=0}^{\infty} \left(\mathbf{D}^{-1} \left(\mathbf{D} - \mathbf{H} \right) \right)^n \mathbf{D}^{-1} \mathbf{e}_i, \quad \forall i \quad (68)$$

$$= \sum_{n=0}^{\infty} C_n^{j,i} \quad (69)$$

where we have defined

$$C_n^{j,i} \triangleq \mathbf{e}_j^H \left(\mathbf{D}^{-1} \left(\mathbf{D} - \mathbf{H} \right) \right)^n \mathbf{D}^{-1} \mathbf{e}_i, \quad \forall i, n. \quad (70)$$

Writing explicitly the matrix products in (70), it follows from basic algebra that

$$\mathbb{E} \left[\left| C_n^{j,i} \right|^2 \right] \leq \left(P^{-\min_{i \neq j} \Gamma_{i,j}} \right)^n. \quad (71)$$

Hence, the $C_{j,i}^n$ with n such that

$$n \min_{i \neq j} \Gamma_{i,j} > 1 \quad (72)$$

can be neglected without any impact over the DoF. This translates to truncating the infinite summation to a finite summation up to n_0 defined as

$$n_0 \triangleq \left\lceil \frac{1}{\min_{i \neq j} \Gamma_{i,j}} \right\rceil. \quad (73)$$

$$\left| \frac{e_j^H \mathbf{H}^{-1} \mathbf{e}_i}{\|\mathbf{H}^{-1} \mathbf{e}_i\|} - \frac{e_j^H (\mathbf{H}^{(j)})^{-1} \mathbf{e}_i}{\|(\mathbf{H}^{(j)})^{-1} \mathbf{e}_i\|} \right| \leq \|\mathbf{H}^{-1} \mathbf{e}_i\| \left| e_j^H \mathbf{H}^{-1} \mathbf{e}_i - e_j^H (\mathbf{H}^{(j)})^{-1} \mathbf{e}_i \right| + |e_j^H (\mathbf{H}^{(j)})^{-1} \mathbf{e}_i| \left| \|\mathbf{H}^{-1} \mathbf{e}_i\| - \|(\mathbf{H}^{(j)})^{-1} \mathbf{e}_i\| \right| \quad (64)$$

A. Analysis of $|e_j^H \mathbf{H}^{-1} \mathbf{e}_i - e_j^H (\mathbf{H}^{(j)})^{-1} \mathbf{e}_i|$

We focus first on the first term, and since the norm $\|\mathbf{H}^{-1} \mathbf{e}_i\|$ is equivalent to H_{ii}^{-1} as the SNR P increases, we omit it in the following calculation for the sake of clarity.

We further introduce $\forall i, n$,

$$\mathbf{D}^{(j)} \triangleq \text{diag}(\mathbf{H}^{(j)}), \quad (74)$$

$$C_n^{j,i,(j)} \triangleq e_j^H ((\mathbf{D}^{(j)})^{-1} (\mathbf{D}^{(j)} - \mathbf{H}^{(j)}))^n (\mathbf{D}^{(j)})^{-1} \mathbf{e}_i. \quad (75)$$

We can then write

$$\begin{aligned} & \mathbb{E}[|e_j^H \mathbf{H}^{-1} \mathbf{e}_i - e_j^H (\mathbf{H}^{(j)})^{-1} \mathbf{e}_i|^2] \\ & \doteq \mathbb{E} \left[\left| \sum_{n=1}^{n_0} C_n^{j,i} - C_n^{j,i,(j)} \right|^2 \right] \end{aligned} \quad (76)$$

$$\leq \mathbb{E} \left[\left(\sum_{n=1}^{n_0} |C_n^{j,i} - C_n^{j,i,(j)}| \right)^2 \right] \quad (77)$$

$$\leq \sum_{n=1}^{n_0} \mathbb{E}[|C_n^{j,i} - C_n^{j,i,(j)}|^2] \quad (78)$$

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

$$\mathbb{E}[|C_n^{j,i} - C_n^{j,i,(j)}|^2] \leq P^{-1}, \quad \forall i, n. \quad (79)$$

As a starting point, we consider the first coefficients.

$$\mathbb{E}[|C_0^{j,i} - C_0^{j,i,(j)}|^2] = \mathbb{E} \left[\left| \frac{e_j^H \mathbf{e}_i}{H_{i,i}} - \frac{e_j^H \mathbf{e}_i}{H_{i,i}^{(j)}} \right|^2 \right] \quad (80)$$

$$\doteq 2^{-B_{i,i}^{(j)}} \delta_{j,i}. \quad (81)$$

Thus, we have

$$B_{j,j}^{(j)} \geq \lceil \log_2(P) \rceil. \quad (82)$$

This ensures to fulfill (79) for $n = 0$. The error done over $H_{j,j}$ becomes then negligible in terms of DoF (i.e., in terms of exponential equality). For $n \geq 1$, it holds that $C_n^{j,j} = 0$ such that we assume that $i \neq j$ in the following,

$$\mathbb{E}[|C_1^{j,i} - C_1^{j,i,(j)}|^2] = \mathbb{E} \left[\left| \frac{\{\mathbf{D} - \mathbf{H}\}_{j,i}}{H_{j,j} H_{i,i}} - \frac{\{\mathbf{D}^{(j)} - \mathbf{H}^{(j)}\}_{j,i}}{H_{j,j}^{(j)} H_{i,i}^{(j)}} \right|^2 \right] \quad (83)$$

$$\doteq \mathbb{E} \left[\left| \frac{H_{i,i}^{(j)} \tilde{H}_{j,i} - H_{i,i} \tilde{H}_{j,i}^{(j)}}{H_{j,j} H_{i,i} H_{i,i}^{(j)}} \right|^2 \right] P^{-\Gamma_{j,i}} \quad (84)$$

$$\doteq (2^{-B_{i,i}^{(j)}} + 2^{-B_{j,i}^{(j)}}) P^{-\Gamma_{j,i}} \quad (85)$$

Setting

$$B_{i,i}^{(j)} \geq \lceil [1 - \Gamma_{j,i}]^+ \log_2(P) \rceil, \quad \forall i \neq j \quad (86)$$

$$B_{j,i}^{(j)} \geq \lceil [1 - \Gamma_{j,i}]^+ \log_2(P) \rceil, \quad \forall i \neq j \quad (87)$$

ensures to fulfill (79) for all streams i for $n = 1$. Going further, we consider then $C_2^{j,i}$,

$$\mathbb{E}[|C_2^{j,i} - C_2^{j,i,(j)}|^2] \quad (88)$$

$$\begin{aligned} & = \mathbb{E} \left[\left| e_j^H (\mathbf{D}^{-1} (\mathbf{D} - \mathbf{H}))^2 \mathbf{D}^{-1} \mathbf{e}_i \right. \right. \\ & \quad \left. \left. - e_j^H ((\mathbf{D}^{(j)})^{-1} (\mathbf{D}^{(j)} - \mathbf{H}^{(j)}))^2 (\mathbf{D}^{(j)})^{-1} \mathbf{e}_i \right|^2 \right] \end{aligned} \quad (89)$$

$$\begin{aligned} & = \mathbb{E} \left[\left| \sum_{k=1, k \neq i, k \neq j}^K \frac{1}{H_{j,j}^2 H_{i,i}} \tilde{H}_{j,k} \tilde{H}_{k,i} \right. \right. \\ & \quad \left. \left. - \frac{1}{(H_{j,j}^{(j)})^2 H_{i,i}^{(j)}} \tilde{H}_{j,k}^{(j)} \tilde{H}_{k,i}^{(j)} \right|^2 \right] P^{-(\Gamma_{j,k} + \Gamma_{k,i})} \end{aligned} \quad (90)$$

$$\leq \mathbb{E} \left[\sum_{k=1, k \neq i, k \neq j}^K \left| \tilde{H}_{j,k} \tilde{H}_{k,i} - \tilde{H}_{j,k}^{(j)} \tilde{H}_{k,i}^{(j)} \right|^2 \right] P^{-(\Gamma_{j,k} + \Gamma_{k,i})} \quad (91)$$

$$\doteq \sum_{k=1, k \neq i, k \neq j}^K (2^{-B_{j,k}^{(j)}} - 2^{B_{k,i}^{(j)}}) P^{-(\Gamma_{j,k} + \Gamma_{k,i})} \quad (92)$$

$$\doteq \sum_{k=1, k \neq i, k \neq j}^K 2^{-B_{k,i}^{(j)}} P^{-(\Gamma_{j,k} + \Gamma_{k,i})} \quad (93)$$

where we could remove $2^{-B_{j,k}^{(j)}}$ in the last exponential inequality because of (87). Setting

$$B_{k,i}^{(j)} \geq \lceil [1 - (\Gamma_{j,k} + \Gamma_{k,i})]^+ \log_2(P) \rceil, \quad \forall k, i \quad (94)$$

allows to fulfill (79) for $n = 2$. Going to arbitrary value of n , we obtain equation (95) at the top of next page. It is clear from (95) that when considering the i th stream, the coefficient $\tilde{H}_{k,\ell}$ appears weighted with P^{-1} at a coefficient at least equal to $\Gamma_{k \rightarrow j} + \Gamma_{k,\ell} + \Gamma_{i \rightarrow \ell}$. Yet, this condition has to be fulfilled for every stream, hence every value of i . Since $\Gamma_{i \rightarrow \ell} \geq 0$, taking $i = \ell$ gives the most tight constraint. In total, this gives the CSIT allocation

$$B_{k,\ell}^{(j)} \geq \lceil [1 - (\Gamma_{k \rightarrow j} + \Gamma_{k,\ell})]^+ \log_2(P) \rceil, \quad \forall k, \ell. \quad (96)$$

This CSIT allocation can be seen to fulfill (79) for all i and all n .

$$\begin{aligned} & \mathbb{E}[|C_n^{j,i} - C_n^{j,i,(j)}|^2] \\ &= \mathbb{E}\left[\left|\sum_{k_1 \neq j}^K \sum_{k_2 \neq k_1}^K \dots \sum_{\substack{k_{n-1} \neq k_{n-2} \\ k_{n-1} \neq i}}^K \left(\frac{\tilde{H}_{j,k_1} \tilde{H}_{k_1,k_2} \dots \tilde{H}_{k_{n-1},i}}{H_{j,j}^n H_{i,i}} - \frac{\tilde{H}_{j,k_1}^{(j)} \tilde{H}_{k_1,k_2}^{(j)} \dots \tilde{H}_{k_{n-1},i}^{(j)}}{(H_{j,j}^{(j)})^n H_{i,i}^{(j)}}\right)\right|^2\right] P^{-(\Gamma_{j,k_1} + \Gamma_{k_1,k_2} + \dots + \Gamma_{k_{n-1},i})}. \quad (95) \end{aligned}$$

B. Analysis of $\|e_j^H(\mathbf{H}^{(j)})^{-1}e_i\|$ vs $\|\mathbf{H}^{-1}e_i\|$ and $\|(\mathbf{H}^{(j)})^{-1}e_i\|$

From the norm properties, it holds that

$$\left|\|\mathbf{H}^{-1}e_i\| - \|(\mathbf{H}^{(j)})^{-1}e_i\|\right|^2 \leq \left\|\mathbf{H}^{-1}e_i - (\mathbf{H}^{(j)})^{-1}e_i\right\|^2 \quad (97)$$

$$= \sum_{k=1}^K \left|e_k^H \mathbf{H}^{-1}e_i - e_k^H (\mathbf{H}^{(j)})^{-1}e_i\right|^2 \quad (98)$$

Hence, we can study the CSIT requirements ensuring that

$$\mathbb{E}[|C_n^{k,i} - C_n^{k,i,(j)}|^2] \leq P^{-1}, \quad \forall k, n. \quad (99)$$

It can be seen that the difference between (99) and (79) comes from the fact that (79) had to be fulfilled for every stream i , whereas (99) has to be fulfilled for every k , with the stream index fixed. Proceeding similarly as in the first part, we obtain the following condition

$$B_{k,\ell}^{(j)} \geq 1 - \Gamma_{k,\ell} - \Gamma_{i \rightarrow \ell}, \quad \forall k, \ell. \quad (100)$$

However, we have also shown in the previous calculations that

$$|e_j^H (\mathbf{H}^{(j)})^{-1}e_i|^2 \leq P^{-\Gamma_{j,i}}. \quad (101)$$

Taking this further attenuation into account and the fact that the CSIT requirements have to be fulfilled for every stream i , we obtain the CSIT allocation

$$B_{k,\ell}^{(j)} \geq 1 - \Gamma_{k,\ell} - \min_i (\Gamma_{i \rightarrow \ell} + \Gamma_{j,i}). \quad (102)$$

Putting together (96) and (102) gives the CSIT allocation in Theorem 1. \square

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] P. Marsch and G. Fettweis, "On multicell cooperative transmission in backhaul-constrained cellular systems," *Annals of Telecommunications*, vol. 63, no. 5, 2008.
- [4] 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.
- [5] B. L. Ng, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theo.*, vol. 54, no. 12, pp. 5491–5499, 2008.
- [6] S. Shamai (Shitz) and M. Wigger, "Rate-limited transmitter-cooperation in Wyner's asymmetric interference network," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2011.
- [7] S. Park, O. Simeone, O. Sahin, and S. Shamai (Shitz), "Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5646–5658, 2013.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] A. Giovanidis, J. Krolkowski, 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] B. Özbek and D. Le Ruyet, "Adaptive limited feedback for intercell interference cancellation in cooperative downlink multicell networks," in *Proc. IEEE International Symposium on Wireless Communication Systems (ISWCS)*, 2010.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] B. Khoshnevis, W. Yu, and Y. Lohan, "Two-stage channel feedback for beamforming and scheduling in network MIMO systems," in *Proc. International Conference on Communications (ICC)*, 2012.
- [22] 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.
- [23] D. Tuninetti, "On interference channel with generalized feedback (IFC-GF)," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2007.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.

- [28] 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.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.
- [33] 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.
- [34] T. Cover and A. Thomas, *Elements of information theory*. Wiley-Interscience, Jul. 2006.
- [35] B. Girod. (2013, Aug.) Lecture on image and video compression (EE398A). [Online]. Available: <http://www.stanford.edu/class/ee398a/handouts/lectures/05-Quantization.pdf>
- [36] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] 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.
- [41] 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.
- [42] 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.



David Gesbert (IEEE Fellow) is Professor and Head of the Mobile Communications Department, EURECOM, France, where he also heads the Communications Theory Group. He obtained the Ph.D degree from Ecole Nationale Supérieure des Télécommunications, 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 over 200 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 was 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 2012 SPS Signal Processing Magazine Best Paper Award, 2004 IEEE Best Tutorial Paper Award (Communications Society), 2005 Young Author Best Paper Award for Signal Proc. Society journals, and paper awards at conferences 2011 IEEE SPAWC, 2004 ACM MSWiM workshop. He co-authored the book Space time wireless communications: From parameter estimation to MIMO systems, Cambridge Press, 2006. In 2013, he was a General Chair for the IEEE Communications Theory Workshop, and a Technical Program Chair for the Communications Theory Symposium of ICC2013. He is a Technical Program Chair for IEEE ICC 2017, to be held in Paris.



Paul de Kerret (IEEE Student Member) graduated in 2009 from Ecole Nationale Supérieure des Télécommunications 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 Université 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. In december

2013, he obtained a Ph.D. degree in the Mobile Communications Department at EURECOM, France, under the supervision of David Gesbert. He is the first author of several journal papers in prestigious journals and a magazine. With David Gesbert, he will give a tutorial in ICASSP 2014 on the challenges behind the cooperation of transmitters in wireless networks. He is interested in communication theory, information theory, and game theory.