Cooperative MIMO Precoding with Distributed CSI: A Hierarchical Approach

Italo Atzeni and David Gesbert

EURECOM, Communication Systems Department, Sophia Antipolis, France

Emails: {italo.atzeni, david.gesbert}@eurecom.fr

Abstract—The problem of network multiple-input multiple-output precoding under distributed channel state information is a notoriously challenging question, for which optimal solutions with reasonable complexity remain elusive. In this context, we assess the value of hierarchical information exchange, whereby an order is established among the transmitters (TXs) in such a way that a given TX has access not only to its local channel estimate but also to the estimates available at the less informed TXs. Assuming regularized zero forcing (RZF) precoding at the TXs, we propose naive, locally robust, and globally robust suboptimal strategies for the joint precoding design. Numerical results show that hierarchical information exchange brings significant performance gains, with the locally and globally robust algorithms performing remarkably close to the optimal RZF strategy. Lastly, the cost of hierarchical information exchange relative to the cooperation gain is examined and the optimal tradeoff is numerically evaluated.

Index Terms—Cooperative communications, distributed CSI, limited feedback, network MIMO, robust precoding.

I. Introduction

Network multiple-input multiple-output (MIMO) systems, whereby distributed transmitters (TXs) sharing user data symbols and channel state information (CSI) cooperatively serve several receivers (RXs) by cooperatively designing their downlink precoding strategy, are regarded as a promising solution to enhance data rates and to meet the quality-of-service requirements of future cellular networks [1]. A practical limitation of decentralized network MIMO systems, which makes the joint precoding optimization challenging, is that the CSI is actually known imperfectly and differently across the TXs due to limited and uneven feedback [2]: this occurs, for instance, when the TXs are not connected to a perfect backhaul or are mounted on mobile devices such as vehicles or drones [3]. Under such a *distributed CSI* (D-CSI) setting, each TX needs to design its precoding strategy solely on the basis of its local CSI without any further information exchange with the other TXs. This problem falls into the category of so-called *team decision problems* [4], where multiple decentralized decision makers aim at coordinating their strategies to maximize the system-level performance while not being able to accurately predict the actions taken by the other decision makers.

Under centralized CSI, the network MIMO TXs can be virtually combined into a unique antenna array serving the RXs in a multi-antenna broadcast channel fashion, for which there is a large body of literature dealing with robust precoding

in presence of CSI imperfections (see, e.g., [5], [6]). On the other hand, fewer results are available for the D-CSI case. Among these, the work in [7] proposes a robust distributed precoding method that relies on the quantization of the CSI space to enforce coordination between the TXs. In [8], D-CSI arises from combining periodical feedback via backhaul links, equal for all TXs, with local CSI exchanges from neighboring RXs, generating partially new local CSI at a given TX between backhaul updates. Furthermore, [9] proposes a D-CSI structure where the TXs are ordered by increasing level of CSI, i.e., in such a way that a given TX has access not only to its local CSI but also to the CSI available at the less informed TXs.

In this paper, we formulate the general joint precoding optimization problem under D-CSI as a team decision problem and particularize it to a *deterministically hierarchical D-CSI* scenario [9]. Hierarchical D-CSI can be enforced by a suitable information exchange mechanism between the TXs at a certain signaling/power cost: here, we show how such a hierarchical information exchange can be leveraged to yield implementable and efficient distributed precoding solutions. In particular, restricting the structure of the precoding strategies to *regularized zero forcing* (RZF) precoding [10] and considering the ergodic sum rate as performance metric, we propose *naive*, *locally robust*, and *globally robust* suboptimal strategies (in increasing order of both performance and computational complexity) for the joint precoding design. Numerical results show that the deterministically hierarchical D-CSI configuration yields significant gains over the classical non-hierarchical D-CSI counterpart, with the locally and globally robust algorithms performing remarkably close to the optimal RZF strategy. Lastly, the cost of hierarchical information exchange is examined and is shown to be outweighed by the resulting cooperation gain.

II. System Model

Let us consider a network MIMO system where N distributed multi-antenna TXs cooperatively serve K single-antenna RXs in the downlink. Each TX n is equipped with M_n antennas and the total number of transmit antennas among the N TXs is $M \triangleq \sum_{n=1}^{N} M_n$. We use $\mathbf{h}_{kn} \in \mathbb{C}^{M_n \times 1}$, $\mathbf{h}_k \triangleq [\mathbf{h}_{k1}^{\mathrm{T}} \ldots \mathbf{h}_{kN}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{M \times 1}$, and $\mathbf{H} \triangleq [\mathbf{h}_1 \ldots \mathbf{h}_K] \in \mathbb{C}^{M \times K}$ to denote the channel vector between TX n and RX k, the channel vector between the N TXs and RX k, and the channel matrix between the N TXs and the K RXs, respectively. Furthermore, we assume that $\mathbf{h}_{kn} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_{kn})$, with $\boldsymbol{\Sigma}_{kn} \in \mathbb{C}^{M_n \times M_n}$ being the covariance matrix of $\mathbf{h}_{kn}$: hence, it follows that $\mathbf{h}_k \sim$

$$\mathbf{W}_\star^{(n)} \triangleq \operatorname*{argmax}_{\mathbf{W}^{(n)}} \; \mathbb{E}_{\{\hat{\mathbf{H}}^{(\ell)}\}_{\ell \neq n}|\hat{\mathbf{H}}^{(n)}} \left[\max_{\{\mathbf{W}^{(\ell)}\}_{\ell \neq n}} \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}^{(n)}} \left[R\Big(\mathbf{H}, \mathbf{W}^{(n)}(\hat{\mathbf{H}}^{(n)}), \{\mathbf{W}^{(\ell)}(\hat{\mathbf{H}}^{(\ell)})\}_{\ell \neq n}\Big) \right] \right] \tag{6}$$
$$\text{s.t.} \quad \big\| \mathbf{W}^{(\ell)}(\hat{\mathbf{H}}^{(\ell)}) \big\|_{\mathrm{F}}^2 \leq P_\ell, \qquad \ell = 1, \dots, N$$

$$\mathbf{W}_{\star,\mathrm{h}}^{(n)} \triangleq \operatorname*{argmax}_{\mathbf{W}^{(n)}} \; \mathbb{E}_{\{\hat{\mathbf{H}}^{(\ell)}\}_{\ell=n+1}^{N}|\hat{\mathbf{H}}^{(n)}} \left[\max_{\{\mathbf{W}^{(\ell)}\}_{\ell=n+1}^{N}} \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}^{(n)}} \left[R\Big(\mathbf{H}, \{\mathbf{W}_{\star,\mathrm{h}}^{(\ell)}\}_{\ell=1}^{n-1}, \mathbf{W}^{(n)}(\hat{\mathbf{H}}^{(n)}), \{\mathbf{W}^{(\ell)}(\hat{\mathbf{H}}^{(\ell)})\}_{\ell=n+1}^{N}\Big) \right] \right] \tag{11}$$
$$\text{s.t.} \quad \big\| \mathbf{W}^{(\ell)}(\hat{\mathbf{H}}^{(\ell)}) \big\|_{\mathrm{F}}^2 \leq P_\ell, \qquad \ell = n, \dots, N$$

$\mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_k)$, with $\boldsymbol{\Sigma}_k \triangleq \operatorname{blkdiag}(\boldsymbol{\Sigma}_{k1}, \dots, \boldsymbol{\Sigma}_{kN}) \in \mathbb{C}^{M \times M}$ being the covariance matrix of $\mathbf{h}_k$.

Let $\mathbf{W} \in \mathbb{C}^{M \times K}$ denote the multiuser precoding matrix given by

$$\mathbf{W} \triangleq [\mathbf{w}_1 \dots \mathbf{w}_K] = \begin{bmatrix} \mathbf{W}^{(1)} \\ \vdots \\ \mathbf{W}^{(N)} \end{bmatrix} \tag{1}$$

where $\mathbf{w}_k \in \mathbb{C}^{M \times 1}$ is the beamforming vector used by the N TXs to serve RX k and $\mathbf{W}^{(n)} \in \mathbb{C}^{M_n \times K}$ is the precoding submatrix used by TX n; a per-TX power constraint is assumed such that $\|\mathbf{W}^{(n)}\|_{\mathrm{F}}^2 \leq P_n$. The receive signal at RX k is then expressed as

$$y_k \triangleq \mathbf{h}_k^{\mathrm{H}} \mathbf{x} + z_k \tag{2}$$

where $\mathbf{x} \in \mathbb{C}^{M \times 1}$ is the transmit signal obtained from the user data symbol vector $\mathbf{s} \triangleq [s_1 \dots s_K]^{\mathrm{T}} \in \mathbb{C}^{K \times 1}$ as

$$\mathbf{x} \triangleq \mathbf{W}\mathbf{s} = \sum_{k=1}^{K} \mathbf{w}_k s_k \tag{3}$$

and $z_k \sim \mathcal{CN}(0, \sigma^2)$ is the noise at RX k. Finally, the sum rate of the network MIMO system is given by

$$R(\mathbf{H}, \mathbf{W}) \triangleq \sum_{k=1}^{K} \log_2 \left(1 + \frac{|\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_k|^2}{\sum_{j \neq k} |\mathbf{h}_k^{\mathrm{H}} \mathbf{w}_j|^2 + \sigma^2} \right). \tag{4}$$

III. DISTRIBUTED CSI MODEL

In practice, not only is the channel matrix known imperfectly but also differently across the network nodes due to limited and uneven feedback. In this paper, we thus consider a D-CSI scenario [2], where each TX n has a different estimate of the channel matrix $\mathbf{H}$,[1] denoted by $\hat{\mathbf{H}}^{(n)} = [\hat{\mathbf{h}}_1^{(n)} \dots \hat{\mathbf{h}}_K^{(n)}] \in \mathbb{C}^{M \times K}$. The imperfect CSI at TX n is modeled as

$$\hat{\mathbf{H}}^{(n)} \triangleq \sqrt{1 - \epsilon_n^2} \mathbf{H} + \epsilon_n \mathbf{E}^{(n)} \tag{5}$$

where $\epsilon_n \in [0, 1]$ describes the quality of the channel estimation and $\mathbf{E}^{(n)} \triangleq [\mathbf{e}_1^{(n)} \dots \mathbf{e}_K^{(n)}] \in \mathbb{C}^{M \times K}$ is the error matrix, where $\mathbf{e}_k^{(n)} \sim \mathcal{CN}(0, \boldsymbol{\Upsilon}^{(n)})$, $\forall k = 1, \dots, K$, with $\boldsymbol{\Upsilon}^{(n)} \in \mathbb{C}^{M \times M}$ being the error covariance matrix of TX n.

Hence, in a D-CSI scenario, it is meaningful to formulate a team decision problem (see [4]) where each TX n computes

its precoding submatrix with the objective of maximizing the ergodic sum rate given the local channel estimate $\hat{\mathbf{H}}^{(n)}$, as shown in (6) at the top of the page, where we have expressed the precoding submatrix computed by each TX as a function of its local channel estimate. The conditional distributions of $\mathbf{H}|\hat{\mathbf{H}}^{(n)}$ and $\{\hat{\mathbf{H}}^{(\ell)}\}_{\ell \neq n}|\hat{\mathbf{H}}^{(n)}$, by which TX n can make a prediction on the real channel matrix and on the CSI available at the other TXs, respectively, are derived in the following proposition.

Proposition 1. *Given the unconditional channel* $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_k)$ *and the channel estimation model in* (5)*, the following hold:*

i) *The channel* $\mathbf{h}_k$ *conditioned on the channel estimate* $\hat{\mathbf{h}}_k^{(n)}$ *is distributed as* $\mathbf{h}_k|\hat{\mathbf{h}}_k^{(n)} \sim \mathcal{CN}(\boldsymbol{\mu}_k^{(n)}, \boldsymbol{\Sigma}_k^{(n)})$*, with*

$$\boldsymbol{\mu}_k^{(n)} \triangleq \sqrt{1 - \epsilon_n^2} \boldsymbol{\Sigma}_k \big((1 - \epsilon_n^2) \boldsymbol{\Sigma}_k + \epsilon_n^2 \boldsymbol{\Upsilon}^{(n)} \big)^{-1} \hat{\mathbf{h}}_k^{(n)}, \tag{7}$$
$$\boldsymbol{\Sigma}_k^{(n)} \triangleq \boldsymbol{\Sigma}_k - (1 - \epsilon_n^2) \boldsymbol{\Sigma}_k \big((1 - \epsilon_n^2) \boldsymbol{\Sigma}_k + \epsilon_n^2 \boldsymbol{\Upsilon}^{(n)} \big)^{-1} \boldsymbol{\Sigma}_k. \tag{8}$$

ii) *The channel estimate* $\hat{\mathbf{h}}_k^{(\ell)}$ *conditioned on the channel estimate* $\hat{\mathbf{h}}_k^{(n)}$ *is distributed as* $\hat{\mathbf{h}}_k^{(\ell)}|\hat{\mathbf{h}}_k^{(n)} \sim \mathcal{CN}(\boldsymbol{\mu}_k^{(\ell|n)}, \boldsymbol{\Sigma}_k^{(\ell|n)})$*, with*

$$\boldsymbol{\mu}_k^{(\ell|n)} \triangleq \sqrt{1 - \epsilon_\ell^2} \boldsymbol{\mu}_k^{(n)}, \tag{9}$$
$$\boldsymbol{\Sigma}_k^{(\ell|n)} \triangleq (1 - \epsilon_\ell^2) \boldsymbol{\Sigma}_k^{(n)} + \epsilon_\ell^2 \boldsymbol{\Upsilon}^{(\ell)} \tag{10}$$

with $\boldsymbol{\mu}_k^{(n)}$ *and* $\boldsymbol{\Sigma}_k^{(n)}$ *defined in* (7) *and* (8)*, respectively.*

In Proposition 1, $\boldsymbol{\mu}_k^{(n)}$ (resp. $\boldsymbol{\mu}_k^{(\ell|n)}$) is computed as the minimum mean square error (MMSE) estimate of $\mathbf{h}_k|\hat{\mathbf{h}}_k^{(n)}$ (resp. $\hat{\mathbf{h}}_k^{(\ell)}|\hat{\mathbf{h}}_k^{(n)}$), whereas $\boldsymbol{\Sigma}_k^{(n)}$ (resp. $\boldsymbol{\Sigma}_k^{(\ell|n)}$) is the corresponding MMSE covariance matrix. In the rest of the paper, we assume that the channel covariance matrices $\{\boldsymbol{\Sigma}_k\}_{k=1}^{K}$, the error covariance matrices $\{\boldsymbol{\Upsilon}^{(n)}\}_{n=1}^{N}$, and the coefficients $\{\epsilon_n\}_{n=1}^{N}$ are perfectly known across the network[2] so that each TX n can derive the conditional distributions of $\mathbf{H}|\hat{\mathbf{H}}^{(n)}$ and $\{\mathbf{H}^{(\ell)}\}_{\ell \neq n}|\hat{\mathbf{H}}^{(n)}$ using (7)–(8) and (9)–(10), respectively.

A. Deterministically Hierarchical D-CSI Model

In this paper, we analyze a deterministically hierarchical network MIMO system, whereby an order is established among the TXs in such a way that TX n has access not only to $\hat{\mathbf{H}}^{(n)}$ but also to $\{\hat{\mathbf{H}}^{(\ell)}\}_{\ell=1}^{n-1}$. Hence, such a deterministically hierarchical D-CSI structure allows each TX to determine exactly the strategies

---

[1]Even though the CSI is distributed, it is still reasonable to assume that the user data symbol vector $\mathbf{s}$ is perfectly known at all TXs.

[2]Observe that this is a realistic assumption since these parameters depend on long-term statistics.

$$\alpha_{\star,\mathrm{h}}^{(n)} \triangleq \underset{\alpha^{(n)} \in [0,1]}{\mathrm{argmax}} \; \mathbb{E}_{\{\hat{\mathbf{H}}^{(\ell)}\}_{\ell=n+1}^{N} | \hat{\mathbf{H}}^{(n)}} \left[\underset{\{\alpha^{(\ell)} \in [0,1]\}_{\ell=n+1}^{N}}{\max} \; \mathbb{E}_{\mathbf{H} | \hat{\mathbf{H}}^{(n)}} \left[R\Big(\mathbf{H}, \big\{ \mathbf{W}_{\mathrm{rzf}}^{(\ell)}(\hat{\mathbf{H}}^{(\ell)}, \alpha_{\star,\mathrm{h}}^{(\ell)}) \big\}_{\ell=1}^{n-1}, \right.\right.$$

$$\left.\left. \mathbf{W}_{\mathrm{rzf}}^{(n)}(\hat{\mathbf{H}}^{(n)}, \alpha^{(n)}), \big\{ \mathbf{W}_{\mathrm{rzf}}^{(\ell)}(\hat{\mathbf{H}}^{(\ell)}, \alpha^{(\ell)}) \big\}_{\ell=n+1}^{N} \Big) \right] \right] \qquad (13)$$

$$\alpha_{\mathrm{NA,h}}^{(n)} \triangleq \underset{\alpha^{(n)} \in [0,1]}{\mathrm{argmax}} \; R\Big(\hat{\mathbf{H}}^{(n)}, \big\{ \mathbf{W}_{\mathrm{rzf}}^{(\ell)}(\hat{\mathbf{H}}^{(\ell)}, \alpha_{\mathrm{NA,h}}^{(\ell)}) \big\}_{\ell=1}^{n-1}, \mathbf{W}_{\mathrm{rzf}}^{(n)}(\hat{\mathbf{H}}^{(n)}, \alpha^{(n)}), \big\{ \mathbf{W}_{\mathrm{rzf}}^{(\ell)}(\hat{\mathbf{H}}^{(n)}, \alpha^{(n)}) \big\}_{\ell=n+1}^{N} \Big) \qquad (14)$$

$$\alpha_{\mathrm{LR,h}}^{(n)} \triangleq \underset{\alpha^{(n)} \in [0,1]}{\mathrm{argmax}} \; \mathbb{E}_{\mathbf{H} | \hat{\mathbf{H}}^{(n)}} \left[R\Big(\mathbf{H}, \big\{ \mathbf{W}_{\mathrm{rzf}}^{(\ell)}(\hat{\mathbf{H}}^{(\ell)}, \alpha_{\mathrm{LR,h}}^{(\ell)}) \big\}_{\ell=1}^{n-1}, \mathbf{W}_{\mathrm{rzf}}^{(n)}(\hat{\mathbf{H}}^{(n)}, \alpha^{(n)}), \big\{ \mathbf{W}_{\mathrm{rzf}}^{(\ell)}(\hat{\mathbf{H}}^{(n)}, \alpha^{(n)}) \big\}_{\ell=n+1}^{N} \Big) \right] \qquad (15)$$

$$\alpha_{\mathrm{GR,h}}^{(n)} \triangleq \underset{\alpha^{(n)} \in [0,1]}{\mathrm{argmax}} \; \mathbb{E}_{\{\hat{\mathbf{H}}^{(\ell)}\}_{\ell=n+1}^{N} | \hat{\mathbf{H}}^{(n)}} \left[\mathbb{E}_{\mathbf{H} | \hat{\mathbf{H}}^{(n)}} \left[R\Big(\mathbf{H}, \big\{ \mathbf{W}_{\mathrm{rzf}}^{(\ell)}(\hat{\mathbf{H}}^{(\ell)}, \alpha_{\mathrm{GR,h}}^{(\ell)}) \big\}_{\ell=1}^{n-1}, \mathbf{W}_{\mathrm{rzf}}^{(n)}(\hat{\mathbf{H}}^{(n)}, \alpha^{(n)}), \big\{ \mathbf{W}_{\mathrm{rzf}}^{(\ell)}(\hat{\mathbf{H}}^{(\ell)}, \alpha^{(n)}) \big\}_{\ell=n+1}^{N} \Big) \right] \right]$$
$$(16)$$

computed by the less informed TXs; nevertheless, the strategies used by the more informed TXs can only be predicted imperfectly since their channel estimates are not known.

In this setting, we have a team decision problem where each TX n computes its precoding submatrix with the objective of maximizing the ergodic sum rate given the local channel estimate $\hat{\mathbf{H}}^{(n)}$ and the precoding submatrices computed by the less informed TXs, as shown in (11) at the top of the previous page.

IV. REGULARIZED ZERO FORCING PRECODING

We assume that RZF precoding [10] is adopted at each TX. The RZF precoding submatrix used by TX n has the form[3]

$$\mathbf{W}_{\mathrm{rzf}}^{(n)}(\hat{\mathbf{H}}^{(n)}, \alpha^{(n)}) \triangleq \sqrt{P_n}$$
$$\times \frac{\boldsymbol{\Delta}_n^{\mathrm{T}} \hat{\mathbf{H}}^{(n)} \big((1-\alpha^{(n)})(\hat{\mathbf{H}}^{(n)})^{\mathrm{H}} \hat{\mathbf{H}}^{(n)} + \alpha^{(n)} \mathbf{I}_K \big)^{-1}}{\big\| \boldsymbol{\Delta}_n^{\mathrm{T}} \hat{\mathbf{H}}^{(n)} \big((1-\alpha^{(n)})(\hat{\mathbf{H}}^{(n)})^{\mathrm{H}} \hat{\mathbf{H}}^{(n)} + \alpha^{(n)} \mathbf{I}_K \big)^{-1} \big\|_{\mathrm{F}}} \quad (12)$$

where $\boldsymbol{\Delta}_n \triangleq \big[\mathbf{0}_{M_n \times \sum_{\ell=1}^{n-1} M_\ell} \; \mathbf{I}_{M_n} \; \mathbf{0}_{M_n \times \sum_{\ell=n+1}^{N} M_\ell} \big]^{\mathrm{T}} \in \mathbb{C}^{M \times M_n}$ is a block selection matrix and $\alpha^{(n)} \in [0,1]$ is the regularization factor. The advantage of RZF precoding stems from the fact that only a one-dimensional real parameter, i.e., the regularization factor, needs to be optimized at each TX n instead of a $(M_n \times K)$-dimensional complex matrix.

With RZF precoding, we have a team decision problem where each TX n computes its regularization factor with the objective of maximizing the ergodic sum rate given the local channel estimate $\hat{\mathbf{H}}^{(n)}$ and the precoding submatrices computed by the less informed TXs, as shown in (13) at the top of the page. In the following, we refer to (13) as *optimal approach*. By comparing (11) and (13), it is straightforward to note that adopting RZF at the TXs greatly reduces the complexity of the computation of the precoding submatrices.

A. Lower-Complexity Algorithms

Deriving $\alpha_{\star,\mathrm{h}}^{(n)}$ as in (13) is still impractical due to the expectation over the channel estimates at the more informed

[3]Note that this way of enforcing the per-TX power normalization is not necessarily optimal; however, it does not require any additional information exchange between the TXs.

TXs conditioned on the channel estimate at TX n, i.e., $\{\hat{\mathbf{H}}^{(\ell)}\}_{\ell=n+1}^{N} | \hat{\mathbf{H}}^{(n)}$, and the maximization at the objective over the regularization factors of the more informed TXs, i.e., $\{\alpha^{(\ell)}\}_{\ell=n+1}^{N}$, within the aforementioned expectation. Hence, in the following, we present three suboptimal approaches with the aim of reducing the complexity in the computation of the regularization factors. The algorithms are presented in increasing order of both performance and computational complexity.

- *Naive approach.* Each TX n assumes that its local CSI is perfect and shared by the more informed TXs, i.e., $\{\hat{\mathbf{H}}^{(\ell)} = \mathbf{H}\}_{\ell=n}^{N}$: in this setting, $\alpha_{\star,\mathrm{h}}^{(n)}$ simplifies to $\alpha_{\mathrm{NA,h}}^{(n)}$ in (14), shown at the top of the page.
- *Locally robust approach.* Each TX n assumes that its local CSI is imperfect and shared by the more informed TXs, i.e., $\{\hat{\mathbf{H}}^{(\ell)} = \hat{\mathbf{H}}^{(n)}\}_{\ell=n+1}^{N}$: in this setting, $\alpha_{\star,\mathrm{h}}^{(n)}$ simplifies to $\alpha_{\mathrm{LR,h}}^{(n)}$ in (15), shown at the top of the page.
- *Globally robust approach.* Each TX n assumes that its local CSI is imperfect and not shared by the more informed TXs; however, in order to reduce the computational complexity with respect to (13), it neglects the possibly different regularization factors adopted by the more informed TXs: in this setting, $\alpha_{\star,\mathrm{h}}^{(n)}$ simplifies to $\alpha_{\mathrm{GR,h}}^{(n)}$ in (16), shown at the top of the page.

In the naive approach (14), neither the local nor the global CSI imperfections are taken into account. On the other hand, compared with the latter, the locally robust approach (15) is robust to local CSI imperfections, although it does not deal with the possibly different channel estimates at the more informed TXs. Lastly, the globally robust approach (16) adds some global robustness to the local robustness of (15) but, unlike (13), it does not involve the maximization of the objective over $\{\alpha^{(\ell)}\}_{\ell=n+1}^{N}$ within the expectation over $\{\hat{\mathbf{H}}^{(\ell)}\}_{\ell=n+1}^{N} | \hat{\mathbf{H}}^{(n)}$.

V. THE CASE OF 2 TXS

In this section, we consider a simple network MIMO system with $N = 2$ TXs, where TX 2 has perfect CSI, i.e., $\hat{\mathbf{H}}^{(2)} = \mathbf{H}$. In this scenario, the regularization coefficients at TX 1 and TX 2 are computed building on both the optimal and the lower-complexity approaches (13)–(16) as follows.

- Optimal approach. (cf. (13)):

$$\alpha_{\star,\mathrm{h}}^{(1)} = \underset{\alpha^{(1)} \in [0,1]}{\arg\max} \, \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}^{(1)}} \Bigg[\underset{\alpha^{(2)} \in [0,1]}{\max} R\big(\mathbf{H},$$

$$\mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)}), \mathbf{W}_{\mathrm{rzf}}^{(2)}(\mathbf{H}, \alpha^{(2)})\big) \Bigg], \quad (17)$$

$$\alpha_{\star,\mathrm{h}}^{(2)} = \underset{\alpha^{(2)} \in [0,1]}{\arg\max} R\big(\mathbf{H}, \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha_{\star,\mathrm{h}}^{(1)}), \mathbf{W}_{\mathrm{rzf}}^{(2)}(\mathbf{H}, \alpha^{(2)})\big). \quad (18)$$

- Naive approach (cf. (14)):

$$\alpha_{\mathrm{NA},\mathrm{h}}^{(1)} = \underset{\alpha^{(1)} \in [0,1]}{\arg\max} R\big(\hat{\mathbf{H}}^{(1)}, \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)}),$$

$$\mathbf{W}_{\mathrm{rzf}}^{(2)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)})\big), \quad (19)$$

$$\alpha_{\mathrm{NA},\mathrm{h}}^{(2)} = \underset{\alpha^{(2)} \in [0,1]}{\arg\max} R\big(\mathbf{H}, \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha_{\mathrm{NA},\mathrm{h}}^{(1)}), \mathbf{W}_{\mathrm{rzf}}^{(2)}(\mathbf{H}, \alpha^{(2)})\big). \quad (20)$$

- Locally robust approach (cf. (15)):

$$\alpha_{\mathrm{LR},\mathrm{h}}^{(1)} = \underset{\alpha^{(1)} \in [0,1]}{\arg\max} \, \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}^{(1)}} \Big[R\big(\mathbf{H}, \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)}),$$

$$\mathbf{W}_{\mathrm{rzf}}^{(2)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)})\big) \Big], \quad (21)$$

$$\alpha_{\mathrm{LR},\mathrm{h}}^{(2)} = \underset{\alpha^{(2)} \in [0,1]}{\arg\max} R\big(\mathbf{H}, \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha_{\mathrm{LR},\mathrm{h}}^{(1)}), \mathbf{W}_{\mathrm{rzf}}^{(2)}(\mathbf{H}, \alpha^{(2)})\big). \quad (22)$$

- Globally robust approach (cf. (16)):

$$\alpha_{\mathrm{GR},\mathrm{h}}^{(1)} = \underset{\alpha^{(1)} \in [0,1]}{\arg\max} \, \mathbb{E}_{\mathbf{H}|\hat{\mathbf{H}}^{(1)}} \Big[R\big(\mathbf{H}, \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)}),$$

$$\mathbf{W}_{\mathrm{rzf}}^{(2)}(\mathbf{H}, \alpha^{(1)})\big) \Big], \quad (23)$$

$$\alpha_{\mathrm{GR},\mathrm{h}}^{(2)} = \underset{\alpha^{(2)} \in [0,1]}{\arg\max} R\big(\mathbf{H}, \mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha_{\mathrm{GR},\mathrm{h}}^{(1)}), \mathbf{W}_{\mathrm{rzf}}^{(2)}(\mathbf{H}, \alpha^{(2)})\big). \quad (24)$$

A. Numerical Results

In the above setting, we provide numerical results with the purpose of: *i)* evaluating the gains brought by the deterministically hierarchical D-CSI configuration over the classical non-hierarchical D-CSI setup; and *ii)* assessing the performance of the proposed lower-complexity algorithms. Under the assumption of uniform linear arrays (ULAs) at the TXs, the channel covariance matrices are constructed using the angle spread model (see, e.g., [11]): hence, for each RX k, we have $\mathbf{\Sigma}_k = \mathrm{blkdiag}(\mathbf{\Sigma}_{k1}, \mathbf{\Sigma}_{k2})$ with $\mathbf{\Sigma}_{kn} = \beta_{kn}^2 \mathbb{E}[\mathbf{a}(\theta_{kn})\mathbf{a}^{\mathrm{H}}(\theta_{kn})]$, where $\beta_{kn} > 0$ is the average attenuation of $\mathbf{h}_{kn}$, $\mathbf{a}(\theta) \in \mathbb{C}^{Mn \times 1}$ is the steering vector given by

$$\mathbf{a}(\theta) \triangleq \big[1 \ e^{-j2\pi\delta\cos\theta} \ e^{-j2\pi(M_n-1)\delta\cos\theta} \big] \quad (25)$$

with δ being the ratio between the antenna spacing at the TX and the signal wavelength, and $\theta_{kn} \in [0, \pi]$ represents the random angle of departure (AoD) between TX n and RX k. Without loss of generality, we fix $\delta = 0.5$ and assume a uniform distribution for the AoDs such that $\theta_{kn} \in [\bar{\theta}_{kn} - \Delta\theta, \bar{\theta}_{kn} + \Delta\theta]$, where $\bar{\theta}_{kn}$ is the average AoD between TX n and RX k and $\Delta\theta$ denotes the angle spread. We examine a setup with the two
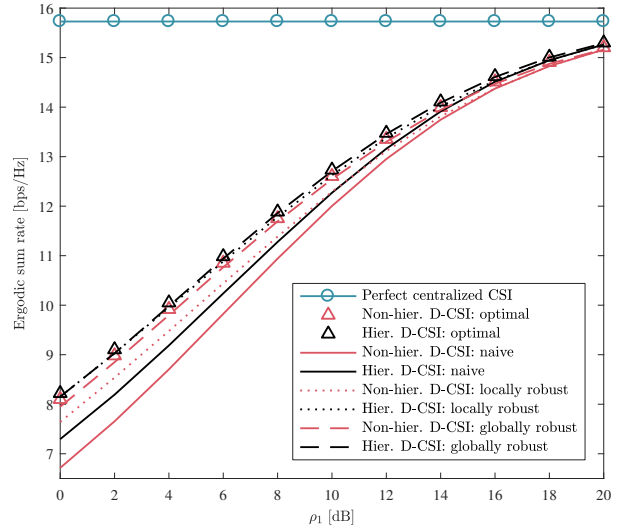

Figure 1. Ergodic sum rate VS feedback SNR of TX 1, with $P_1 = P_2 = 10$ dBW.

TXs facing each other at a distance of $d = 40$ m and $K = 5$ angularly equispaced RXs in $[\pi/4, 3\pi/4]$ between the TXs. Moreover, we consider $\beta_{kn} = r_{kn}^{-\eta/2}$, where r_{kn} represents the distance between TX n and RX k and $\eta = 2$ is the pathloss exponent. Lastly, we assume $M_1 = M_2 = 4$ for the number of transmit antennas, $\Delta\theta = \pi/8$ for the angle spread, $\sigma^2 = 0$ dBm for the noise power at the RXs, and $\{\mathbf{\Upsilon}^{(n)} = \mathbf{I}\}_{n=1}^{N}$ for the error covariance matrices at the TXs.

We analyze the ergodic sum rate as performance metric, which is computed via Monte Carlo simulations with 10^3 realizations of the channel $\mathbf{H}$ and of the channel estimate at TX 1 $\hat{\mathbf{H}}^{(1)}$. Figure 1 considers $P_1 = P_2 = 10$ dBW and plots the ergodic sum rate against the quality of the channel estimation expressed in terms of feedback SNR at TX 1, defined as $\rho_1 \triangleq (1 - \epsilon_1^2)/\epsilon_1^2$ (cf. (5)); here, the performance obtained with perfect and centralized CSI is also included for comparison. First of all, the hierarchical D-CSI setting always outperforms the non-hierarchical D-CSI counterpart in terms of ergodic rate; furthermore, in this hierarchical D-CSI setup, the locally robust approach nearly achieves the performance of the globally robust one, which is in turn remarkably close to the optimal RZF strategy (note that the same does not hold for non-hierarchical D-CSI). The performance gap between the different algorithms appears more evident from Figure 2, which illustrates the ergodic sum rate against the per-TX power constraint with $\rho_1 = 0$ dB. It is straightforward to observe that the overall performance gain brought by the hierarchical D-CSI structure becomes larger as the transmit power increases.

B. Information Exchange VS Cooperation Gain

In this section, we briefly address the following question: *what is the cost of the information exchange implied by the deterministically hierarchical D-CSI structure?* Suppose that TX 2 receives only a quantized version of the precoding submatrix $\mathbf{W}_{\mathrm{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)})$ from TX 1 via an out-of-band single-input single-output (SISO) feedback channel with bandwidth B, and that the per-TX power budget P_1 has to accommodate both
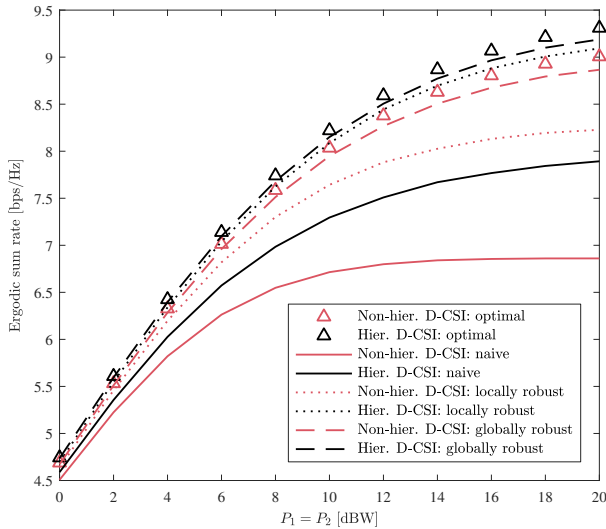
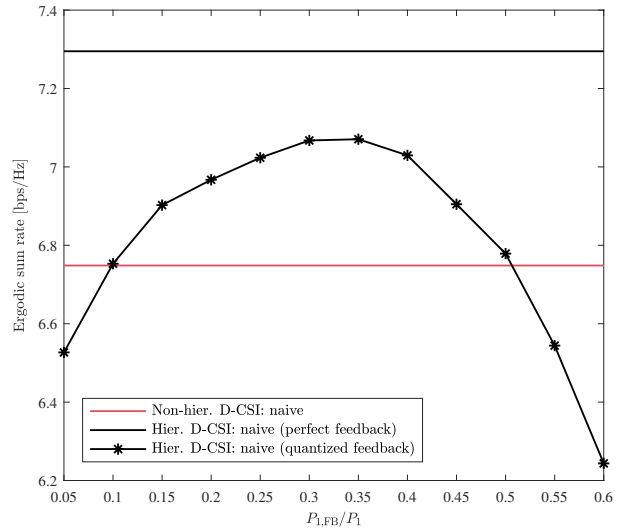Figure 2. Ergodic sum rate VS per-TX power constraint, with $\rho_1 = 0$ dB.

Figure 3. Ergodic sum rate VS feedback power, with $P_1 = P_2 = 10$ dBW, $P_{1,\text{tx}} = 10 \log_{10}(P_1 - P_{1,\text{fb}})$ dBW, and $\rho_1 = 0$ dB.

feedback and downlink transmission: in this regard, we use $P_{1,\text{fb}}$ and $P_{1,\text{tx}}$ to denote the feedback and transmit power, respectively, with $P_{1,\text{fb}} + P_{1,\text{tx}} = P_1$. Note that sharing $\mathbf{W}_{\text{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)})$ rather than $\hat{\mathbf{H}}^{(1)}$ is preferable not only to reduce the information exchange (the former is a $(M_1 \times K)$-dimensional matrix whereas the latter is $(M \times K)$-dimensional) but also to ease the computational burden at TX 2.

Denoting by ξ the number of feedback bits that can be transmitted by TX 1 during the coherence time T, let us assume that the two TXs share a common codebook $\mathcal{W} \triangleq \{\hat{\mathbf{W}}_q^{(1)}\}_{q=1}^{2^\xi}$, where each matrix $\hat{\mathbf{W}}_q^{(1)} \in \mathbb{C}^{(M \times K)}$ has $\|\hat{\mathbf{W}}_q^{(1)}\|_{\text{F}}^2 = P_1$: then, TX 1 computes $\mathbf{W}_{\text{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)})$ and sends the index

$$\hat{q} \triangleq \underset{q \in [1, 2^\xi]}{\operatorname{argmin}} \|\hat{\mathbf{W}}_q^{(1)} - \mathbf{W}_{\text{rzf}}^{(1)}(\hat{\mathbf{H}}^{(1)}, \alpha^{(1)})\|_{\text{F}} \qquad (26)$$

to TX 2, and $\sqrt{P_{1,\text{tx}}}\hat{\mathbf{W}}_{\hat{q}}^{(1)}$ is taken into account by the latter for the computation of $\mathbf{W}_{\text{rzf}}^{(2)}(\mathbf{H}, \alpha^{(2)})$. Hence, the number of feedback bits is determined by the feedback power as

$$\xi \triangleq \left\lfloor BT \log_2 \left(1 + d^{-\eta} \frac{P_{1,\text{fb}}}{\sigma^2} \right) \right\rfloor . \qquad (27)$$

Assuming $B = 1$ kHz and $T = 5$ ms, Figure 3 compares the hierarchical D-CSI setup with the non-hierarchical D-CSI counterpart using the naive approach. Interestingly, the former outperforms the latter for $P_{1,\text{fb}}/P_1 \in [0.05, 0.5]$; more specifically, the ergodic sum rate is maximized when approximately 35% of the power budget is dedicated to the feedback.

VI. CONCLUSIONS

Enforcing a hierarchical information structure is a promising solution to boost the performance of network MIMO systems in presence of distributed channel state information (D-CSI). In this paper, we formulate the general joint precoding optimization problem under D-CSI as a team decision problem and particularize it to a deterministically hierarchical D-CSI scenario. Imposing a specific structure on the precoding matrices, based on regularized zero forcing (RZF) precoding,

we propose naive, locally robust, and globally robust suboptimal strategies for the joint precoding design. Focusing on the simple case of two TXs, we show that the deterministically hierarchical D-CSI setup yields significant gains over the classical non-hierarchical D-CSI counterpart (larger gains are expected for a higher number of TXs) and that the locally and globally robust approaches perform remarkably close to the optimal RZF strategy.

REFERENCES

[1] 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.

[2] 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.

[3] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Drone small cells in the clouds: Design, deployment and performance analysis," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, Dec. 2015.

[4] R. Radner, "Team decision problems," *Ann. Math. Statist.*, vol. 33, no. 3, 1962.

[5] M. B. Shenouda and T. N. Davidson, "On the design of linear transceivers for multiuser systems with channel uncertainty," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 6, pp. 1015–1024, Aug. 2008.

[6] N. Vucic, H. Boche, and S. Shi, "Robust transceiver optimization in downlink multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3576–3587, Sept. 2009.

[7] P. de Kerret and D. Gesbert, "Quantized team precoding: A robust approach for network MIMO under general CSI uncertainties," in *Proc. IEEE Int. Workshop Signal Process. Adv. in Wireless Commun. (SPAWC)*, Edinburgh, UK, July 2016.

[8] T. R. Lakshmana, A. Tölli, and T. Svensson, "Improved local precoder design for JT-CoMP with periodical backhaul CSI exchange," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 566–569, Mar. 2016.

[9] P. de Kerret, R. Fritzsche, D. Gesbert, and U. Salim, "Robust precoding for network MIMO with hierarchical CSIT," in *Proc. IEEE Int. Symp. Wireless Commun. Syst. (ISWCS)*, Barcelona, Spain, Aug. 2014.

[10] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[11] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.