Team decision for the cooperative MIMO channel with imperfect CSIT sharing

Randa Zakhour and David Gesbert

Mobile Communications Department

Eurecom

2229 Route des Crêtes,

06560 Sophia Antipolis, France

{zakhour, gesbert}@eurecom.fr

Abstract—We consider the problem of joint MIMO precoding across multiple distant cooperating transmitters. The transmitters are assumed to be sharing user data and aim at serving a group of users in a distributed MIMO broadcast-like fashion. Among application scenarios, we find the so-called network MIMO setup. The novelty of our setup resides in the fact that each of the transmitters obtains imperfect and importantly, different, estimates of the same global multi-user channel. Despite not sharing the same vision over the CSIT, the transmitters seek to jointly act in a consistent manner in designing the precoders. This problem in facts falls in the class of so-called Team Decision Theory problems. We present some solutions to the problem of beamforming design in this case and illustrate the benefits in practical network scenarios.

I. INTRODUCTION

A major issue in several types of wireless networks is that of interference created to each other by several distant transmitters operating on the same frequency band. In the downlink cellular network example, full reuse of the spectrum across neighboring base stations triggers unacceptable levels of interference in the cell edge area. Recently, a solution to this problem has been proposed in the form of joint MIMO precoding across the distant transmitters, which should cooperatively serve the set of mobile users (so-called network MIMO, dealt with in [1], [2] for example). This situation is illustrated in Fig. 1.

In the downlink scenario, implementation of network MIMO requires both data and channel state information (CSI) to be shared by the transmitters, or to be fed back to some central processor which designs the transmission and informs the base stations of which precoding solutions shall be used. Data and CSI sharing comes, however, at a cost of delay, feedback and backhaul resources. One way to reduce the delays or have a more efficient use of the backhaul is to reduce how much CSI needs to be shared. Thus, assuming that the user data is conveniently routed to all concerned transmitters, but assuming the transmitters obtain local CSIT only, we obtain a MIMO channel with a novel CSIT model where the different transmit antennas do not have the same vision of the downlink channel. To the best of our knowledge this problem has not yet been investigated, despite its strong relevance in practical situations: in fact, previous work on multi-transmitter MIMO precoding assumes that either perfect

[1], [2] or limited CSIT [6] is available and shared among all transmitters. Two instances of such a new distributed CSIT situation are described below.

In the first instance of a feedback model, we consider that the different receivers broadcast the CSI estimates (which they have obtained over the downlink) over the air, and each transmitter attempts to decode the said information independently (see the framework proposed in [7]). Thus, depending on the distance between a receiver and each base station, a given base station may decode successfully the totality or part of the CSI feedback CSI. In another instance of partially shared CSIT, a station may decode completely the CSI feedback of a subset of users, and forward subquantized versions of it to neighboring bases. Note that both approaches lead to i) a reduction in CSI exchange, and ii) result in different representations of the same channel at the different transmitters.

Interestingly, the different transmitters must conciliate their views in order to design a consistent set of precoding vectors that will maximize a performance metric at the user side, despite possible differences in their estimated CSIT. This problem can be categorized as a so-called "team-decision problem" or a decentralized statistical decision making problem [8], [9]. More generally, in such problems, i) each decision maker (here, transmitter) has different but correlated information about the underlying uncertainty in the channel state, and ii) the transmitters need to act in a coordinated manner in order to realize the common payoff (which could be for example, the average sum rate). Such a scheme offers the possibility for reduction in communication requirements, at the expense of performance reduction [11], yet is expected to perform better than a framework where the decision makers simply ignore the differences in their view of the channel state.

Specifically, in this paper, our contributions are as follows:

- We propose a new distributed CSIT framework.
- We investigate the best transmit strategy (here beamforming) to be adopted by the transmitters in this framework.
- We study via Monte Carlo simulations the performance gap to a scenario that would have centralized CSIT.

The paper is organized as follows. We start by introducing the system model in section II. Section III provides a Bayesian formulation of the problem considered, and details it for the

2×2 case. Section IV illustrates the performance of our solution.

A. Notation

In what follows we use $\mathcal{CN}(0, a)$ to denote a complex circularly symmetric Gaussian random variable of 0 mean and variance a. Moreover, $f_{\mathbf{X}}(\mathbf{X})$ represents the joint probability distribution of the elements of matrix $\mathbf{X}$, whereas $f_{\mathbf{X}|\hat{\mathbf{X}}}(\mathbf{X}|\hat{\mathbf{X}})$ represents the joint probability distribution of the elements of $\mathbf{X}$ conditioned on those of $\hat{\mathbf{X}}$.

II. SYSTEM MODEL

Consider a set of N distant transmitters communicating with M receivers. In the downlink cellular network setup, the transmitters represent the base stations, whereas the receivers are the mobile stations. In what follows, we assume the transmitters have $N_t \geq 1$ antennas each, whereas the receivers have a single antenna each.
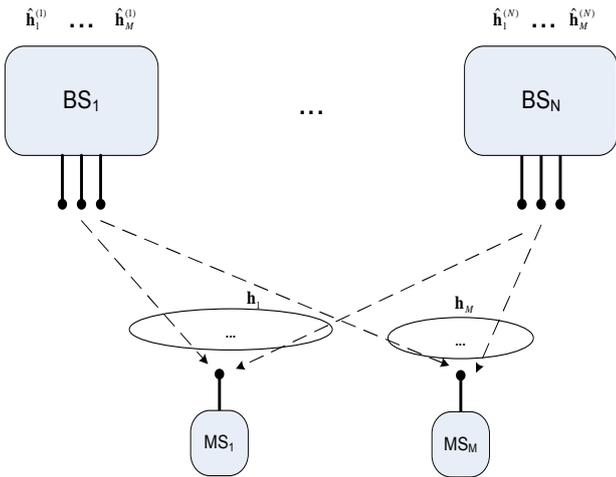
Fig. 1. Cooperative MIMO channel with imperfect CSIT sharing setup, with N base stations, M mobile stations.

Let $\mathbf{h}_{ji}$ denote the N_t-dimensional complex row vector corresponding to the channel between transmitter i and receiver j and let $\mathbf{h}_j$ be defined as $\mathbf{h}_j \triangleq [\mathbf{h}_{j1} \ \ldots \ \mathbf{h}_{jN}] \in \mathbb{C}^{MN_t \times 1}$, i.e. let it correspond to receiver j's whole channel. User channels are assumed to be independent, but in general not identically distributed. This is to cope with the fact that some users may be closer than others to certain transmitters.

The signal received at mobile station j is given by:

$$y_j = \mathbf{h}_j \mathbf{x} + n_j, \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^{NN_t \times 1}$ is the concatenated transmit signal sent by all transmitters and $n_j \sim \mathcal{CN}(0, \sigma^2)$ is the independent complex circularly symmetric AWGN noise at that receiver.

Multi-transmitter cooperative processing in the form of joint linear precoding with per-transmitter power constraints is adopted. Thus, $\mathbf{x}$ can be expressed as:

$$\mathbf{x} = \mathbf{W}\mathbf{s}, \tag{2}$$

where $\mathbf{s} \in \mathbb{C}^{M \times 1}$ is the vector of transmit symbols, its entries are assumed to be independent and $\mathcal{CN}(0, 1)$. The precoding matrix $\mathbf{W} = [\mathbf{w}_1 \ \ldots \mathbf{w}_M] \in \mathbf{C}^{NN_t \times M}$, where $\mathbf{w}_j = [\mathbf{w}_{j,1}; \ldots; \mathbf{w}_{j,N}]$ is the beamforming vector corresponding to user j's symbol, $\mathbf{w}_{j,i} \in \mathbb{C}^{N_t \times 1}$ corresponding to Transmitter i's precoding. Defining $\mathbf{V}_i$ as $[\mathbf{w}_{1,i} \ \ldots \ \mathbf{w}_{M,i}]$, i.e. as the precoding matrix used at transmitter i, $\mathbf{W}$ may be alternatively written as

$$\begin{bmatrix} \mathbf{V}_1 \\ \ldots \\ \mathbf{V}_N \end{bmatrix}. \tag{3}$$

The per-transmitter power constraint is given by:

$$\|\mathbf{V}_i\|_{\mathrm{F}}^2 \leq P, \ \forall i = 1, \ldots, N. \tag{4}$$

A. Full message sharing

As implied by the above formulation, the transmit symbols in $\mathbf{s}$ are known at all transmitters. This corresponds for instance to a situation where backhaul links pre-exist which are destined to routing the message (symbol) information of each user to the multiple base stations. Such a setup is envisioned in LTE Advanced under the name of COMP. However, as stated in the introduction, the CSI is not fully shared, and the design of the precoding will need to take this into consideration. The problem of partial sharing of the user messages is also relevant and was addressed recently in [3], [4], [5] for example. This problem is however beyond the scope of this paper. Details of the distributed CSI knowledge now follow.

B. Distributed CSIT

Previous work on multi-transmitter MIMO precoding assumes that either i) perfect CSIT is shared and available at all transmitters [1], [2] or ii) limited CSIT is available, yet common to all transmitters e.g. [6]. Here we argue that a more general and realistic setup is one where the CSI feedback is designed in such a way that different transmitters end up with different representations of the channel: for instance it is likely that users which are relatively closer to some transmitters will be able to convey more precise information about their channel state to these. The benefit of such a scheme is a reduction in signaling with respect to the scheme where all transmitters must achieve the same state of CSI knowledge, hence a greater scalability of multi-transmitter MIMO cooperation.

The distributed CSI model is shown in Figure 2, where transmitter i's knowledge of $\mathbf{h}_j$ is represented by its quantized version $\hat{\mathbf{h}}_j^{(i)}$.
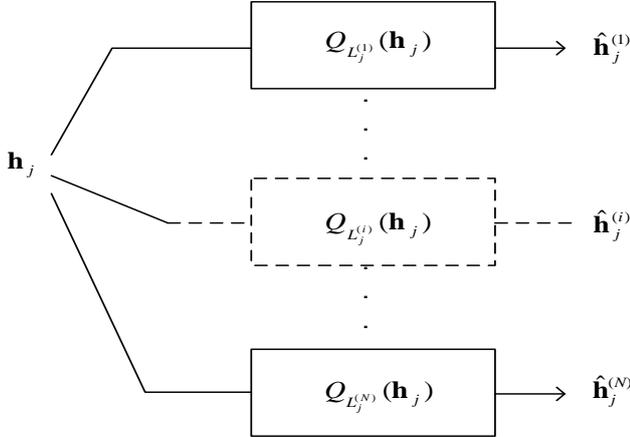
Fig. 2. Distributed CSI model: each CSI vector is seen through a different quantization filter at each base station. The quantization codebooks are designed to be hierarchical to offer additional structure.

1) Hierarchical CSI structure: To provide help in solving the problem, we propose that the true channel corresponding to user j, $\mathbf{h}_j$, be quantized using a hierarchical codebook, such that different transmitters know the channel up to different levels of said codebook. Each base station knows, in addition:

- Each user's channel statistics,
- The hierarchical codebook for each user,
- the hierarchy in their knowledge; this is detailed below.

Thus, for each $\mathbf{h}_j$, we define a degrees of accuracy mapping

$$L_j : \{1, \ldots, N\} \mapsto \{1, l_{j,max}\}, \quad (5)$$

which maps each of the transmitters to the number of bits it can decode from the feedback information sent by user j, $j = 1, \ldots, M$, in other words to its level of knowledge in user j's hierarchical codebook; $l_{j,max}$ corresponds to the most accurate level (the hierarchical codebook has $2^{l_{j,max}}$ codewords). Thus transmitter i is assumed to decode vector $\mathbf{h}_j$ up to level $L_j(i)$ of quantization accuracy yielding estimate $\hat{\mathbf{h}}_j^{(i)}$. One interesting advantage of this hierarchical information structure is that, if $L_j(i_1) > L_j(i_2)$, then transmitter i_1 knows exactly what is known by transmitter i_2, i.e. $\hat{\mathbf{h}}_j^{(i_2)}$, in addition to its own estimate $\hat{\mathbf{h}}_j^{(i_1)}$. On the other hand, transmitter i_2 does not know precisely what is decoded by transmitter i_2 however it does know that $\hat{\mathbf{h}}_j^{(i_1)}$ must belong to a certain subset of codewords located in the Voronoi region centered at $\hat{\mathbf{h}}_j^{(i_2)}$.

III. DECENTRALIZED BEAMFORMING

The N transmitters may be viewed as members of a team which need to take decisions in order to attain a common payoff, but who do not have access to the same information. Each Transmitter i chooses $\mathbf{V}_i$ based on its local CSI,

$\hat{\mathbf{H}}^{(i)} \triangleq \left[\hat{\mathbf{h}}_1^{(i)}; \ldots; \hat{\mathbf{h}}_M^{(i)} \right]$ and the extra information stemming from the hierarchical structure of the codebook. Thus the decisions at the different transmitters are based on different information; however, the performance (SINR, rate, BER for example) depends on all of these decisions. This is taken into account by the following Bayesian formulation.

A. Bayesian Formulation

We define the common goal of the considered team of N transmitters as the maximization of the expected value of a utility function, the sum rate for example. This utility function U is a function of the true channel states $\mathbf{h}_1, \ldots, \mathbf{h}_M$, as well as the decisions made at each transmitter, the beamforming vectors $\mathbf{w}_j$, since linear precoding is considered. We can write the objective function as:

$$\mathcal{U} = \mathbb{E}\left[U\left(\mathbf{H}, \mathbf{V}_1\left(\hat{\mathbf{H}}^{(1)} \right), \ldots, \mathbf{V}_N\left(\hat{\mathbf{H}}^{(N)} \right) \right) \right], \quad (6)$$

where $\mathbf{H} \triangleq [\mathbf{h}_1; \ldots; \mathbf{h}_M]$, and the dependence of the decisions at each agent (transmitter) as a function of his knowledge is made explicit. Moreover, we assume that the objective function may be decoupled as a sum of utilities over the users:

$$U\left(\mathbf{H}, \mathbf{V}_1\left(\hat{\mathbf{H}}^{(1)} \right), \ldots, \mathbf{V}_N\left(\hat{\mathbf{H}}^{(N)} \right) \right)$$
$$= \sum_{j=1}^{M} U_j\left(\mathbf{h}_j, \mathbf{V}_1\left(\hat{\mathbf{H}}^{(1)} \right), \ldots, \mathbf{V}_N\left(\hat{\mathbf{H}}^{(N)} \right) \right) \quad (7)$$

This limits the utility function within a subset of the general class of utilities, but is not too restrictive: a weighted sum rate fits into this model for example.

Restricting ourselves to deterministic decisions, in the sense that there will be a single $\mathbf{V}_i$ corresponding to each state of channel knowledge at transmitter i, $\hat{\mathbf{H}}^{(i)}$, $\mathcal{U}$ can be expanded into:

$$\mathcal{U} = \int \ldots \int d\mathbf{H} f_{\mathbf{H}}\left(\mathbf{H} \right) U\left(\mathbf{H}, \tilde{\mathbf{V}}_1\left(\mathbf{H} \right), \ldots, \tilde{\mathbf{V}}_N\left(\mathbf{H} \right) \right), \quad (8)$$

where

$$\tilde{\mathbf{V}}_i\left(\mathbf{H} \right) \triangleq \mathbf{V}_i\left(\hat{\mathbf{H}}^{(i)} \Big| \mathbf{H} \right)$$

is the beamforming strategy at transmitter i given the local knowledge at that transmitter corresponding to a true channel $\mathbf{H}$.

B. Global Optimization

A globally optimal set of beamforming decisions consists of sets of beamforming matrices $\{\mathbf{V}_i\}$, $i = 1, \ldots, N$, (one set per user, consisting of as many matrices as there are possible states of knowledge at that user), which jointly maximize $\mathcal{U}$. As stated in [8], [10] for example, it is often intractable to find the globally optimal strategies at the different team members. In such cases, a suboptimal solution may be obtained by finding strategies that are *person-by-person optimal*, as specified next.

C. Person-by-person Optimization

One can always find strategies which are *person-by-person optimal*: this corresponds to a strategy which is optimal for a given team member given that the other team members' strategies are fixed. Clearly, the globally optimal strategies are person-by-person optimal, but the converse is in general not true. In our particular setup of distributed CSIT, an optimal strategy for transmitter i, given that the other transmitters' strategies are fixed may be characterized, for a local channel knowledge equal to $\hat{\mathbf{H}}^{(i)}$, as follows:

$$\mathbf{V}_i^* \left(\hat{\mathbf{H}}^{(i)} \right)$$
$$= \arg \max_{\|\mathbf{V}_i\|_{\mathrm{F}}^2 \leq P} \int \ldots \int d\mathbf{H} f_{\mathbf{H}|\hat{\mathbf{H}}^{(i)}} \left(\mathbf{H}|\hat{\mathbf{H}}^{(i)} \right) \tilde{U}\left(\mathbf{H}, \mathbf{V}_i \right) \tag{9}$$

where

$$\tilde{U}\left(\mathbf{H}, \mathbf{V}_i \right) \triangleq U\left(\mathbf{H}, \tilde{\mathbf{V}}_1\left(\mathbf{H} \right), \ldots, \mathbf{V}_i, \ldots, \tilde{\mathbf{V}}_N\left(\mathbf{H} \right) \right). \tag{10}$$

Since $\hat{\mathbf{H}}^{(i)}$ corresponds to a quantized version of the channel, we define $\mathbf{R}\left(\hat{\mathbf{H}}^{(i)} \right)$ the Voronoi region corresponding to this state of knowledge at transmitter i:

$$f_{\mathbf{H}|\hat{\mathbf{H}}^{(i)}}\left(\mathbf{H}|\hat{\mathbf{H}}^{(i)} \right) = \begin{cases} \frac{1}{\Pr[\mathbf{R}(\hat{\mathbf{H}}^{(i)})]} f_{\mathbf{H}}\left(\mathbf{H} \right) & \mathbf{H} \in \mathbf{R}\left(\hat{\mathbf{H}}^{(i)} \right) \\ 0 & \mathbf{H} \notin \mathbf{R}\left(\hat{\mathbf{H}}^{(i)} \right) \end{cases}, \tag{11}$$

where

$$\Pr\left[\mathbf{R}\left(\hat{\mathbf{H}}^{(i)} \right) \right] = \int \ldots \int_{\mathbf{R}(\hat{\mathbf{H}}^{(i)})} d\mathbf{H} f_{\mathbf{H}}. \tag{12}$$

Thus, (9) is equivalent to:

$$\mathbf{V}_i^* \left(\hat{\mathbf{H}}^{(i)} \right)$$
$$= \arg \max_{\|\mathbf{V}_i\|_{\mathrm{F}}^2 \leq P} \int \ldots \int_{\mathbf{R}(\hat{\mathbf{H}}^{(i)})} d\mathbf{H} f_{\mathbf{H}}\left(\mathbf{H} \right) \tilde{U}\left(\mathbf{H}, \mathbf{V}_i \right). \tag{13}$$

Such a person-by-person optimization approach may be useful if the number of decisions to be determined is too large, or if the knowledge at the different transmitters does not satisfy our hierarchical assumption. For the case when $M = N = 2$, which we consider next, we formulate the problem in a way so as to try to find the globally optimal transmitter strategies.

D. Decentralized Beamforming, for $M = N = 2$

To simplify exposition of the solution to the problem, we focus on the $M = N = 2$ case. The hierarchy in the knowledge at the two transmitters, and as a result the beamforming strategies to follow, fall into one of three cases, which may be characterized as follows:

Common knowledge: In this case, $L_1(1) = L_1(2)$ and $L_2(1) = L_2(2)$. It corresponds to the *traditional* assumption under limited CSIT, where both transmitters have the same knowledge. This corresponds, for instance, to users being at the cell edge, as represented in Figure 3(a). This is equivalent to having centralized beamforming decisions being made.
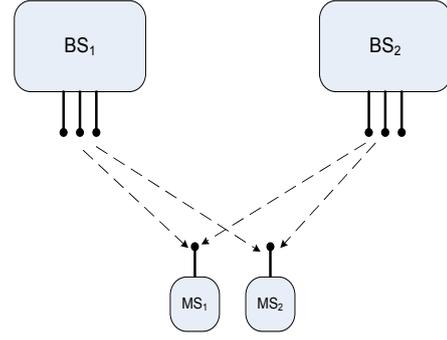
(a) Two users at edge of cell
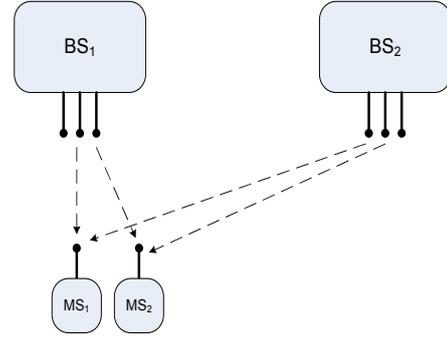
(b) Two users in same cell
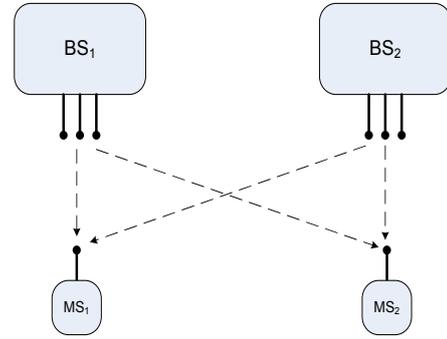
(c) Two users inside respective cells

Fig. 3. Different cell setups corresponding to different CSI hierarchies.

Degraded knowledge: In this case, $L_1(1) \geq L_1(2)$ and $L_2(1) \geq L_2(2)$, or $L_1(1) \leq L_1(2)$ and $L_2(1) \geq L_2(2)$. In other words, one of the transmitters has a better representation of both channels, and will adapt its beamforming on a finer scale than the other transmitter. Such a situation would arise, for example, if the two users being served lie in the same 'cell', as in Figure 3(b).

Symmetric knowledge: Here, $L_1(1) > L_1(2)$ and $L_2(1) < L_2(2)$, or $L_1(1) < L_1(2)$ and $L_2(1) > L_2(2)$. So one of the transmitters has a better representation of the channel of a given user and a worse one for the other user, with the

reverse occuring at the other transmitter. This corresponds, for instance, to the base stations serving users each situated within their own 'cell', as in Figure 3(c). As will be detailed below one needs to jointly optimize sets of beamforming decisions at the two transmitters corresponding to a given common coarse state of channel knowledge.

We now focus on the symmetric case where $L_1(1) > L_1(2)$ and $L_2(1) < L_2(2)$: this represents the more common setup among the ones described and shown in figure 3 above and is also the more challenging to formulate; the remaining cases can be dealt with in a similar manner. We characterize each user's quantized CSI by a pair $\mathbf{i}_1 = (i_{1,2}, i_{1,1})$ for user 1, and another $\mathbf{i}_2 = (i_{2,1}, i_{2,2})$ for user 2. The first index in each pair corresponds to the coarse knowledge (hence is shared by both users), i.e. the index of the codeword in the coarsest codebook, to which the channel is quantized, $Q_{\min_i L_j(i)}(\mathbf{h}_j)$ (see Figure 2), and the second index provides the missing bits to locate the finer codeword around the coarsest one, $Q_{\max_i L_j(i)}(\mathbf{h}_j)$. Given the structure of the distributed CSI, the beamforming matrix decisions may be parametrized in terms of these indices, so that $\mathbf{V}_1$ varies with $(\mathbf{i}_1, i_{2,1})$, whereas $\mathbf{V}_2$ is a function of $(i_{1,2}, \mathbf{i}_2)$.

Taking this into consideration, we expand (8):

$$\sum_{i_{1,2}=1}^{2^{L_1(2)}} \sum_{i_{2,1}=1}^{2^{L_2(1)}} S(i_{1,2}, i_{2,1}) \qquad (14)$$

where $S(i_{1,2}, i_{2,1})$ is given by

$$\sum_{i_{1,1}=1}^{I_1} \sum_{i_{2,2}=1}^{I_2} \int_{\mathbf{R}_1(\mathbf{i}_1)} \int_{\mathbf{R}_2(\mathbf{i}_2)} d\mathbf{h}_1 d\mathbf{h}_2$$
$$f_{\mathbf{H}}(\mathbf{H}) U(\mathbf{H}, \mathbf{V}_1(\mathbf{i}_1, i_{2,1}), \mathbf{V}_2(i_{1,2}, \mathbf{i}_2)), \qquad (15)$$

where $I_1 = 2^{L_1(1)-L_1(2)}$, $I_2 = 2^{L_2(2)-L_2(1)}$, $\mathbf{R}_1(\mathbf{i}_1)$ and $\mathbf{R}_2(\mathbf{i}_2)$ correspond to the Voronoi regions associated with the indexed codewords.

It is easy to verify that the beamforming decisions for each $S(i_{1,2}, i_{2,1})$ term may be optimized separately. For given $i_{1,2}$ and $i_{2,1}$, we optimize the corresponding $S(i_{1,2}, i_{2,1})$. To simplify notation we remove the dependence on $i_{1,2}$ and $i_{2,1}$ from the expressions. The problem is thus:

$$\max. \sum_{i_{1,1}=1}^{I_1} \sum_{i_{2,2}=1}^{I_2} \int_{\mathbf{R}_1(i_{1,1})} \int_{\mathbf{R}_2(i_{2,2})} d\mathbf{h}_1 d\mathbf{h}_2$$
$$[f_{\mathbf{H}}(\mathbf{H}) U(\mathbf{H}, \mathbf{V}_1(i_{1,1}), \mathbf{V}_2(i_{2,2}))] \qquad (16)$$
$$\text{s.t. } \|\mathbf{V}_1(i_{1,1})\|_F^2 \le P, i_{1,1} = 1, \dots, I_1 \qquad (17)$$
$$\|\mathbf{V}_2(i_{2,2})\|_F^2 \le P, i_{2,2} = 1, \dots, I_2. \qquad (18)$$

Recalling the separable nature of our utility function (refer to equation (7)), this can be reformulated as:

$$\max. \sum_{i_{1,1}=1}^{I_1} \sum_{i_{2,2}=1}^{I_2} \sum_{j=1}^{2} \Pr\left[\mathbf{R}_{\bar{j}}(i_{\bar{j},\bar{j}})\right] \int_{\mathbf{R}_j(i_{j,j})} d\mathbf{h}_j$$
$$[f_{\mathbf{h}_j}(\mathbf{h}_j) U_j(\mathbf{h}_j, \mathbf{V}_1(i_{1,1}), \mathbf{V}_2(i_{2,2}))]$$
$$\text{s.t. } \|\mathbf{V}_1(i_{1,1})\|_F^2 \le P, i_{1,1} = 1, \dots, I_1$$
$$\|\mathbf{V}_2(i_{2,2})\|_F^2 \le P, i_{2,2} = 1, \dots, I_2, \qquad (19)$$

where $\bar{j} = \mod(j, 2) + 1$ and

$$\Pr\left[\mathbf{R}_{\bar{j}}(i_{\bar{j},\bar{j}})\right] = \int_{\mathbf{R}_{\bar{j}}(i_{\bar{j},\bar{j}})} d\mathbf{h}_{\bar{j}} f_{\mathbf{h}_{\bar{j}}}(\mathbf{h}_{\bar{j}}), \qquad (20)$$

is the probability of user $\bar{j}$'s channel being quantized to the codeword indexed by the pair $(i_{\bar{j},j}, i_{\bar{j},\bar{j}})$.

1) Application to sum rate maximization: The above problem may be approximately solved via a projected gradient ascent method. Moreover, to avoid integration, we resort to approximations. As we deal with sum rate maximization in our illustrative examples, the following approximation is plugged into problem formulation (19) above:

$$\int_{\mathbf{R}_j(i_{j,j})} d\mathbf{h}_j U_j(\mathbf{h}_j, \mathbf{V}_1(i_{1,1}), \mathbf{V}_2(i_{2,2}))$$
$$= \int_{\mathbf{R}_j(i_{j,j})} d\mathbf{h}_j \log_2\left(1 + \frac{|\mathbf{h}_j \mathbf{w}_j(i_{1,1}, i_{2,2})|^2}{\sigma^2 + |\mathbf{h}_j \mathbf{w}_{\bar{j}}(i_{1,1}, i_{2,2})|^2}\right)$$
$$\approx \log_2\left(1 + \frac{\mathbf{w}_j(i_{1,1}, i_{2,2})^H \mathbf{C}_j^{(i_{j,j})} \mathbf{w}_j(i_{1,1}, i_{2,2})}{\sigma^2 + \mathbf{w}_{\bar{j}}(i_{1,1}, i_{2,2})^H \mathbf{C}_j^{(i_{j,j})} \mathbf{w}_{\bar{j}}(i_{1,1}, i_{2,2})}\right), \qquad (21)$$

where $\mathbf{C}_j^{(i_{j,j})} = \mathbb{E}\left[\mathbf{h}_j \mathbf{h}_j^H \Big| \mathbf{h}_j \in \mathbf{R}_j(i_{j,j})\right]$, and $\mathbf{w}_j(i_{1,1}, i_{2,2})$, $j = 1, 2$ is obtained from $\mathbf{V}_1(i_{1,1})$ and $\mathbf{V}_2(i_{2,2})$ by extracting the appropriate entries as defined in our system model. A similar approximation was used in [12] for example. The quality of this approximation increases and becomes asymptotically optimal with the size of the codebook.

E. Reference Schemes

Simple upper and lower bounds to the proposed schemes correspond to joint beamforming based on the more accurate (unachievable in a distributed CSIT system) and the least accurate (achievable) CSIT. Another decentralized scheme which attempts to use the local channel knowledge would be for each base station to design its transmission assuming all the other base stations share the same knowledge as itself. This is much simpler than the proposed decentralized scheme, and has similar complexity to joint beamforming design based on the coarse CSIT.

IV. NUMERICAL RESULTS

To illustrate the gains from such decentralized scheme, we show the average sum rates achieved for a symmetric $M = N = 2$, $N_t = 1$ channel, where Rayleigh fading is assumed and the covariance matrix of user 1's channel is given by $[1\ 0; 0\ \beta]$, that of user 2 by $[0\ \beta; 0\ 1]$, β being a

simulation parameter. We also vary the number of bits used for the different quantization levels.

The hierarchical codebooks are designed using Lloyd's algorithm: the coarse codebook is initially designed, then for each codeword in it, the corresponding finer codebook.

Figure 4 compares the proposed decentralized scheme to the upper and lower bounds given in III-E for $L_1(2) = L_2(1) = 2$ and $L_1(1) = L_2(2) = 6$. We label the scheme which attempts to use local channel knowledge as if it were shared 'myopic beamforming (BF)'. Thus, the upper bound scheme would require $2(L_1(1) + L_2(2)) = 24$ bits of CSIT being shared, whereas the schemes based on distributed CSIT would require $L_1(1) + L_2(2) + L_1(2) + L_2(1) = 16$ bits. The benefit of the second layer of CSI over the more coarse shared representation of the channel depends on the SNR and on the value of β. At low SNR and for β low, there is little use for the extra information. The myopic BF's performance, even though it relies on more information that the joint beamforming relying on coarse CSI, is significantly worse, highlighting the importance of coordinated action. For reference, we also plot the performance that would be obtained if the knowledge at transmitter i, $i = 1, 2$ were indeed common to both transmitters and joint beamforming would result; clearly this yields more gain that joint beamforming based on coarse CSI.

V. CONCLUSION

In this paper, the problem of cooperation in the multicell MIMO downlink under distributed CSI is formulated as a team decision problem. The solution to this problem for the two-cell two-user case is detailed and numerical results compare it to upper and lower bounds.

ACKNOWLEDGMENT

REFERENCES

[1] K. Karakayali, G. Foschini, R. Valenzuela and R. Yates, *On the Maximum Common Rate Achievable in a Coordinated Network*, in Proc. IEEE International Conference on Communications (ICC), June 2006.

[2] O. Somekh, O. Simeone, Y. Bar-Ness and A.M. Haimovich, *Distributed Multi-Cell Zero-Forcing Beamforming in Cellular Downlink Channels*, in Proc. Global Telecommunications Conference (GLOBECOM), 2006

[3] S. Shamai (Shitz), O. Simeone, O. Somekh and H.V. Poor, *Joint Multi-Cell Processing for Downlink Channels with Limited-Capacity Backhaul*, ITA 2008.

[4] P. Marsch and G. Fettweis, *On Base Station Cooperation Schemes for Downlink Network MIMO under a Constrained Backhaul*, GLOBECOM 2008, Nov.-Dec. 2008.

[5] R. Zakhour and D. Gesbert, *Optimized data sharing in multicell MIMO with finite backhaul capacity*, submitted to ISIT 2010.

[6] M. Kobayashi, M. Debbah and J. Belfiore, *Outage efficient strategies in network MIMO with partial CSIT*, in Proc. IEEE ISIT'09, 2009

[7] A. Papadogiannis, E. Hardouin and D. Gesbert, *Decentralising multi-cell cooperative processing on the downlink : a novel robust framework*, EURASIP Journal on Wireless Communications and Networking, Special Issue on Broadband Wireless Access, August 2009.

[8] Y.-C. Ho, *Team decision theory and information structures*, Proceedings of the IEEE, Vol. 68, No. 6, June 1980.
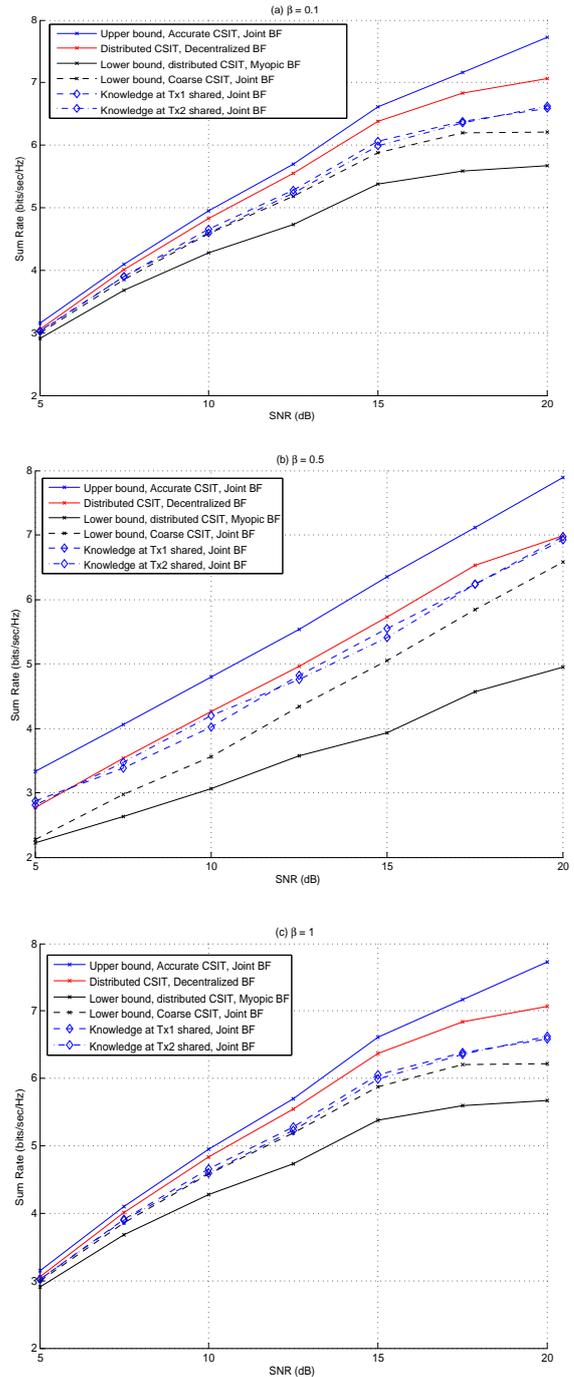
Fig. 4. Sum Rate Comparison for $L_1(2) = L_2(1) = 2$, $L_1(1) = L_2(2) = 6$ bits and different β.

[9] R. Radner, *Team Decision Problems*, Ann. Math. Statist., Vol. 33, No. 3, 1962.

[10] J. N. Tsitsiklis and M. Athans, *On the complexity of decentralized decision making and detection problems*, The 23rd IEEE Conference on Decision and Control, Dec. 1984.

[11] J.N. Tsitsiklis, *Decentralized Detection*, 1983.

[12] D. Hammarwall, M. Bengsston and B. Ottersten, *Acquiring Partial CSI for Spatially Selective Transmission by Instantaneous Channel Norm Feedback*, IEEE Transactions on Signal Processing, Vol. 56, No. 3, 2008.