

© © 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

# Distributed User-Centric Clustering and Precoding Design for CoMP Joint Transmission

Sandra Lagen\*, Adrian Agustin\*, Josep Vidal\*, Beatriz Soret<sup>†</sup>, Klaus I. Pedersen<sup>†‡</sup>

\*Dept. Signal Theory and Communications - Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

<sup>†</sup>Nokia Networks, Aalborg, Denmark, <sup>‡</sup>Aalborg University, Aalborg, Denmark

Emails: {sandra.lagen, adrian.agustin, josep.vidal}@upc.edu, {beatriz.soret, klaus.pedersen}@nokia.com

**Abstract**—This paper addresses distributed interference management in downlink of multi-cell multiple-input multiple-output (MIMO) time division duplex (TDD) networks. Each user is associated with a user-centric cluster of base stations (BSs), which cooperatively serve the user through CoMP joint transmission. Clusters of different users possibly share some BSs such that they may overlap, being coupled by interference and transmit power constraints at each BS. Our objective is the design of BSs clustering and precoding matrices per-user in order to maximize the weighted sum-rate of the system by controlling the interference and the power spent at BSs. In contrast to previous works where all channel matrices in the system are needed, we propose a distributed procedure whereby only channel matrices towards a limited number of candidate BSs per user are required while interference is still controlled by using the signal received from an uplink transmission. For an LTE-compliant dense deployment of  $2 \times 2$  MIMO BSs/users, results show gains of 6-16% in terms of sum-rate and 49-84% in terms of 5%-tile per-user rate (varying according to the maximum cluster size) as compared to distributed BS-disjoint clustering schemes.

## I. INTRODUCTION

In order to address the massive data demand in wireless networks, today's most advanced cellular networks (e.g. 3GPP LTE-A [1]) envision the use of MIMO systems, full frequency reuse, and a dense deployment of base stations (BSs). In such scenario, efficient interference management techniques are expected to play a crucial role. Among them, *coordinated multi-point joint transmission* (CoMP-JT) [2] has been shown to be a solution that significantly improves the system spectral efficiency and the cell-edge performance in low mobility conditions [3], where long channel coherence time allows a reliable estimation of the propagation channels to serving BSs.

In CoMP-JT multiple geographically separated BSs transmit data to each user, such that data has to be shared among the coordinated BSs that will jointly transmit. Although large theoretical capacity gains are obtained, in practical implementations the gains of CoMP-JT are saturated with the number of cooperating BSs due to the overhead required to acquire knowledge of the channel matrices and due to the impact of channel estimation errors [4]. For that reason, the number of BSs in which CoMP-JT can take place (called the *cluster size* or CoMP set) has to be limited. In LTE-A the maximum cluster size is 3, see [3], being suitable for classical regular deployments. However, in a dense deployment of small cells how to form the BSs clustering may not be trivial.

There are two types of BSs clustering schemes for CoMP-JT: BS-disjoint clustering and user-centric clustering. In the BS-disjoint clustering scheme, non-overlapping clusters of

BSs are formed according to the deployment and BSs in each cluster jointly serve all users within their coverage area. This way, users at the cluster-edge still suffer from considerable interference and other kinds of coordination might be needed (see [5]). Therefore, it is not the most suitable scheme from the users perspective and for dense and irregular deployments. In contrast, in the *user-centric clustering* scheme, a cluster of BSs is selected per user such that different clusters of different users might overlap. In this case, the concept of cluster-edge users disappears but not the interference in the network.

Previous works in the literature have addressed the problem of joint BSs clustering and precoding for CoMP-JT with user-centric clustering scheme assuming that all BSs in a given area are candidates to form the cluster of each user [6][7]. In [6], the problem is faced by introducing sparsity over the precoding matrices and penalizing the objective function so as to avoid larger cluster sizes. In [7], per-BS backhaul rate constraints are included in the problem in order to limit the cluster size according to the data sharing condition among BSs involved for CoMP-JT. However, in both cases the problem is solved in a centralized manner assuming that knowledge of all channel matrices from all BSs towards all users is available. In [6] distributed implementation is discussed, but it requires again knowledge of all channel matrices at BSs.

In this work we propose a distributed procedure for joint BSs clustering and precoding (DJCP) that avoids estimation of all channel matrices. Our objective is the maximization of the downlink (DL) weighted sum-rate (WSR) under per-BS power constraints (per-antenna power constraints could be imposed easily). We assume that each user selects a limited number of *candidate* BSs for its cluster and that one of them adopts the role of the BS *master* of the cluster. We decompose the maximum WSR problem into parallel subproblems to be solved at each master, which decide the BSs clustering and precoding for CoMP-JT in a coordinated manner with the rest of neighboring masters (see Fig. 1). The decomposition takes into account both the inter-cluster interference and the coupling per-BS power constraints, as one BS can be cooperating in the transmission towards different clusters of different users. To tackle interference in a distributed manner we use the *interference-cost* concept [8] and propose a procedure to acquire the interference-cost by exploiting an uplink (UL) pilot-based transmission in TDD, such that estimation of the interfering channel matrices is not needed while interference in the network can be managed. Finally, to further reduce the cluster size, a penalizing term is introduced in the formulation

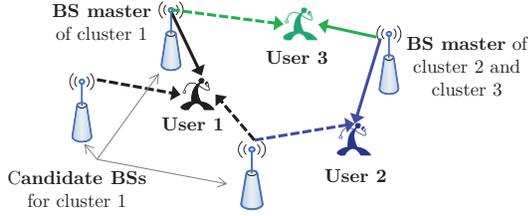


Fig. 1: Example of user-centric clustering for CoMP-JT in a cellular network of  $K=4$  BSs and  $I=3$  users. For distributed implementation, each user selects a subset of candidate BSs to form its cluster, which will be controlled by the master of that cluster.

as a weighted sum of the power spent by candidate BSs in the cluster and we propose a rule to update the weights iteratively so as to take out of the cluster those BSs using too low power. Different from previous works, [6][7], in the proposed DJCP each master only requires knowledge of the channel matrices between the user and the candidate BSs in the cluster and estimation of the interfering channel matrices is not needed either at BSs or at users. Consequently, we avoid: *i*) the computational cost associated to channel estimation, *ii*) the network planning required for pilot signals, and *iii*) the performance loss due to imperfect estimation of channel matrices that are estimated with a low signal-to-noise ratio.

## II. SYSTEM MODEL

Consider a DL multi-cell network composed of a set of  $\mathcal{K} \triangleq \{1, \dots, K\}$  BSs equipped with  $M_k$  antennas each ( $k = 1, \dots, K$ ) and a set of  $\mathcal{I} \triangleq \{1, \dots, I\}$  users with  $N_i$  antenna elements each ( $i = 1, \dots, I$ ). An example is shown in Fig. 1 for  $K=4$  BSs and  $I=3$  users. Assume that each  $i$ -th user selects a subset of potential *candidate* BSs to form its user-centric cluster, denoted by  $\mathcal{C}_i \subseteq \mathcal{K}$ . The subset of candidate BSs can be selected with different criteria: a maximum number per user can be established; or each user can choose the BSs from which the received power from the reference signals exceeds a certain threshold; or even a combination of both criteria could be adopted. Assume that one of the candidate BSs for the  $i$ -th user (e.g. the nearest BS) adopts the role of the BS *master* for the  $i$ -th cluster (see Fig. 1). The master is assumed to be connected through high-speed backhaul links with the candidate BSs in the cluster, and it will decide the BSs clustering and precoding design to serve the  $i$ -th user through CoMP-JT in coordination with the rest of masters. The final cluster of BSs for the  $i$ -th user is denoted by  $\mathcal{B}_i \subseteq \mathcal{C}_i$ .

To represent the BSs clustering and precoding in a compact form, we use network-wide matrices defined in the following. Let  $\mathbf{H}_i^k \in \mathbb{C}^{N_i \times M_k}$  denote the channel matrix between the transmitting antennas at the  $k$ -th BS and the receiving antennas at the  $i$ -th user. Let  $\mathbf{T}_i^k \in \mathbb{C}^{M_k \times m_i}$  denote the transmit precoder that the  $k$ -th BS uses to transmit  $m_i$  streams towards the  $i$ -th user. Define:

- $\mathbf{H}_i \triangleq [\mathbf{H}_i^1, \dots, \mathbf{H}_i^K] \in \mathbb{C}^{N_i \times M_T}$  as the channel matrix between the transmitting antennas at all BSs and the receiving antennas at the  $i$ -th user, being  $M_T = \sum_{k \in \mathcal{K}} M_k$ .
- $\mathbf{T}_i \triangleq [(\mathbf{T}_i^1)^T, \dots, (\mathbf{T}_i^K)^T]^T \in \mathbb{C}^{M_T \times m_i}$  as the collection of all transmit precoders intended for the  $i$ -th user.

Note that the precoders corresponding to non-candidate BSs for the  $i$ -th user are:  $\mathbf{T}_i^k = \mathbf{0}$ ,  $\forall k \notin \mathcal{C}_i$ ; while after optimization

the number of precoders among  $k \in \mathcal{C}_i$  that are different from  $\mathbf{0}$  will determine the BSs clustering (i.e.  $\mathcal{B}_i$ ).

We assume that for each  $i$ -th user only the channel matrices towards the candidate BSs (i.e.  $\mathbf{H}_i^k, \forall k \in \mathcal{C}_i$ ) are known at the master of its cluster. Note that  $\mathbf{H}_i$  contains both the candidate channels and the interfering channels (i.e.  $\mathbf{H}_i^k, \forall k \notin \mathcal{C}_i$ ).

### A. Signal model

Let  $\mathbf{x}$  denote the collection of transmitted signals of all BSs:  $\mathbf{x} \triangleq [(\mathbf{x}^1)^T, \dots, (\mathbf{x}^K)^T]^T \in \mathbb{C}^{M_T \times 1}$ , being  $\mathbf{x}^k \in \mathbb{C}^{M_k \times 1}$  the signal transmitted by the  $k$ -th BS, i.e.

$$\mathbf{x} = \sum_{i \in \mathcal{I}} \mathbf{T}_i \mathbf{b}_i, \quad \mathbf{x}^k = \sum_{i \in \mathcal{I}} \mathbf{T}_i^k \mathbf{b}_i, \quad (1)$$

where  $\mathbf{b}_i \in \mathbb{C}^{m_i \times 1}$  contains the unit power independent Gaussian symbols of the  $i$ -th user (i.e.  $\mathbf{b}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ ). The total power spent at the  $k$ -th BS is:

$$P^k = \sum_{i \in \mathcal{I}} \text{Tr}(\mathbf{T}_i^k (\mathbf{T}_i^k)^H), \quad (2)$$

where  $\text{Tr}(\cdot)$  denotes the trace operator.

Assuming narrow-band transmissions, the equivalent base-band signal  $\mathbf{y}_i \in \mathbb{C}^{N_i \times 1}$  observed at the  $i$ -th user is:

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{T}_i \mathbf{b}_i + \sum_{j \neq i, j \in \mathcal{I}} \mathbf{H}_i \mathbf{T}_j \mathbf{b}_j + \mathbf{v}_i, \quad (3)$$

where  $\mathbf{v}_i$  refers to the additive zero-mean white Gaussian noise with distribution  $\mathbf{v}_i \sim \mathcal{CN}(\mathbf{0}, \sigma_i^2 \mathbf{I})$ . The symbols are estimated at the  $i$ -th user assuming that interference is treated as noise and that a linear receive filter  $\mathbf{R}_i \in \mathbb{C}^{N_i \times m_i}$  is applied:

$$\hat{\mathbf{b}}_i = \mathbf{R}_i^H \mathbf{y}_i. \quad (4)$$

The mean square error (MSE) for the symbols transmitted towards the  $i$ -th user can be expressed through the so-called *MSE-matrix*  $\mathbf{E}_i = \mathbb{E}_{\mathbf{b}, \mathbf{v}}[(\hat{\mathbf{b}}_i - \mathbf{b}_i)(\hat{\mathbf{b}}_i - \mathbf{b}_i)^H] \in \mathbb{C}^{m_i \times m_i}$ , being  $\mathbb{E}[\cdot]$  the expectation operator. Under the independence assumption of  $\{\mathbf{b}_i\}_{\forall i}$  and  $\mathbf{v}_i$ , the MSE-matrix results:

$$\mathbf{E}_i = \mathbf{I} + \mathbf{R}_i^H \mathbf{C}_{\mathbf{y}_i} \mathbf{R}_i - \mathbf{R}_i^H \mathbf{H}_i \mathbf{T}_i - \mathbf{T}_i^H \mathbf{H}_i^H \mathbf{R}_i, \quad (5)$$

where  $\mathbf{C}_{\mathbf{y}_i} = \mathbb{E}_{\mathbf{b}, \mathbf{v}}[\mathbf{y}_i \mathbf{y}_i^H] \in \mathbb{C}^{N_i \times N_i}$  is the covariance matrix of the received signal at the  $i$ -th user:

$$\mathbf{C}_{\mathbf{y}_i} = \mathbf{H}_i \mathbf{T}_i \mathbf{T}_i^H \mathbf{H}_i^H + \mathbf{N}_i = \sum_{j \in \mathcal{I}} \mathbf{H}_i \mathbf{T}_j \mathbf{T}_j^H \mathbf{H}_i^H + \sigma_i^2 \mathbf{I}, \quad (6)$$

and  $\mathbf{N}_i$  is the covariance matrix of the received interference-plus-noise signal at the  $i$ -th user. The receive filter that minimizes the MSE, i.e.  $\text{Tr}(\mathbf{E}_i)$ , is the MMSE receiver [9]:

$$\mathbf{R}_i^{\text{mmse}} = \mathbf{C}_{\mathbf{y}_i}^{-1} \mathbf{H}_i \mathbf{T}_i, \quad (7)$$

such that the MSE-matrix in (5) with  $\mathbf{R}_i^{\text{mmse}}$  in (7) results:

$$\mathbf{E}_i^{\text{mmse}} = \mathbf{I} - \mathbf{T}_i^H \mathbf{H}_i^H \mathbf{C}_{\mathbf{y}_i}^{-1} \mathbf{H}_i \mathbf{T}_i = (\mathbf{I} + \mathbf{T}_i^H \mathbf{H}_i^H \mathbf{N}_i^{-1} \mathbf{H}_i \mathbf{T}_i)^{-1}. \quad (8)$$

The achievable rate of the  $i$ -th user is:

$$R_i = \log_2 |\mathbf{I} + \mathbf{H}_i \mathbf{T}_i \mathbf{T}_i^H \mathbf{H}_i^H \mathbf{N}_i^{-1}| = -\log_2 |\mathbf{E}_i^{\text{mmse}}|, \quad (9)$$

being  $\mathbf{N}_i$  shown in (6) and  $|\cdot|$  the determinant operator.

### B. Problem formulation

With the objective of maximizing the total DL weighted sum-rate (WSR) of the system with a maximum power constraint per BS as well as further reducing the cluster size if desirable, the joint BSs clustering and precoding are obtained as the solution to the following problem (inspired by [6]):

$$\begin{aligned}
\text{(P0)} : & \underset{\{\mathbf{T}_i\}_{\forall i}}{\text{maximize}} \sum_{i \in \mathcal{I}} \left( \mu_i R_i - \alpha \sum_{k \in \mathcal{C}_i} \gamma_i^k \text{Tr}(\mathbf{T}_i^k (\mathbf{T}_i^k)^H) \right) \\
& \text{subject to} \sum_{i \in \mathcal{I}} \text{Tr}(\mathbf{T}_i^k (\mathbf{T}_i^k)^H) \leq P_{\max}^k \quad \forall k, \\
& \mathbf{T}_i^k = \mathbf{0} \quad \forall k \notin \mathcal{C}_i, \forall i,
\end{aligned} \tag{10}$$

where  $\mu_i$  is a weighting coefficient associated to the priority of the  $i$ -th user,  $R_i$  is the achievable rate shown in (9), and  $P_{\max}^k$  is the available transmit power at the  $k$ -th BS.  $\alpha \geq 0$  is a parameter that trades-off between maximum WSR performance and minimum weighted power spent at BSs: if  $\alpha = 0$  it reduces to plane maximum WSR, while if  $\alpha > 0$  the power spent at the BSs is taken into account with the goal of reducing the cluster size. In addition to [6], we introduce the parameter  $\gamma_i^k$  (per user and per BS), which will allow taking out of the cluster of the  $i$ -th user those BSs devoting little power towards that user (see details in Section III-D).

Due to interference, problem (P0) in (10) is not convex on  $\{\mathbf{T}_i\}_{\forall i}$  and the optimal solution cannot be guaranteed. Nevertheless, one solution attaining a local optimum of (P0) in (10) can be obtained by solving the minimization of the total weighted sum of MSEs [9] while still keeping the minimization of the cluster size in the problem:

$$\begin{aligned}
\text{(P1)} : & \underset{\{\mathbf{T}_i\}, \{\mathbf{R}_i\}, \{\mathbf{W}_i\}_{\forall i}}{\text{minimize}} \sum_{i \in \mathcal{I}} \left( \text{Tr}(\mathbf{W}_i \mathbf{E}_i) - \mu_i \log_2 \left| \frac{\ln(2)}{\mu_i} \mathbf{W}_i \right| \right. \\
& \quad \left. + \alpha \sum_{k \in \mathcal{C}_i} \gamma_i^k \text{Tr}(\mathbf{T}_i^k (\mathbf{T}_i^k)^H) \right) \tag{11} \\
& \text{subject to} \sum_{i \in \mathcal{I}} \text{Tr}(\mathbf{T}_i^k (\mathbf{T}_i^k)^H) \leq P_{\max}^k \quad \forall k, \\
& \mathbf{T}_i^k = \mathbf{0} \quad \forall k \notin \mathcal{C}_i, \forall i,
\end{aligned}$$

where  $\mathbf{W}_i \in \mathbb{C}^{m_i \times m_i}$  is a weighting matrix associated to the  $i$ -th user and  $\mathbf{E}_i$  corresponds to the MSE-matrix shown in (5). Problem (P1) in (11) is convex for each set of variables ( $\{\mathbf{T}_i\}_{\forall i}$ ,  $\{\mathbf{R}_i\}_{\forall i}$ , and  $\{\mathbf{W}_i\}_{\forall i}$ ) separately. Therefore, a block coordinate descent (BCD) approach [10] can be followed to find a local optimum by alternating the optimization between  $\{\mathbf{T}_i\}_{\forall i}$ ,  $\{\mathbf{R}_i\}_{\forall i}$ , and  $\{\mathbf{W}_i\}_{\forall i}$  if all channel matrices are known. The attained solution is a local optimal solution to problem (P0) in (10), see [6][9]. However, this procedure requires all channel matrices from all BSs to all users to be collected.

### III. DISTRIBUTED JOINT CLUSTERING AND PRECODING

We focus on solving (P1) in (11) in a distributed manner. To do so, we split the problem in such a way that each master solves a subproblem in coordination with other masters. More specifically, first, problem (P1) in (11) is decomposed into two problems: (P2) <sub>$i$</sub>  to be solved at the master of the  $i$ -th cluster so as to find the optimal BSs clustering and precoding (included

in  $\mathbf{T}_i$ ) when  $\mathbf{R}_i$  and  $\mathbf{W}_i$  are fixed, and (P3) <sub>$i$</sub>  to be solved at the  $i$ -th user so as to find  $\mathbf{R}_i$  and  $\mathbf{W}_i$  when  $\mathbf{T}_i$  is fixed. Second, how to acquire the required parameters for distributed design is detailed. Finally, the iterative algorithm is presented, subsuming the acquisition of the required parameters at the master of each cluster and the simultaneous optimizations.

Let us define the following parameters that will allow us to decompose problem (P1) in (11).

- The *interference-cost matrix* [5] is defined as:

$$\mathbf{Y}_i = \sum_{j \neq i, j \in \mathcal{I}} \mathbf{H}_j^H \mathbf{R}_j \mathbf{W}_j \mathbf{R}_j^H \mathbf{H}_j, \tag{12}$$

which reflects the DL interference that could be created by the  $i$ -th cluster towards unintended users (i.e.  $j \neq i$ ,  $j \in \mathcal{I}$ ) and it is seen as a penalizing term for the design of BSs clustering and precoding at the  $i$ -th cluster (see next problem in (16)). Let us partition  $\mathbf{Y}_i$  in (12) as:

$$\mathbf{Y}_i = \begin{bmatrix} \mathbf{Y}_i[1, 1] & \cdots & \mathbf{Y}_i[1, K] \\ \vdots & \ddots & \vdots \\ \mathbf{Y}_i[K, 1] & \cdots & \mathbf{Y}_i[K, K] \end{bmatrix}, \tag{13}$$

such that (as  $\mathbf{H}_j = [\mathbf{H}_j^1, \dots, \mathbf{H}_j^K]$ ):

$$\mathbf{Y}_i[k, l] = \sum_{j \neq i, j \in \mathcal{I}} (\mathbf{H}_j^k)^H \mathbf{R}_j \mathbf{W}_j \mathbf{R}_j^H \mathbf{H}_j^l. \tag{14}$$

In Section III-A we show that not all blocks in (14) are needed to control interference, and Section III-C1 explains how to obtain the needed ones without having to estimate every single interfering channel matrix.

- The power spent by the  $k$ -th BS involved in the  $i$ -th cluster towards users in other clusters ( $j \neq i$ ) is:

$$P_{-i}^k = \sum_{j \neq i, j \in \mathcal{I}} \text{Tr}(\mathbf{T}_j^k (\mathbf{T}_j^k)^H), \tag{15}$$

such that the per-BS power constraint in (11) can be decoupled as:  $\text{Tr}(\mathbf{T}_i^k (\mathbf{T}_i^k)^H) + P_{-i}^k \leq P_{\max}^k$ .

#### A. Optimization at the master of each cluster

Given  $\{P_{-i}^k\}_{\forall i, \forall k \in \mathcal{C}_i}$ ,  $\{\mathbf{R}_i\}_{\forall i}$ , and  $\{\mathbf{W}_i\}_{\forall i}$ , problem (P1) in (11) can be decomposed into  $I$  parallel optimization problems (one per master). The problem to be solved at the master of the  $i$ -th cluster (considering only terms in (11) that are affected by  $\mathbf{T}_i$ ) for given  $\{P_{-i}^k\}_{\forall k \in \mathcal{C}_i}$ ,  $\{\mathbf{R}_i\}_{\forall i}$ , and  $\{\mathbf{W}_i\}_{\forall i}$ , is:

$$\begin{aligned}
\text{(P2)}_i : & \underset{\mathbf{T}_i}{\text{minimize}} \quad f_i + g_{-i} \tag{16} \\
& \text{subject to} \quad \text{Tr}(\mathbf{T}_i^k (\mathbf{T}_i^k)^H) + P_{-i}^k \leq P_{\max}^k \quad \forall k \in \mathcal{C}_i, \\
& \mathbf{T}_i^k = \mathbf{0} \quad \forall k \notin \mathcal{C}_i,
\end{aligned}$$

where  $f_i$  measures the impact over the served  $i$ -th user:

$$\begin{aligned}
f_i = & \text{Tr}(\mathbf{W}_i \mathbf{R}_i^H \mathbf{H}_i \mathbf{T}_i \mathbf{T}_i^H \mathbf{H}_i^H \mathbf{R}_i) - \text{Tr}(\mathbf{W}_i \mathbf{R}_i^H \mathbf{H}_i \mathbf{T}_i) \\
& - \text{Tr}(\mathbf{W}_i \mathbf{T}_i^H \mathbf{H}_i^H \mathbf{R}_i) + \alpha \sum_{k \in \mathcal{C}_i} \gamma_i^k \text{Tr}(\mathbf{T}_i^k (\mathbf{T}_i^k)^H), \tag{17}
\end{aligned}$$

while  $g_{-i}$  in (16) takes into account the generated interference towards unintended users ( $j \neq i$ ) (see (12)):

$$g_{-i} = \text{Tr}(\mathbf{Y}_i \mathbf{T}_i \mathbf{T}_i^H). \tag{18}$$

The main part of the algorithm is to find the optimal BSs clustering and precoding (both included in  $\mathbf{T}_i$ ) from (P2<sub>*i*</sub>). Problem (P2<sub>*i*</sub>) in (16) is convex on  $\mathbf{T}_i$  and the optimal structure for  $\mathbf{T}_i$  could be directly obtained by deriving its Lagrangian function. However, as we are interested in controlling the cluster size, we have to work with the blocks of the network-wide matrix  $\mathbf{T}_i = [(\mathbf{T}_i^1)^H, \dots, (\mathbf{T}_i^K)^H]^H$ , where only the blocks corresponding to candidate BSs selected by the *i*-th user have to be optimized (i.e.  $\mathbf{T}_i^k, \forall k \in \mathcal{C}_i$ ). Accordingly, as only these block matrices of  $\mathbf{T}_i$  can be different from  $\mathbf{0}$ , the blocks of  $\mathbf{Y}_i$  in (13) that are needed for distributed optimization at the master of the *i*-th cluster are those  $\mathbf{Y}_i[k, l]$  such that  $k \in \mathcal{C}_i$  and  $l \in \mathcal{C}_i$  (see (18)). Note that problem (P2<sub>*i*</sub>) in (16) is formulated in a way such that only knowledge of  $\mathbf{Y}_i[k, l]_{\forall k, l \in \mathcal{C}_i}$ ,  $\{P_{-i}^k\}_{\forall k \in \mathcal{C}_i}$ ,  $\mathbf{R}_i$ , and  $\mathbf{W}_i$  is required. So from now on, let us assume they are given at the master of the *i*-th cluster (which is equivalent but less restrictive to assume that  $\{P_{-i}^k\}_{\forall k \in \mathcal{C}_i}$ ,  $\{\mathbf{R}_i\}_{\forall i}$ , and  $\{\mathbf{W}_i\}_{\forall i}$  are given).

Interestingly, problem (P2<sub>*i*</sub>) in (16) is separable among the precoders of different BSs, such that a BCD method [10] can be applied with  $\{\mathbf{T}_i^k\}_{\forall k \in \mathcal{C}_i}$  as block variables. To do so, we follow similar steps as in [6] but extended to the multi-stream case per user and applied to solve problem (P2<sub>*i*</sub>) in (16).

Define the following two sets of variables:

$$\mathbf{J}_i \triangleq \mathbf{H}_i^H \mathbf{R}_i \mathbf{W}_i \mathbf{R}_i^H \mathbf{H}_i, \quad \mathbf{D}_i \triangleq \mathbf{H}_i^H \mathbf{R}_i \mathbf{W}_i. \quad (19)$$

Partition  $\mathbf{J}_i \in \mathbb{C}^{M_T \times M_T}$  and  $\mathbf{D}_i \in \mathbb{C}^{M_T \times m_i}$  in (19) as:

$$\mathbf{J}_i = \begin{bmatrix} \mathbf{J}_i[1, 1] & \cdots & \mathbf{J}_i[1, K] \\ \vdots & \ddots & \vdots \\ \mathbf{J}_i[K, 1] & \cdots & \mathbf{J}_i[K, K] \end{bmatrix}, \quad \mathbf{D}_i = \begin{bmatrix} \mathbf{D}_i[1] \\ \vdots \\ \mathbf{D}_i[K] \end{bmatrix}. \quad (20)$$

It is important to emphasize here that:

$$\mathbf{J}_i[k, l] = (\mathbf{H}_i^k)^H \mathbf{R}_i \mathbf{W}_i \mathbf{R}_i^H \mathbf{H}_i^l, \quad \mathbf{D}_i[k] = (\mathbf{H}_i^k)^H \mathbf{R}_i \mathbf{W}_i. \quad (21)$$

According to the definitions and partitions in (20) and (13), the gradient of the Lagrangian function of problem (P2<sub>*i*</sub>) in (16) ( $\mathcal{L}$ ) with respect to  $\mathbf{T}_i^k$  for  $k \in \mathcal{C}_i$  is:

$$\begin{aligned} \nabla_{\mathbf{T}_i^k} \mathcal{L} = & \mathbf{J}_i[k, k] \mathbf{T}_i^k + \sum_{l \neq k, l \in \mathcal{C}_i} \mathbf{J}_i[k, l] \mathbf{T}_i^l - \mathbf{D}_i[k] + \lambda_k \mathbf{T}_i^k \\ & + \alpha \gamma_i^k \mathbf{T}_i^k + \mathbf{Y}_i[k, k] \mathbf{T}_i^k + \sum_{l \neq k, l \in \mathcal{C}_i} \mathbf{Y}_i[k, l] \mathbf{T}_i^l, \end{aligned} \quad (22)$$

where  $\lambda_k$  is a non-negative dual variable associated to the *k*-th per-BS power constraint in (16). Then, by equating (22) to  $\mathbf{0}$  we obtain the precoding structure  $\mathbf{T}_i^k$  for  $k \in \mathcal{C}_i$  (being  $\mathbf{0}$  otherwise):

$$\mathbf{T}_i^k = (\mathbf{J}_i[k, k] + \mathbf{Y}_i[k, k] + (\lambda_k + \alpha \gamma_i^k) \mathbf{I})^{-1} \mathbf{F}_i^k \quad \text{if } k \in \mathcal{C}_i, \quad (23)$$

$$\mathbf{T}_i^k = \mathbf{0} \quad \text{if } k \notin \mathcal{C}_i, \quad (24)$$

where

$$\mathbf{F}_i^k = \mathbf{D}_i[k] - \sum_{l \neq k, l \in \mathcal{C}_i} \mathbf{J}_i[k, l] \mathbf{T}_i^l + \sum_{l \neq k, l \in \mathcal{C}_i} \mathbf{Y}_i[k, l] \mathbf{T}_i^l. \quad (25)$$

The precoder in (23) is coupled with the precoders of BSs in the *i*-th cluster (i.e.  $\{\mathbf{T}_i^l\}_{\forall l \neq k, l \in \mathcal{C}_i}$ ). So the solution for joint BSs clustering and precoding is achieved at the master of the

cluster of the *i*-th user by applying a BCD method among block variables  $\{\mathbf{T}_i^k\}_{\forall k \in \mathcal{C}_i}$  until a stop condition (e.g. convergence or maximum number of iterations achieved) for given  $\mathbf{R}_i$ ,  $\mathbf{W}_i$ ,  $\mathbf{Y}_i[k, l]_{\forall k, l \in \mathcal{C}_i}$  and  $\{P_{-i}^k\}_{\forall k \in \mathcal{C}_i}$ . Algorithm 1 summarizes the BCD method among block variables  $\{\mathbf{T}_i^k\}_{\forall k \in \mathcal{C}_i}$ . Subsection III-C details how to acquire the required parameters at each master.

As it can be observed in (23)-(25) and (21), for given  $\mathbf{Y}_i[k, l], \forall k, l \in \mathcal{C}_i$ , the BSs clustering and precoding design can be performed in a distributed manner at each master by having knowledge only of the channel matrices towards the candidate BSs selected by the user (i.e.  $\mathbf{H}_i^k, \forall k \in \mathcal{C}_i$ ).

---

**Algorithm 1** Clustering and Precoding design at the master of the *i*-th cluster for given  $\{\mathbf{Y}_i[k, l]\}_{\forall k, l \in \mathcal{C}_i}$ ,  $\{P_{-i}^k\}_{\forall k \in \mathcal{C}_i}$ ,  $\mathbf{R}_i$ , and  $\mathbf{W}_i$

---

- 1: Compute  $\mathbf{J}_i[k, l]$  and  $\mathbf{D}_i[k]$  using (21)  $\forall k, l \in \mathcal{C}_i$
  - 2: **initialize precoders:**  $\mathbf{T}_i^k, \forall k \in \mathcal{C}_i$
  - 3: **iterate** (cyclically pick a candidate BS  $k \in \mathcal{C}_i$ ):
  - 4:   - Compute  $\mathbf{T}_i^k$  using (23) with  $\mathbf{J}_i[k, l], \forall l \in \mathcal{C}_i, \mathbf{D}_i[k], \mathbf{R}_i, \mathbf{W}_i$ , and  $\mathbf{Y}_i[k, l], \forall l \in \mathcal{C}_i$  (compute  $\lambda_k$  with the bisection method [11] such that  $\text{Tr}(\mathbf{T}_i^k (\mathbf{T}_i^k)^H) + P_{-i}^k \leq P_{\max}^k$ )
  - 5: **until stop condition**
- 

## B. Optimization at each user

Given all the precoding matrices,  $\{\mathbf{T}_i\}_{\forall i}$ , problem (P1) in (11) can be easily decomposed into *I* parallel optimization problems (one per user), where the optimal receive filter ( $\mathbf{R}_i$ ) and the optimal weighting matrix ( $\mathbf{W}_i$ ) for the *i*-th user are obtained as the solution to:

$$(P3_i) : \underset{\mathbf{R}_i, \mathbf{W}_i}{\text{minimize}} \quad \text{Tr}(\mathbf{W}_i \mathbf{E}_i) - \mu_i \log_2 \left| \frac{\ln(2)}{\mu_i} \mathbf{W}_i \right|. \quad (26)$$

The optimal receive filter  $\mathbf{R}_i$  to (P3<sub>*i*</sub>) in (26) is given by the MMSE receive filter  $\mathbf{R}_i^{\text{mmse}}$  in (7). As it is done in real deployments [12], each user can compute  $\mathbf{R}_i^{\text{mmse}}$  in (7) based on the estimation of the equivalent channel  $\mathbf{H}_i \mathbf{T}_i$  and  $\mathbf{C}_{\mathbf{Y}_i}$ . It is not needed to estimate the interfering channels to get  $\mathbf{C}_{\mathbf{Y}_i}$  in (6), as it can be evaluated by averaging  $\mathbf{y}_i \mathbf{y}_i^H$  [12].

Once the receive filter  $\mathbf{R}_i$  is designed, the optimal weighting matrix  $\mathbf{W}_i$  to (P3<sub>*i*</sub>) in (26) is given by [9]:

$$\mathbf{W}_i = \frac{\mu_i}{\ln(2)} (\mathbf{E}_i^{\text{mmse}})^{-1}. \quad (27)$$

So each user can compute  $\mathbf{W}_i$  in (27) based on  $\mathbf{H}_i \mathbf{T}_i$  and  $\mathbf{C}_{\mathbf{Y}_i}$ , i.e. the same information needed to compute  $\mathbf{R}_i^{\text{mmse}}$  in (7).

## C. Acquisition of parameters at the master of each cluster

1) *Acquisition of the interference-cost matrix  $\mathbf{Y}_i$* : Similarly as [5], we can get an estimate of the interference-cost matrix for the user-centric clustering scheme by using an UL pilot-based transmission. More specifically, the blocks needed for distributed optimization from  $\mathbf{Y}_i$  in (14) can be estimated from the covariance matrix of the interference-plus-noise received signal in UL when: *i*) channel reciprocity is assumed (as in TDD systems) and *ii*) users in UL transmit with a specific pilot signal that is precoded as a function of the receive filter  $\mathbf{R}_i$  and the weighting matrix  $\mathbf{W}_i$ . This way, we avoid the complex task associated to the estimation of the most harmful interfering channel matrices that would be

needed to compute  $\Upsilon_i$  by following (12) and we avoid also the reporting of all receive filters and weighting matrices to non-serving BSs (i.e.  $\mathbf{R}_j$  and  $\mathbf{W}_j$ ,  $\forall j \neq i$ , see (12)). As pointed out before, the blocks of  $\Upsilon_i$  in (13) that are needed for distributed optimization at the master of the  $i$ -th cluster are those such that  $\Upsilon_i[k, l], \forall k, l \in \mathcal{C}_i$  (see (18)). The procedure to obtain them is described in the following.

From an UL pilot-based transmission in TDD, the covariance matrix of the received interference-plus-noise signal at the BSs ( $\forall k \in \mathcal{C}_i$ ) of the  $i$ -th cluster in UL  $\bar{\mathbf{N}}_i \in \mathbb{C}^{M_{U_i} \times M_{U_i}}$  (being  $M_{U_i} = \sum_{k \in \mathcal{C}_i} M_k$ ) is:

$$\bar{\mathbf{N}}_i = \sum_{j \neq i, j \in \mathcal{I}} \bar{\mathbf{H}}_j^i \bar{\mathbf{T}}_j^i \bar{\mathbf{T}}_j^{iH} (\bar{\mathbf{H}}_j^i)^H + \bar{\sigma}_i^2 \mathbf{I}, \quad (28)$$

where  $\bar{\mathbf{T}}_j^i \in \mathbb{C}^{N_j \times m_j}$  denotes the UL precoder used at the  $j$ -th user,  $\bar{\sigma}_i^2$  the UL noise power, and  $\bar{\mathbf{H}}_j^i$  contains the channel matrices from the  $j$ -th user towards the BSs in the  $i$ -th cluster stacked as:  $\bar{\mathbf{H}}_j^i = [\mathbf{H}_j^k, \dots, \mathbf{H}_j^l]^T, \forall k, l \in \mathcal{C}_i$ . Let us partition  $\bar{\mathbf{N}}_i$  as we did for the interference-cost matrix  $\Upsilon_i$  in (13), but the partition of  $\bar{\mathbf{N}}_i$  only contains the blocks of those candidate BSs in the cluster of the  $i$ -th user. Let us maintain the index of the BSs for referencing the blocks, such that ( $\forall k, l \in \mathcal{C}_i$ ):

$$\bar{\mathbf{N}}_i[k, l] = \sum_{j \neq i, j \in \mathcal{I}} (\mathbf{H}_j^k)^T \bar{\mathbf{T}}_j^i \bar{\mathbf{T}}_j^{iH} (\mathbf{H}_j^l)^* + \bar{\sigma}_i^2 \mathbf{I}. \quad (29)$$

If the UL precoder is designed according to:

$$\bar{\mathbf{T}}_j^i = \sqrt{F} \mathbf{R}_j^* (\mathbf{W}_j^{\frac{1}{2}})^*, \quad (30)$$

being  $F < 1$  a scaling cell-wide factor that allows meeting the UL power constraint (see further details in [5]) and  $\mathbf{W}_j = \mathbf{W}_j^{\frac{1}{2}} (\mathbf{W}_j^{\frac{1}{2}})^H$ , then the blocks of  $\Upsilon_i$  in (13) that are needed for distributed optimization at the master of the  $i$ -th cluster (i.e.  $\Upsilon_i[k, l], \forall k, l \in \mathcal{C}_i$ ) can be estimated as:

$$\hat{\Upsilon}_i[k, l] = F^{-1} (\bar{\mathbf{N}}_i[k, l])^*. \quad (31)$$

As it is shown in [5], the estimation errors in  $\hat{\Upsilon}_i$  are negligible when properly selecting the scaling factor  $F$  in (30). Note that it is not needed to estimate the interfering channels to get  $\bar{\mathbf{N}}_i$  in (28) because it can be estimated by averaging the UL received signal if high-speed backhaul links connect the BSs within each cluster. Further, it is not needed to decode the UL transmitted symbols, so an UL pilot-based transmission is enough to get  $\hat{\Upsilon}_i$ . In LTE-A we could use the already defined sounding reference signals [2] with UL power control.

2) *Acquisition of the power spent by candidate BSs in the cluster towards other clusters*  $\{P_{-i}^k\}_{\forall k \in \mathcal{C}_i}$ : We assume that the master of each  $i$ -th cluster collects from the candidate BSs that form the  $i$ -th cluster the power that they use towards other clusters, i.e.  $P_{-i}^k, \forall k \in \mathcal{C}_i$ . So, only exchange of control information with neighboring BSs is required.

3) *Acquisition of  $\mathbf{R}_i$  and  $\mathbf{W}_i$* : We assume that  $\mathbf{R}_i$  and  $\mathbf{W}_i$  are reported from the  $i$ -th user towards the master of its  $i$ -th cluster through an UL feedback link. As compared to [6] where feedback links are required from each user towards all BSs in the network to report  $\mathbf{R}_i$  and  $\mathbf{W}_i$ , in our case only one feedback link is needed per-user towards its BS master.

#### D. Design of the weighting power coefficients $\gamma_i^k$

The introduction of the weighting power coefficients in (P0) in (10) aims at the reduction of the cluster size by taking out of the cluster those candidate BSs that use too low power. So  $\gamma_i^k$  should decrease with the power spent by the  $k$ -th BS towards the  $i$ -th user. A suitable selection is the following:

$$\gamma_i^k = \frac{1}{\tau + \text{Tr}(\mathbf{T}_i^k (\mathbf{T}_i^k)^H)}, \quad (32)$$

where  $\tau > 0$  is a small constant factor to avoid  $\gamma_i^k \rightarrow \infty$  at any point in the iterations and  $\text{Tr}(\mathbf{T}_i^k (\mathbf{T}_i^k)^H)$  is the power spent by the  $k$ -th BS to the  $i$ -th user in the previous iteration.

#### E. Algorithm

Algorithm 2 summarizes the iterative procedure to solve (P1) in (11) in a distributed manner. First of all, each  $i$ -th user selects the candidate BSs to form its cluster (i.e.  $\mathcal{C}_i$ ) and among them the BS being the master of its cluster (line 1), which acquires the channel matrices from candidate BSs in the cluster towards the user (i.e.  $\{\mathbf{H}_i^k\}_{\forall k \in \mathcal{C}_i}$ ) (line 2). Then, the algorithm starts from an initialization of the precoders  $\{\mathbf{T}_i^k\}_{\forall k \in \mathcal{C}_i}, \forall i$ , that satisfies the per-BS power constraints in (16) (line 3). For simulation purposes, a suitable initialization is obtained by solving (P2<sub>*i*</sub>) in (16) using  $\Upsilon_i = \mathbf{0}$ . Then, a DL transmission is carried out using  $\{\mathbf{T}_i^k\}$  (line 3) where users can evaluate the covariance matrix of the received signal  $\mathbf{C}_{y_i}$  and update  $\mathbf{R}_i$  and  $\mathbf{W}_i$  as shown in Section III-B (line 4).

---

#### Algorithm 2 DJCP to solve (P1) in (11)

---

- 1: All users ( $\forall i$ ): select set of candidate BSs,  $\mathcal{C}_i$ , one of them being the master of the  $i$ -th cluster
  - 2: All masters ( $\forall i$ ): acquire knowledge of the channel matrix from candidate BSs towards the  $i$ -th user,  $\{\mathbf{H}_i^k\}_{\forall k \in \mathcal{C}_i}$
  - 3: All masters ( $\forall i$ ): initialize  $\{\mathbf{T}_i^k\}_{\forall k \in \mathcal{C}_i}$  and transmit in DL
  - 4: All users ( $\forall i$ ): compute  $\mathbf{R}_i$  in (7) and  $\mathbf{W}_i$  in (27)
  - 5: **iterate**:
  - 6:   **# Acquisition of parameters**
  - 7:   - All users ( $\forall i$ ): transmit in UL a pilot signal properly precoded as in (30) using  $\mathbf{R}_i$  and  $\mathbf{W}_i$ , such that each master acquires  $\Upsilon_i[k, l], \forall k, l \in \mathcal{C}_i$  as in (31)
  - 8:   - All masters ( $\forall i$ ): acquire power spent by BSs in the cluster towards other clusters,  $P_{-i}^k, \forall k \in \mathcal{C}_i$
  - 9:   - All users ( $\forall i$ ): report  $\mathbf{R}_i$  and  $\mathbf{W}_i$  to the master of the  $i$ -th cluster
  - 10:   **# Simultaneous optimizations at each master**
  - 11:   - All masters ( $\forall i$ ): do **Algorithm 1** for fixed  $\{\Upsilon_i[k, l]\}_{\forall k, l \in \mathcal{C}_i}, \{P_{-i}^k\}_{\forall k \in \mathcal{C}_i}, \mathbf{R}_i$ , and  $\mathbf{W}_i$
  - 12:   - All masters ( $\forall i$ ): update weights  $\{\gamma_i^k\}_{\forall k \in \mathcal{C}_i}$  as in (32)
  - 13:   - All BSs ( $\forall k$ ): transmit in DL with  $\{\mathbf{T}_i^k\}_{\forall i}$
  - 14:   **# Simultaneous optimizations at each user**
  - 15:   - All users ( $\forall i$ ): compute  $\mathbf{R}_i$  in (7) and  $\mathbf{W}_i$  in (27)
  - 16: **until stop condition**
- 

After the initialization, the iterative procedure is implemented assuming a TDD synchronized system and alternate UL/DL transmissions. First, the required parameters for distributed optimization are acquired at each master, as detailed in Section III-C (i.e.  $\{\Upsilon_i[k, l]\}_{\forall k, l \in \mathcal{C}_i}$  from the UL,  $\{P_{-i}^k\}_{\forall k \in \mathcal{C}_i}$  through inter-BS signaling, and  $\mathbf{R}_i$  and  $\mathbf{W}_i$  through feedback) (lines 7-9). Next, the simultaneous BSs clustering and precoding designs are performed at the

master of each  $i$ -th cluster for DL transmission, as shown in Section III-A (line 11). Let us recall that after the per-cluster optimization the precoders have to be scaled so as to strictly satisfy the per-BS power constraint as, due to the uncoupling of the per-BS transmit power, in a given iteration the constraint could be violated. Then, the weighting power coefficients  $\{\gamma_i^k\}_{\forall k \in \mathcal{C}_i}$  are updated (line 12). Finally, DL transmission is performed (line 13) such that the optimization at users can be done as shown in Section III-B (line 15).

Monotonic convergence of Algorithm 2 can be proven when all parameters are perfectly acquired,  $\gamma_i^k$  is fixed,  $\forall i, \forall k$ , and the optimizations at each cluster are performed in a sequential manner. This can be demonstrated thanks to the special convex properties of problem (P1) in (11) and by following a similar rationale as in [5] (omitted here due to space limit). If all clusters perform the optimization simultaneously, monotonic convergence cannot be guaranteed due to the coupling of the per-BS power constraints among clusters. However, even with simultaneous optimizations, convergence is consistently observed in simulations. Convergence is also observed when including the rule to update  $\gamma_i^k$  in (32).

#### IV. SIMULATION RESULTS

The network scenario consists of a dense synchronized TDD deployment of  $K$  small cells (that act as BSs) in a concentrated area, following specifications for Scenario 2a in [13]. BSs are randomly placed within a circle of 50 m radius with a minimum distance of 20 m among them, and  $I$  users are randomly placed in a concentric 70 m radius circle. All BSs operate on the same carrier frequency at 3.5 GHz with 10 MHz bandwidth. ITU Urban Micro model with 3D distance is used for path loss and shadowing modeling. The antenna pattern is omnidirectional and the transmit power is 24 dBm. Noise spectral density is -174 dBm/Hz. The number of antennas is  $M_k = 2, \forall k$ , and  $N_i = 2, \forall i$  (such that  $m_i = 2, \forall i$ ). Full-load traffic model is adopted. Two different criteria are used: sum-rate (SR) with  $\mu_i = 1, \forall i$ , in (10) and proportional fair (PF) where  $\mathbf{W}_i = R_i^{-1}(\mathbf{E}_i^{\text{mmsc}})^{-1}$  is used in (27) (see [6]). Results are averaged over 100 random deployments.

First, we consider a network of  $K = 8$  BSs and  $I = 8$  users and study the performance of the proposed DJCP when varying the number of candidate BSs selected by the users (i.e.  $|\mathcal{C}_i|$ ) for  $\alpha = 0$  in (10). Note that the larger is the value of  $|\mathcal{C}_i|$ , the more channel matrices have to be estimated.  $K=8$  BSs is used so as to compare DJCP with the distributed 'BS-disjoint clustering' scheme in [5] using disjoint clusters of either 2 BSs or 4 BSs, in which inter-cluster interference management is adopted and only channel matrices towards the BSs in the cluster are needed (as in the proposed DJCP). Fig. 2 shows the convergence of DJCP in a given random deployment for different values of  $|\mathcal{C}_i|$  (1, 2, 3, 4, 8) and SR criterion. Convergence is consistently achieved.

Fig. 3 and Fig. 4 display the sum-rate performance and the 5%-tile of the per-user rates, respectively, versus the number of candidate BSs for SR and PF criteria. As benchmark we use the 'WMMSE' algorithm [9] for the broadcast channel, where the 8 BSs form a giant virtual transmitter that serves all UEs, including per-BS power constraints and knowledge

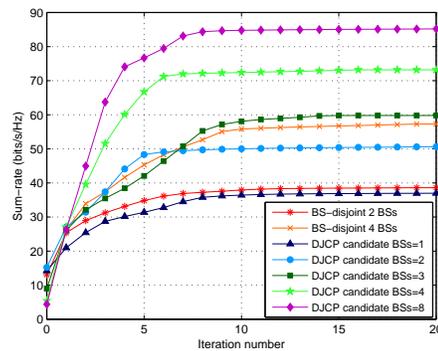


Fig. 2: Convergence of DJCP in terms of sum-rate for different number of candidate BSs selected per user ( $|\mathcal{C}_i|=1, 2, 3, \text{ or } 4$ ).  $K=8, I=8, \alpha=0$ . SR as utility function.

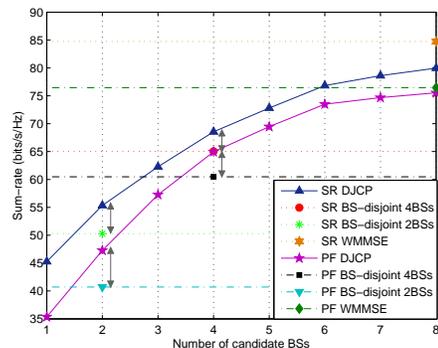


Fig. 3: Sum-rate (in bits/s/Hz) vs. the number of candidate BSs selected per user ( $|\mathcal{C}_i|$ ).  $K=8, I=8, \alpha=0$ . SR and PF as utility function.

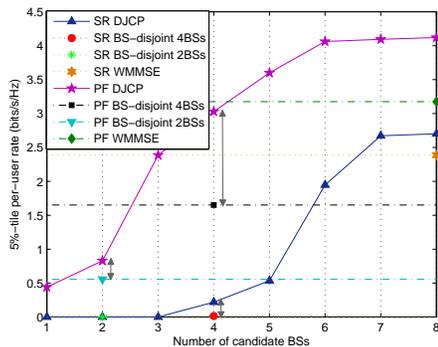
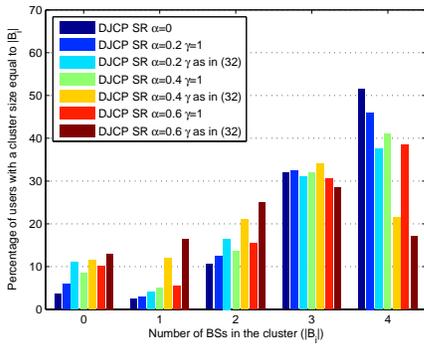


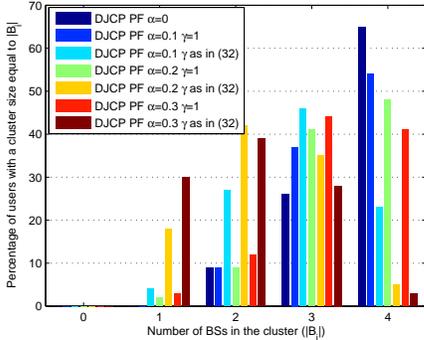
Fig. 4: 5%-tile of the per-user achievable rate (in bits/s/Hz) vs. the number of candidate BSs selected per user ( $|\mathcal{C}_i|$ ).  $K=8, I=8, \alpha=0$ . SR and PF as utility function.

of all channel matrices. By comparing the distributed schemes with the same number of estimated channel matrices (2 BSs or 4 BSs, marked with gray arrows in figures), it can be observed that DJCP outperforms in terms of sum-rate and 5%-tile rate the distributed BS-disjoint clustering for both SR and PF criteria. The gains are: 8% (PF-4BSs), 16% (PF-2BSs), 6% (SR-4BSs), 10% (SR-2BSs) in sum-rate and 84% (PF-4BSs), 49% (PF-2BSs) in 5%-tile rate. The gain is specially remarkable in the 5%-tile rate, where 'PF DJCP' with 3 candidate BSs already outperforms 'PF BS-disjoint 4BSs' and, in the case of 4 BSs in the cluster, 'SR DJCP' gets a 5%-tile rate of 0.218 bits/s/Hz while BS-disjoint clustering gets 0.013 bits/s/Hz (see Fig. 4). This is due to the high flexibility that DJCP offers for dense and irregular deployments of BSs.

Second, we consider a network of  $K = 10$  BSs (as specified in [13]) and  $I = 10$  users and study the performance of DJCP for different values of  $\alpha$  in (10) when the number of candidate



(a) SR as utility function.



(b) PF as utility function.

Fig. 5: Percentage of users (%) having a cluster size ( $|\mathcal{B}_i^k|$ ) of 0, 1, 2, 3, or 4 BSs, for different DJCP optimization cases and SR or PF criterion.  $K=10$ ,  $I=10$ ,  $|\mathcal{C}_i|=4$ .

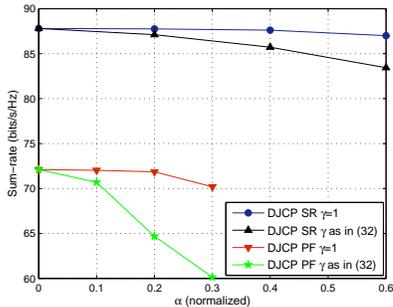


Fig. 6: Sum-rate vs.  $\bar{\alpha}$ .  $K=10$ ,  $I=10$ ,  $|\mathcal{C}_i|=4$ . SR and PF as utility function.

BSs selected by the users is  $|\mathcal{C}_i| = 4$ . When  $\alpha > 0$  we simulate two cases:  $\gamma_i^k = 1, \forall i, k$ , and  $\gamma_i^k$  updated as in (32) using  $\tau = 0.1$ . We normalize  $\alpha$  in (10) by the maximum power at BSs ( $P$ ):  $\alpha = \bar{\alpha}/P$ , and results are shown in terms of  $\bar{\alpha}$ .

Fig. 5 illustrates the percentage of users having a cluster size of 0, 1, 2, 3, or 4 BSs for SR (Fig. 5.a) and PF (Fig. 5.b) criteria. Results are shown when using different values of  $\bar{\alpha}$  for (P0) in (10):  $\bar{\alpha} = 0, 0.2, 0.4$ , or  $0.6$  for SR and  $\bar{\alpha} = 0, 0.1, 0.2$ , or  $0.3$  for PF. Different values of  $\bar{\alpha}$  are considered for SR and PF criteria as the utility function is not of the same magnitude. It can be observed that increasing  $\bar{\alpha}$  reduces the cluster size. The proposed rule to update the parameter  $\gamma_i^k$  in (32) per user and per BS allows further reduction of the cluster size as compared to  $\gamma_i^k = 1$  both for SR and PF criteria. In the case of SR, some users are assigned to a cluster size of 0 BSs, which means they are not scheduled. This effect does not appear with PF criterion (see Fig. 5.b).

Fig. 6 shows the sum-rate versus  $\bar{\alpha}$  when using  $\gamma_i^k = 1$  or  $\gamma_i^k$  updated as in (32) and either SR or PF criterion. In the

case of SR, the performance degradation is very mild, while the number of BSs per cluster is significantly reduced with the proposed design of  $\gamma_i^k$  in (32) (see Fig. 5.a). In the case of PF, the sum-rate is very sensitive to  $\bar{\alpha}$  when  $\gamma_i^k$  is designed as in (32). However, it is worth comparing the case  $\bar{\alpha} = 0.3$  and  $\gamma_i^k = 1$  with respect to  $\bar{\alpha} = 0.1$  and  $\gamma_i^k$  as in (32): both lead to similar sum-rate while the number of BSs per cluster is significantly reduced in the later case (see Fig. 5.b).

## V. CONCLUSIONS

This article proposes a distributed joint BSs clustering and precoding for CoMP-JT in DL of multi-cell MIMO TDD systems. A user-centric clustering scheme is used, being the cluster size limited to a maximum number of candidate BSs selected by the user. The problem is distributed among clusters by uncoupling the per-BS power constraints and using the UL received signal as a way of reporting the impact of interference. This way, only channel matrices towards candidate BSs need to be reported. Further, an additional term is included in the problem so as to reduce the cluster size by taking out of the cluster those candidate BSs that use too low power. Results in dense deployments of BSs show significant gains as compared to distributed BS-disjoint clustering schemes and an effective reduction of the cluster size with the proposed design.

## VI. ACKNOWLEDGMENT

This work has been supported by projects TROPIC FP7 ICT-2011-8-318784 (European Commission), FPU12/00828 grant (Ministerio Educación, Cultura y Deporte, Spanish Government), DISNET TEC2013-41315-R (Ministerio Economía y Competitividad, Spanish Government and ERDF), and 2014SGR-60 (Catalan Government).

## REFERENCES

- [1] P. Bhat, S. Nagata, L. Campoy, I. Berberana, T. Derham, G. Liu, X. Shen, P. Zong, and J. Yang, "LTE-Advanced: An Operator Perspective," *IEEE Communications Magazine*, vol. 50, pp. 104–114, Feb. 2012.
- [2] 3GPP TR 36.819, *Coordinated multi-point operation for LTE physical layer aspects*, Release 11, v11.2.0, Sep. 2013.
- [3] J. Lee et al., "Coordinated Multipoint Transmission and Reception in LTE-Advanced Systems," *IEEE Communication Magazine*, vol. 50, pp. 44–50, Nov. 2012.
- [4] A. Lozano, J. G. Andrews, and R. W. Heath, "Spectral Efficiency Limits in Pilot-Assisted Cooperative Communications," *IEEE Int. Symp. on Information Theory Proceedings*, pp. 1132–1136, Jul. 2012.
- [5] S. Lagen, A. Agustín, and J. Vidal, "Distributed Inter-Cluster Interference Management for CoMP-based Cellular Networks," *IEEE Global Communications Conference*, pp. 4204–4209, Dec. 2013.
- [6] M. Hong, R. Sun, H. Baligh, and Z.-Q. Luo, "Joint Base Station Clustering and Beamformer Design for Partial Coordinated Transmission in Heterogeneous Networks," *IEEE J. Sel. Areas in Commun.*, vol. 31, pp. 226–240, Feb. 2013.
- [7] B. Dai and W. Yu, "Sparse Beamforming and User-Centric Clustering for Downlink Cloud Radio Access Network," *IEEE Access: Special section on recent advances in C-RAN*, vol. 2, pp. 1326–1339, Oct. 2014.
- [8] C. Shi, D. A. Schmidt, R. A. Berry, M. L. Honig, and W. Utschick, "Distributed Interference Pricing for the MIMO Interference Channel," in *IEEE Int. Conf. on Communications*, Jun. 2009.
- [9] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel," *IEEE Trans. on Signal Processing*, vol. 59, pp. 4331–4340, Sep. 2011.
- [10] D. P. Bertsekas, *Nonlinear Programming*. 2nd ed. Belmont, MA: Athena Scientific, 1999.
- [11] S. Boyd and L. Vandenbergue, *Convex optimization*. Cambridge, 2004.
- [12] 3GPP TR 36.829, *Enhanced performance requirement for LTE User Equipment (UE)*, Release 11, v11.1.0, Dec. 2012.
- [13] 3GPP TR 36.872, *Small cell enhancements for E-UTRA and E-UTRAN - Physical layer aspects*, Release 12, v12.1.0, Dec. 2013.