Decentralized Joint Precoding for WSRMax with Pilot Aided Beamformer Estimation

Jarkko Kaleva†, Antti Tölli†, Markku Juntti†, Randall Berry‡ and Michael Honig‡

†Centre for Wireless Communications, University of Oulu P.O. Box 4500, 90014 University of Oulu, Finland
‡Northwestern University, Dept. of Elect. Engineering and Comp. Science, Evanston, IL, 60208, U.S.
Email: jarkko.kaleva@oulu.fi

Abstract—Downlink weighted sum rate maximizing beamformer design is considered for joint processing (JP) coordinated multi-point (CoMP) transmission. Global channel state information exchange, required by the centralized JP CoMP processing, is in many scenarios impractical due to the backhaul latency and capacity requirements. Low overhead decentralized processing enables JP even with limited backhaul capacity. The proposed best response (BR) design also takes into account the possibly non-orthogonal pilot design and noisy pilot estimation. By allowing spatially overlapping pilot sequences, the transceiver processing requires only the channel state information of locally served users. This enables more flexible pilot design and makes the proposed approach realizable in practical time correlated and noisy channel conditions. Furthermore, the proposed BR algorithm provides low signaling overhead for systems with limited backhaul capacity. Robustness for fading channel conditions and limited pilot resources is shown by numerical examples.

I. INTRODUCTION

Spatial interference coordination and coordinated multiple-input multiple-output (MIMO) transceiver design are key concepts in deployment of modern dense and heterogeneous wireless systems. Single cell MIMO beamforming is already part of current standards [1]. Furthermore, the basic operation of multi-cell MIMO transceiver processing has been preliminarily covered by the Long Term Evolution Advanced (LTE-A) standards. However, there are still numerous practical limitations in inter-cell backhaul connectivity that are hindering efficient implementation of advanced joint processing (JP) coordinated multi-point (CoMP) schemes, where each user is served with the same data by multiple base stations (BSs) simultaneously. Conventionally, in cloud radio access network (C-RAN) systems, JP is done in a central unit (CU) assuming global channel state information (CSI) [2].

In this paper, we focus on JP CoMP with weighted sum rate maximization (WSRMax) in a downlink (DL) scenario, where the backhaul limitations and delay constraints of the changing channel conditions prevent accurate CSI sharing and beamformer training within the channel coherence window. Thus, in order to utilize JP, the BSs need to perform independent beamformer design within the JP clusters based on locally available CSI. This motivates us to design decentralized transceiver processing with limited training overhead that is robust in the presence of fading channel conditions. Still, we assume that the transmitted data can be shared among the serving BSs. The data is assumed to be queued and prioritized by the CU, which then distributes it to the serving BSs. Furthermore, we assume non-orthogonal and noisy pilot estimation, which is expected in dense deployments. The performance of the proposed method is evaluated in a cellular multi-user network with a time correlated channel model.

WSRMax for multi-cell systems has been extensively studied with respect to decentralized inter-cell interference coordination [3], [4]. The weighted minimum mean-squared error (WMMSE) method solves the WSRMax problem via an equivalent weighted mean-squared error (MSE) minimization problem [3]. The WMMSE has been shown to have a convenient structure for decentralized processing in cellular time division duplexing (TDD) MIMO systems [3], [4]. A more general best response (BR) framework, which allows straightforward parallel processing for varying performance objectives, was proposed in [5]. Unfortunately, JP inherently couples the beamformers among the cooperating BSs. Thus, the aforementioned decentralized coordinated beamforming (CB) methods cannot be used as is. In [6], heuristic JP CoMP schemes were provided depending only on local CSI. However, since the beamforming schemes are solely based on locally available CSI (without any CSI sharing), they cannot utilize all available degrees of freedom (DoF).

In centralized JP, the backhaul information can be done by *data sharing*, where the CU exchanges the data and joint beamformers separately [7]. Another approach is *compression*, where only the compressed version of the analog beamformer is informed to the BSs [8]. Sparse joint beamformer designs have been proposed in an effort to limit the backhaul signaling overhead [9]. These designs try to limit the JP cluster sizes and, thus, implicitly reduce the backhaul overhead. The impact of partial or imperfect CSI feedback to CB and JP has been studied, e.g., in [10]. Pilot contamination in TDD based transceiver training for coordinated beamforming has been considered in [11]. The direct least squares (LS) beamformer estimation from the contaminated uplink (UL)/DL pilots was shown to provide good performance as opposed to estimating the channels separately and explicitly constructing the beamformers from the estimated channels.

In the sequel, we show that, by carefully designed backhaul and pilot signaling, decentralized JP WSRMax is feasible with low signaling overhead even in moderately fast fading channel conditions. Estimating the interference channels separately in

noisy and dense JP systems causes significant issues, when the pilot sequences are not fully orthogonal. Estimating the beamformers directly from the overlapping pilot training sequences provides more robust approach and improved performance. We assume a data sharing scenario, where the transmit beamformers are locally designed in each BS based on locally available CSI, while some reduced information is exchanged over the backhaul.

II. SYSTEM MODEL

We consider a multi-cell system with B BSs each equipped with N_T transmit antennas. There are, in total, K user equipments (UEs) each with N_R receive antennas. Each UE $k = 1, \ldots, K$ is coherently served by $|\mathcal{B}_k|$ BSs, where the set $\mathcal{B}_k$ defines the JP cluster (set of phase-coherently serving BSs) for UE k. Similarly, the set of UE indices served by BS $b = 1, \ldots, B$ is denoted by $\mathcal{C}_b = \{k | b \in \mathcal{B}_k\}$. The set of all UE indices is given by $\mathcal{K} = \{1, \ldots, K\}$. The maximum number of spatial data streams allocated to UE $k = 1, \ldots, K$ is denoted by $L_k \leq \min(|\mathcal{B}_k| N_T, N_R)$. To simplify the notation in various places, we use the following set abbreviations: $(k,l) \triangleq \{(k,l) | k \in \mathcal{K}, l = 1, \ldots, L_k\}$ and $(b,k,l) \triangleq \{(b,k,l) | k \in \mathcal{K}, l = 1, \ldots, L_k, b \in \mathcal{B}_k\}$. The system model is illustrated in Fig. 1.
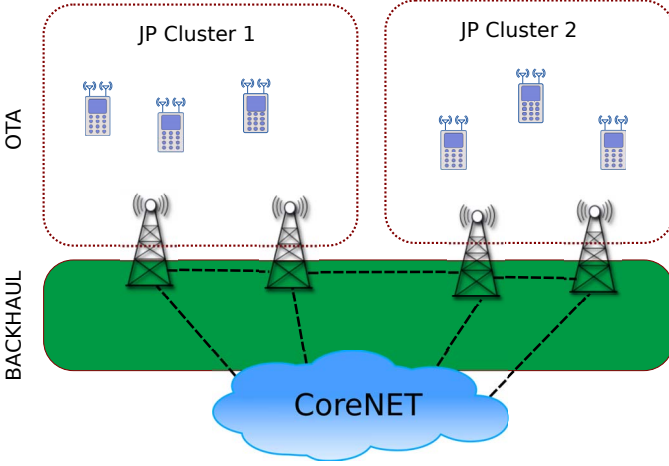

Fig. 1. Simplified system with 4 BSs and 6 UEs in two JP clusters.

Without loss of generality, the DL transmission within the JP set is considered to be symbol synchronous in the sense that the transmitted symbols from $\mathcal{B}_k$ are coherently combined at each UE. Only the local CSI knowledge is assumed, that is, each BS $b = 1, \ldots, B$ is only aware of the channel matrix $\mathbf{H}_{b,k} \in \mathbb{C}^{N_R \times N_T} \ \forall \ k = 1, \ldots, K$, the data sharing is assumed within each serving set $\mathcal{B}_k$. Furthermore, we assume TDD, which is used to exchange the effective UL/DL CSI.

The received signal at UE $k = 1, \ldots, K$ is given as

$$\mathbf{y}_k = \sum_{i=1}^{K} \sum_{b \in \mathcal{B}_i} \sum_{j=1}^{L_i} \mathbf{H}_{b,k} \mathbf{m}_{b,i,j} d_{i,j} + \mathbf{n}_k, \quad (1)$$

where $\mathbf{m}_{b,i,j} \in \mathbb{C}^{N_T}$ is the beamformer vector for the jth spatial data stream for UE i from BS b and $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma_k^2 \mathbf{I})$ denotes the receiver noise. The complex data symbols $d_{k,l}, k = 1, \ldots, K, l = 1, \ldots, L_k$ are assumed to be independent and identically distributed (i.i.d.) with $\mathbb{E}\{|d_{k,l}|^2\} = 1$.

The estimated symbol at UE over stream l, after the applying receive beamformer $\mathbf{u}_{k,l} \in \mathbb{C}^{N_R}$, is given as $\hat{d}_{k,l} = \mathbf{u}_{k,l}^H \mathbf{y}_k$. The resulting signal-to-interference-plus-noise ratio (SINR) is

$$\Gamma_{k,l} = \frac{|\sum_{b \in \mathcal{B}_k} \mathbf{u}_{k,l}^H \mathbf{H}_{b,k} \mathbf{m}_{b,k,l}|^2}{\sum_{i=1}^{K} \sum_{\substack{j=1, \\ (i,j) \neq (k,l)}}^{L_i} |\sum_{b \in \mathcal{B}_i} \mathbf{u}_{k,l}^H \mathbf{H}_{b,k} \mathbf{m}_{b,i,j}|^2 + \|\mathbf{u}_{k,l}\|^2 \sigma_k^2}, \quad (2)$$

and the corresponding MSE is

$$
\begin{aligned}
\epsilon_{k,l} &\triangleq \mathbb{E}[|d_{k,l} - \hat{d}_{k,l}|^2] \\
&= |\sum_{b \in \mathcal{B}_k} \mathbf{u}_{k,l}^H \mathbf{H}_{b,k} \mathbf{m}_{b,k,l} - 1|^2 + \|\mathbf{u}_{k,l}\|^2 \sigma_k^2 + \\
&\quad \sum_{i=1}^{K} \sum_{\substack{j=1, \\ (i,j) \neq (k,l)}}^{L_i} |\sum_{b \in \mathcal{B}_i} \mathbf{u}_{k,l}^H \mathbf{H}_{b,k} \mathbf{m}_{b,i,j}|^2.
\end{aligned} \quad (3)
$$

Note that (3) is a convex function in terms of the transmit and receive beamformers but not jointly convex in both.

III. PROBLEM FORMULATION & CENTRALIZED SOLUTION

We consider WSRMax subject to BS-specific sum transmit power constraints. The general problem can be stated as

$$
\begin{aligned}
\max_{\mathbf{u}_{k,l}, \mathbf{m}_{b,k,l}} \quad & \sum_{k=1}^{K} \sum_{l=1}^{L_k} \mu_k \log_2(1 + \Gamma_{k,l}) \\
\text{s. t.} \quad & \sum_{k \in \mathcal{C}_b}^{K} \sum_{l=1}^{L_k} \|\mathbf{m}_{b,k,l}\|^2 \leq P_b, \ b = 1, \ldots, B,
\end{aligned} \quad (4)
$$

where $\mu_k, k = 1, \ldots, K$ are the user priority weights. The problem is non-convex and known to be difficult [12]. The optimal receive beamformers for (4) are the minimum mean-squared error (MMSE) receivers

$$\mathbf{u}_{k,l} = \mathbf{K}_k^{-1} \left(\sum_{b \in \mathcal{B}_k} \mathbf{H}_{b,k} \mathbf{m}_{b,k,l} \right), \quad (5)$$

where $\mathbf{K}_k = \sum_{i=1}^{K} \sum_{j=1}^{L_i} \sum_{b \in \mathcal{B}_i} \mathbf{H}_{b,k} \mathbf{m}_{b,i,j} \mathbf{m}_{b,i,j}^H \mathbf{H}_{b,k}^H + \mathbf{I} \sigma_k^2$.

It is well-known that, when the MMSE receive beamformers are applied, there is an inverse relation between the SINR and the corresponding MSE [3]

$$\epsilon_{k,l}^{-1} = 1 + \Gamma_{k,l}. \quad (6)$$

Now, applying (6) to (4) we can formulate the weighted sum rate maximization problem as

$$
\begin{aligned}
\min_{\mathbf{u}_{k,l}, \mathbf{m}_{b,k,l}} \quad & \sum_{k=1}^{K} \sum_{l=1}^{L_k} \mu_k \log_2(\epsilon_{k,l}) \\
\text{s. t.} \quad & \sum_{k \in \mathcal{C}_b}^{K} \sum_{l=1}^{L_k} \|\mathbf{m}_{b,k,l}\|^2 \leq P_b, \ b = 1, \ldots, B.
\end{aligned} \quad (7)
$$

Since (7) is not jointly convex for the transmit and receive beamformers, we alternate between solving for the transmit and receive beamformers. This is, particularly, convenient for TDD processing as the DL and UL transmissions are temporally separated [4]. For fixed transmit beamformers $\mathbf{m}_{b,k,l} \; \forall \; (b,k,l)$, the optimal receive beamformer can be obtained from (5). However, (7) remains non-convex in terms of the transmit beamformers. We use the first-order approximation of the non-convex objective to transform (7) into tractable convex optimization problem. The objective is separable in terms of $\epsilon_{k,l} \; \forall \; (k,l)$. Thus, we can approximate each term individually around the point $\epsilon_{k,l}^{(n)}$ as

$$\log_2\left(\epsilon_{k,l}\right) \approx \frac{1}{\log(2)\epsilon_{k,l}^{(n)}}\left(\epsilon_{k,l} - \epsilon_{k,l}^{(n)}\right) + \log_2\left(\epsilon_{k,l}^{(n)}\right). \quad (8)$$

Ignoring the constant terms, the approximated transmit beamformer design subproblem can be restated as a WMMSE problem

$$\min_{\mathbf{m}_{b,k,l}} \quad \sum_{k=1}^{K}\sum_{l=1}^{L_k} w_{k,l}^{(n)} \epsilon_{k,l}$$
$$\text{s. t.} \quad \sum_{k \in \mathcal{C}_b}\sum_{l=1}^{L_k} \|\mathbf{m}_{b,k,l}\|^2 \leq P_b, \; b = 1, \dots, B, \quad (9)$$

where $w_{k,l}^{(n)} = \frac{\mu_k}{\log(2)\epsilon_{k,l}^{(n)}} \; \forall \; (k,l)$. Since the MSE terms in (3) are convex for fixed receivers, (9) is a convex problem and can be formulated as a second order cone program (SOCP) for efficient computation as in [4]. The centralized algorithm with perfect CSI is outlined in Alg. 1. As shown in [3], the successive approximation algorithm provides monotonic convergence of the objective function and convergence to a local stationary point of the original problem (4).

Algorithm 1 Centralized WMMSE algorithm.

1: Initialize feasible $\mathbf{m}_{b,k,l} \; \forall \; (b,k,l)$ and $n = 1$.
2: **repeat**
3:     Generate the MMSE receivers $\mathbf{u}_{k,l} \; \forall \; (k,l)$ from (5).
4:     Compute MSE $\epsilon_{k,l}^{(n)}$ and weights $w_{k,l}^{(n)} \; \forall \; (k,l)$ from (3).
5:     Solve the precoders $\mathbf{m}_{b,k,l} \; \forall \; (b,k,l)$ from (9).
6:     Set $n = n + 1$.
7: **until** Desired level of convergence has been reached.

We now relax the pilot orthogonality requirement and formulating the centralized beamformer design problem with pilot estimation noise and interference from non-orthogonal pilot sequences.

Downlink beamformer estimation

Let $\mathbf{b}_{k,l} \in \mathbb{C}^S$ denote the UL pilot training sequence for the lth data stream of UE $k = 1, \dots, K$, where S is the length of the pilot sequence. Then, the composite of the precoded UL pilot training matrices received at BS b is

$$\mathbf{R}_b = \sum_{k=1}^{K}\sum_{l=1}^{L_k} \mathbf{H}_{b,k}^{\mathsf{H}}\mathbf{u}_{k,l}\sqrt{w_{k,l}}\mathbf{b}_{k,l}^{\mathsf{H}} + \mathbf{N}_b, \quad (10)$$

where $\mathbf{N}_b \in \mathbb{C}^{N_{\mathrm{T}} \times S}$ is the estimation noise matrix for all pilot symbols. We employ precoded training pilots, where the weighted receive beamformers serve as pilot precoders.

Assuming only the uplink training signal (10) at the BSs, we can rewrite (9) as

$$\min_{\mathbf{m}_{b,k,l}} \quad \sum_{(k,l)} \left(1 - 2\mathrm{Re}\{\sum_{b \in \mathcal{B}_k} \sqrt{w_{k,l}}\mathbf{m}_{b,k,l}^{\mathsf{H}}\mathbf{R}_b\mathbf{b}_{k,l}\} + \right.$$
$$\left. \left(\sum_{b \in \mathcal{B}_k} \mathbf{m}_{b,k,l}^{\mathsf{H}}\mathbf{R}_b\right)\left(\sum_{b \in \mathcal{B}_k} \mathbf{R}_b^{\mathsf{H}}\mathbf{m}_{b,k,l}\right) \right) \quad (11)$$
$$\text{s. t.} \quad \sum_{k \in \mathcal{C}_b}\sum_{l=1}^{L_k} \|\mathbf{m}_{b,k,l}\|^2 \leq P_b, \; b = 1, \dots, B.$$

When we substitute (10) into (11), we have

$$\min_{\mathbf{m}_{b,k,l}} \quad \frac{1}{S}\sum_{(k,l)} \left(1 - 2\mathrm{Re}\{w_{k,l}\mathbf{u}_{k,l}^{\mathsf{H}}\sum_{b \in \mathcal{B}_k}\mathbf{H}_{b,k}\mathbf{m}_{b,k,l}\} + \psi_{k,l} \right.$$
$$\left. + \sum_{b \in \mathcal{B}_k} \mathbf{m}_{b,k,l}^{\mathsf{H}}\left(\sum_{i=1}^{K}\sum_{z=1}^{L_i}\mathbf{H}_{b,i}^{\mathsf{H}}\mathbf{u}_{i,z}w_{i,z}\mathbf{u}_{i,z}^{\mathsf{H}}\mathbf{H}_{b,i}\right)\mathbf{m}_{b,k,l} \right), \quad (12)$$

where $\psi_{k,l} \geq 0$ indicates the weighted pilot cross interference. It is easy to see that (12) corresponds to the WMMSE objective in (9) with imperfect pilot estimation. In fact, it is equivalent to (9), if the pilot sequences are orthogonal and $\mathbf{N}_b \to \mathbf{0}$. In the rest of the paper, we consider (11) as the main problem.

Problem (11) requires knowledge of the received training matrices $\mathbf{R}_b$, training sequences $\mathbf{b}_{k,l}$ and the weights $w_{k,l}$ only for the served users. All of this can be gathered with carefully designed TDD pilots and feedback for the weights [4]. The transmit beamformers can be written in a closed form expressions by evaluating the first-order optimality conditions as

$$\mathbf{m}_{b,k,l} = \left(\mathbf{R}_b\mathbf{R}_b^{\mathsf{H}} + \mathbf{I}\nu_b\right)^{-1}\mathbf{R}_b\left(\mathbf{b}_{k,l}\sqrt{w}_{k,l} - [\bar{\mathbf{c}}_{b,k,l}]^{\mathsf{H}}\right), \quad (13)$$

where

$$\bar{\mathbf{c}}_{b,k,l} = \sum_{j \in \mathcal{B}_k \setminus \{b\}} [\mathbf{R}_j]^{\mathsf{H}}\mathbf{m}_{j,k,l}. \quad (14)$$

From (13), we can see that the transmit beamformers can be solved by relying solely on the composite pilot training matrices $(\mathbf{R}_b)$ and local pilot sequences (pilot of the intended user). That is, we do not require any additional information on the interfering DL transmission. It is also easy to see how the beamformer generation is coupled within the cooperating clusters. In the following section, we will exploit this relation to derivate an efficient decentralized JP beamforming algorithm.

An alternative approach to the formulation in (11), is to extract the effective intended signal and interference channels from (10) locally in each BS b as

$$\mathbf{R}_b\mathbf{b}_{i,j} = \mathbf{H}_{b,i}^{\mathsf{H}}\mathbf{u}_{i,j}\sqrt{w_{i,j}} + \delta_{b,i,j} \; \forall \; (i,j), \quad (15)$$

where $\delta_{b,i,j}$ denotes the estimation and pilot contamination noise. These estimates can be used in the original formulation (9) [13]. We call this approach stream specific estimation (SSE). SSE works well, when $\delta_{b,i,j} \to 0$, i.e., the pilot sequences of the dominant interference sources are orthogonal

and pilot noise levels are manageable. On the other hand, proper interference management requires that the pilot sequenses are known for all users not only the served users. We compare the performance of both approaches by numerical examples in Section V.

Uplink beamformer estimation

In analogy with the UL, let the received composite DL pilot training matrix at UE $k = 1, \ldots, K$ be given as

$$\mathbf{T}_k = \sum_{i=1}^{K} \sum_{l=1}^{L_i} \left(\sum_{b \in \mathcal{B}_i} \mathbf{H}_{b,k} \mathbf{m}_{b,i,l} \right) \mathbf{b}_{i,l} + \mathbf{N}_k. \quad (16)$$

The rate optimal receive beamformers are the MSE minimizing receivers, given by

$$\mathbf{u}_{k,l} = \left(\mathbf{T}_k \mathbf{T}_k^{\mathrm{H}} + \mathbf{I} N_0 \right)^{-1} \mathbf{T}_k \mathbf{b}_{k,l}^{\mathrm{H}}. \quad (17)$$

Here, we assume that the UL and DL pilots are the same. This does not have to be case, and the UL/DL pilots can be separately designed. In the sequel, we consider decentralized beamforming techniques for solving (11). MMSE receive beamformer estimation is readily decentralized and, thus, we will focus on DL transmit beamformer estimation.

IV. Decentralized Beamformer Design

In this section, we consider decentralized JP beamformer design, when SSE is not applicable and the effective channels (15) cannot be accurately estimated. The beamformer signaling relies crucially on the channel reciprocity of TDD. In [3] and [4], it was shown that CB using the WMMSE algorithm has inherently decoupled interference processing. As such, it can be easily decentralized with low signaling overhead. However, the JP transmit beamformer design in (11) is coupled among the BSs due to the coherent signal reception, which prevents us from directly applying the same decentralized processing method.

The BR design employs the parallel optimization scheme proposed in [5] to decentralize the beamformer design. This parallel framework is based on solving for the beamformers locally in each BS, while assuming that the cooperating BSs keep their transmitters fixed. Since each BS relies only on the knowledge of the coupled transmissions from the previous iteration, the beamforming problem becomes decoupled. It was shown in [5] that, if the local problems are strongly convex, the beamformer updates can be made monotonic with respect to the original WSRMax objective function. Note that the strong convexity of (9) follows directly from the strong convexity of the individual MSE functions (3). However, due to the pilot estimation noise and non-orthogonal pilots, monotonic convergence cannot always be guaranteed. In Section V, we show by numerical examples that BR updates do provide, on average, monotonic performance improvement with static channels.

We start by considering the transmit beamformer design for BS b, while assuming that the transmissions from the other BSs are fixed. The transmit beamformers can be solved relying only on

$$\mathbf{c}_{j,k,l}^{(n)} = [\mathbf{R}_j^{(n)}]^{\mathrm{H}} \mathbf{m}_{j,k,l}^{(n)} \in \mathbb{C}^S \quad (18)$$

and locally estimated (10). Here (n) denotes the iteration index. Again in (13), we can group the fixed terms together as $\bar{\mathbf{c}}_{b,k,l}^{(n)} = \sum_{j \in \mathcal{B}_k \setminus \{b\}} \mathbf{c}_{j,k,l}^{(n)}$. Since we fix the coupled variables in (13), the optimal transmit beamformers can be determined from

$$\mathbf{m}_{b,k,l} = \left(\mathbf{R}_b \mathbf{R}_b^{\mathrm{H}} + \mathbf{I} \nu_b \right)^{-1} \mathbf{R}_b \left(\mathbf{b}_{k,l} \sqrt{w}_{k,l} - [\bar{\mathbf{c}}_{b,k,l}^{(n)}]^{\mathrm{H}} \right) \quad (19)$$

by bisection search over ν_b to satisfy the transmit power constraints $\sum_{k \in \mathcal{C}_b} \sum_{l=1}^{L_k} \|\mathbf{m}_{b,k,l}\|^2 \leq P_b$. Note that, if $\sum_{k \in \mathcal{C}_b} \sum_{l=1}^{L_k} \|\mathbf{m}_{b,k,l}\|^2 < P_b$ for $\nu_b = 0$, then this is the optimal solution. Note that (19) involves only one matrix inversion. The bisection search algorithms are known to converge quickly to any practically meaningful accuracy [14]. Furthermore, the dimensions of $(\mathbf{R}_b \mathbf{R}_b^{\mathrm{H}} + \mathbf{I} \nu_b)$ depend only on the number of antennas in BS b (N_{T}) and not on the dimensions of the joint beamformer ($|\mathcal{B}_k| N_{\mathrm{T}}, k = 1, \ldots, K$). This makes the per iteration computational complexity, at each BS, comparable to CB [4]. The iterated BR algorithm is summarized in Alg. 2.

After each iteration n, the fixed terms are signaled within the JP clusters and beamformers are updated as

$$\mathbf{m}_{b,k,l}^{(n+1)} = \mathbf{m}_{b,k,l}^{(n)} + \alpha \left(\mathbf{m}_{b,k,l}^{*} - \mathbf{m}_{b,k,l}^{(n)} \right) \ \forall \ (b,k,l), \quad (20)$$

where α is a sufficiently small step-size and $\mathbf{m}_{b,k,l}^{*}$ is the optimal solution for (19). However, the convergence cannot be guaranteed because of the pilot estimation noise, which introduces a random component into each iteration of the beamforming problem. For further details on the convergence properties and step-size selection see [5].[1]

The signaling requirements are apparent from (19). Each BS b requires the knowledge of $\mathbf{c}_{j,k,l}^{(n)}$ from the cooperating BSs $j \in \mathcal{B}_k$ for each stream (k,l). This cumulates into $|\mathcal{C}_b| \sum_{k \in \mathcal{C}_b} L_k S$ complex terms. Note that vector $\mathbf{c}_{j,k,l}^{(n)}$ has length S and, thus, there is a tradeoff between signaling overhead and performance.

V. Numerical Examples

The simulations are carried out using a 7-cell wrap around model, where the distance between the BSs is 600m. The path loss exponent for the user terminals is fixed to 3. The number of transmit and receive antennas are set to $N_{\mathrm{T}} = 4$ and $N_{\mathrm{R}} = 2$, respectively. There are $K_b = 3$ user terminals that are evenly distributed on the cell edge around each BS. In total, there are $K = BK_b = 21$ users in the network. We assume full cooperation, i.e., all users are coherently served by every BS in the system. In practice, practical constraints such as pilot contamination will limit the number of active users per-BS. The number of active spatial streams per users is limited

---

[1] For constant channels and no estimation noise, the largest α that guarantees convergence can be analytically bounded with respect to the Lipschitz constant of the objective [5].

Algorithm 2 Decentralized BR algorithm for WSRMax.

1: Initialize feasible $\mathbf{m}_{b,k,l}$ $\forall$ (b,k,l) and $n=1$.
2: **repeat**
3:   *UE*: Generate the MMSE receivers $\mathbf{u}_{k,l}$ $\forall$ (k,l) from (17).
4:   *UE*: Compute the MSE $\epsilon_{k,l}^{(n)}$ $\forall$ (k,l) from (3).
5:   *UE*: Set the weights $w_{k,l}^{(n)}$ $\forall$ (k,l) from (III).
6:   *BS*: Solve the precoders $\mathbf{m}_{b,k,l}$ $\forall$ (b,k,l) from (19).
7:   *BS*: Exchange the fixed terms (18) among the cooperating BSs.
8:   *BS*: Update the next iteration precoders from (20).
9:   Set $n=n+1$.
10: **until** Desired level of convergence has been reached.

to one. The simulation environment is illustrated in Fig. 2. The signal-to-noise ratio (SNR) is defined on the cell edge
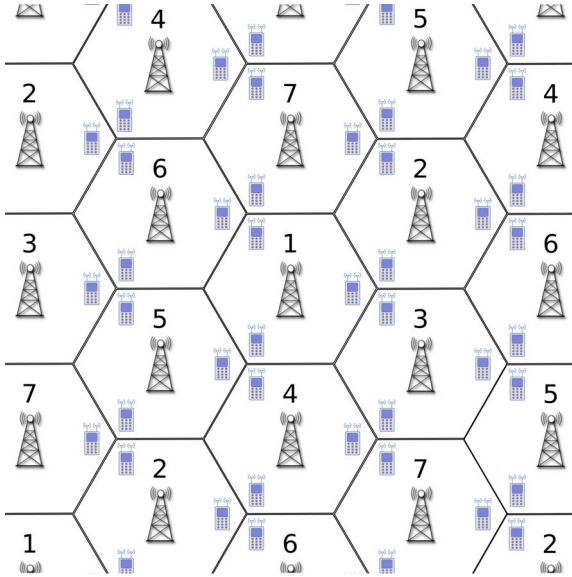

Fig. 2. An illustrative figure of the BS and cell edge UE deployment in 7-cell wrap around model with $K_b = 3$ in each cell.

from the closest BS b, i.e, SNR $= \frac{g_{b,k}P_b}{\sigma_k^2}$, where $g_{b,k}$ denotes the corresponding path loss. The SNR is fixed to 15dB. The pilot training sequences are random binary (± 1) sequences with 10dB pilot power gain over the SNR. This reflects a worst case scenario, where none of the pilot resources are made orthogonal. In systems with more limited cooperation, the practical approach would be to make the pilot sequences orthogonal within the JP clusters. On the other hand, this requires that the clusters are not heavily overlapping and have tight coordination over the pilot resources. If not otherwise stated, the pilot training sequence length is 41. The parameters for different methods are not optimized for specific training sequence lengths or fading conditions. Rather, the parameters are chosen such that the best overall performance is achieved. Parameter α for the BR design is fixed to 0.25.

The channels are generated with Jakes' Doppler spectrum

TABLE I
SIMULATION PARAMETERS.

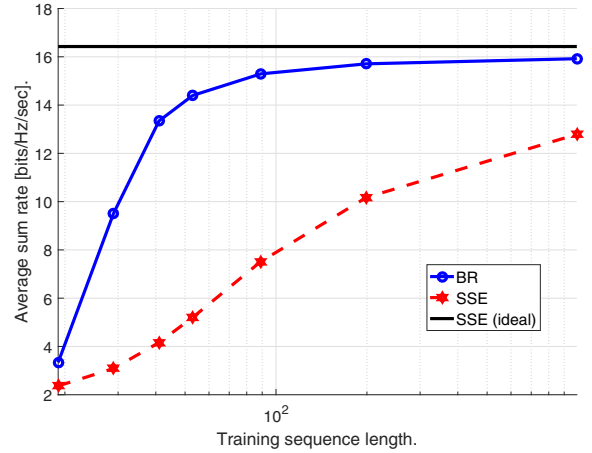| Parameter	Value
Number of UEs (K) / cells (B)	21 (3 per cell) / 7
BS antennas (N_{T}) / UE antennas (N_{R})	4/2
SNR / Pilot power	15dB/25dB
Distance between adjacent BSs	600m
The path loss exponent	3

Fig. 3. Behavior for varying training sequence lengths with constant channels.

model. The channel coherence time is defined by normalized user terminal velocity $t_{\mathrm{s}}f_{\mathrm{d}}$, where t_{s} and f_{d} are the backhaul signaling rate and the maximum Doppler shift, respectively. For example, the 5G systems are expected to have frame size close to 0.25ms and information is exchanged at least once per frame [15]. With 5 GHz carrier frequency, normalized velocity $t_{\mathrm{s}}f_{\mathrm{d}} = 0.005$ equals to 10km/h. The block fading model assumes that the channels remain constant during the transmission of each frame, and the changes occur in-between the frames. The SSE reference scheme uses the SSE method from (15), where all interference terms are individually estimated. As a general performance upper bound, we use SSE with 21 fully orthogonal pilots and without pilot estimation noise. This is denoted by SSE (ideal). Summary of the simulation parameters is listed in Table I.

Fig. 3 demonstrates the gain provided by taking the pilot interference into account as the length of the pilot training sequence is varied. Here, SSE denotes a stream specific beamformer design, where the pilot non-orthogonality and pilot estimation errors are completely ignored. It is easy to confirm that proposed designs have a clear advantage, when the pilot interference levels are high. On the other hand, it should be noted that, with sufficiently long pilot sequences, the pilots can be made fully orthogonal, which reduces the performance gap with large pilot lengths. Even in that case, the pilot estimation noise should be included in the estimate.

The convergence behavior in constant and time correlated

VI. Conclusions

We have proposed decentralized transceiver design for JP CoMP WSRMax in the presence of non-orthogonal pilot resources and pilot estimation noise. Emphasis was given to design that enables the use of JP CoMP in realistic channel fading conditions. Decentralized JP was shown to be feasible even with limited pilot resources, i.e., with limited CSI accuracy, by using carefully designed backhaul and pilot signaling. The proposed beamformer estimation approach was shown to have significant gain versus purely channel based estimation in noisy and dense systems. The BR algorithm was used to provide JP with attractive convergence and performance properties. The numerical results indicated that our algorithm provided good performance and stability even in moderately fast fading channel conditions.

References

[1] E. Dahlman, S. Parkvall, and J. Sköld, *4G LTE / LTE-Advanced for Mobile Broadband*. Academic Press, 2011.

[2] C. L. I, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, "Toward green and soft: a 5G perspective," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 66–73, Feb. 2014.

[3] 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. Signal Processing*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[4] P. Komulainen, A. Tölli, and M. Juntti, "Effective CSI Signaling and Decentralized Beam Coordination in TDD Multi-Cell MIMO Systems," *IEEE Trans. Signal Processing*, vol. 61, no. 9, pp. 2204–2218, 2013.

[5] G. Scutari, F. Facchinei, P. Song, D. Palomar, and J.-S. Pang, "Decomposition by partial linearization: Parallel optimization of multi-agent systems," *IEEE Trans. Signal Processing*, vol. 62, no. 3, pp. 641–656, Feb. 2014.

[6] E. Björnson, R. Zakhour, D. Gesbert, and B. Ottersten, "Cooperative multicell precoding: Rate region characterization and distributed strategies with instantaneous and statistical CSI," *IEEE Trans. Signal Processing*, vol. 58, no. 8, pp. 4298–4310, Aug. 2010.

[7] 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. Select. Areas Commun.*, vol. 31, no. 2, pp. 226–240, Feb. 2013.

[8] S. H. Park, O. Simeone, O. Sahin, and S. Shamai, "Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks," *IEEE Trans. Signal Processing*, vol. 61, no. 22, pp. 5646–5658, Nov. 2013.

[9] W.-C. Liao, M. Hong, Y.-F. Liu, and Z.-Q. Luo, "Base station activation and linear transceiver design for optimal resource management in heterogeneous networks," *IEEE Trans. Signal Processing*, vol. 62, no. 15, pp. 3939–3952, Aug. 2014.

[10] D. Jaramillo-Ramirez, M. Kountouris, and E. Hardouin, "Coordinated multi-point transmission with imperfect CSI and other-cell interference," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 1882–1896, Apr. 2015.

[11] C. Shi, R. Berry, and M. Honig, "Bi-directional training for adaptive beamforming and power control in interference networks," *IEEE Trans. Signal Processing*, vol. 62, no. 3, pp. 607–618, Feb. 2014.

[12] Z. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Select. Areas Commun.*, vol. 2, no. 1, pp. 57–73, Feb. 2008.

[13] J. Kaleva, R. Berry, M. Honig, A. Tölli, and M. Juntti, "Decentralized sum MSE minimization for coordinated multi-point transmission," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, May 2014, pp. 469–473.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[15] E. Lähetkangas, K. Pajukoski, J. Vihriälä, G. Berardinelli, M. Lauridsen, E. Tiirola, and P. Mogensen, "Achieving low latency and energy consumption by 5G TDD mode optimization," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2014, pp. 1–6.
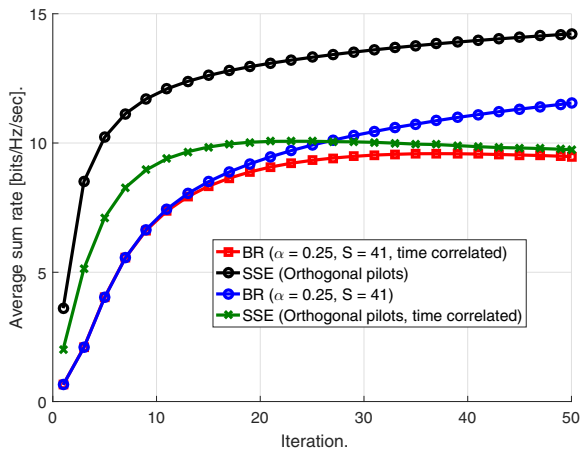
Fig. 4. BR performance for varying training sequence lengths in constant and time correlated (normalized UE velocity $t_s f_d = 0.005$) channels.
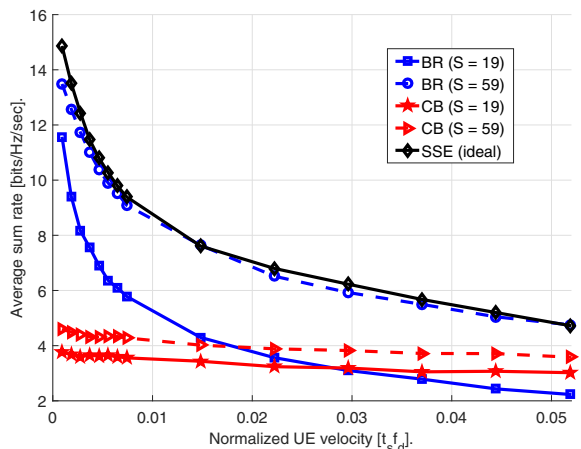

Fig. 5. Saturated performance comparison with normalized UE velocities.

fading ($t_s f_d = 0.005$) channels is shown in Fig. 4. We have focused only on the first 50 beamformer iterations. This is not enough to let the constant channel curves fully converge. However, in practice, 50 iterations should cover the feasible number of over-the-air beamformer training iterations. In time correlated fading channels, the trend is that the performance saturates after 20 iterations. This is due to averaging over multiple channel realizations. For any single channel realization, the performance varies in time as the channel changes.

Robustness to UE mobility can be seen from Fig. 5. Here, the proposed JP BR algorithm is compared to coordinated beamforming. It can be seen that the BR design is robust to UE mobility. It is evident that, as the mobility grows, improved beamformer training is highly beneficial. Even with limited training sequence length ($S = 19$), the BR algorithm achieves JP gain when UE velocity $t_s f_d$ is less than 0.015. For higher mobility, the sequence length needs to increased. Then the performance is dominated by the channel fading and the ideal SSE and BR perform equivalently well already with $S = 59$.