Decentralized Joint Precoding with Pilot Aided Beamformer Estimation

Jarkko Kaleva, *Student Member, IEEE,* Antti Tölli, *Senior Member, IEEE,* Markku Juntti, *Senior Member, IEEE,* Randall Berry, *Fellow, IEEE* and Michael Honig *Fellow, IEEE*

***Abstract*—Downlink beamforming techniques with low beamformer training overhead are proposed for joint processing (JP) coordinated multi-point transmission (CoMP). The objective is to maximize the weighted sum rate within joint transmission clusters without centralized beamformer processing, while accounting for uncertainty in the underlying channels. The proposed methods use time division duplexing and pilot based training with, possibly, non-orthogonal pilot sequences. The beamformer training is done without explicit channel state information estimation, which greatly improves the robustness to pilot contamination. Best response and gradient-based decentralized algorithms are proposed and provide a trade-off between computational complexity and fast convergence rate. The impact of feedback/backhaul quantization is also considered. The results show that JP CoMP is feasible with slow fading conditions and with limited backhaul capacity by employing decentralized beamformer processing.**

***Index Terms*—coordinated multi-point transmission, decentralized beamforming, joint processing, pilot training, weighted sum rate maximization.**

I. Introduction

Cooperative transmission schemes and spatial domain interference management are the foundation of modern cellular and heterogeneous wireless systems. Current wireless standards already support efficient single cell multiple-input multiple-output (MIMO) beamforming, which allows smart beamformer design to exploit multi-user diversity in the spatial domain [3]. Also, the basic operation of multi-cell MIMO transceiver processing has been incorporated in the Long Term Evolution Advanced (LTE-A) standards. Multi-cell coordinated beamforming (CB) is still in the initial stages with respect to the current LTE-A standard [3]. On the other hand, research effort has been invested in CB for multi-cell systems. Much of this research has focused on decentralized coordination strategies [4]–[7]. However, the available degrees of freedom (DoF) of CB are limited due to the capabilities for inter-cell interference suppression at each base station (BS). Joint processing (JP) coordinated multi-point transmission (CoMP) allows joint transmission from cooperating BSs, which greatly increases the available DoF and alleviates the detrimental impact of interference [8]–[11]. Still, practical limitations in BS fronthaul/backhaul connectivity are hindering effective implementation of advanced JP CoMP schemes.

In a cloud radio access network (C-RAN), joint beamforming is fully centralized and the BSs act merely as virtual remote radio heads (RRHs) [12]. As such, the BSs are connected to a remote central processor (RCP) in the cloud over high capacity and low latency links. The RCP can then use JP to simultaneously utilize multiple RRHs for beamforming. Although the fronthaul/backhaul limitations of CoMP transmission systems have been addressed in various publications, most of the JP CoMP research assumes full channel state information (CSI) exchange and centralized processing, which greatly simplifies the beamformer design [6], [10], [13]–[15]. While perfect CSI across multiple RRHs is a common assumption for JP designs, it may not be feasible in practice. Latency and mobility requirements often prevent accurate CSI exchange even in a modest scale [11].

In this paper, we assume an architecture where the beamformer computation is brought close to the radio. The BSs are connected via a limited capacity backhaul. The backhaul supports the centralized data sharing and control signaling between the BSs, but the delay constraints of the changing channel conditions prevent centralized CSI sharing and beamformer training within the channel coherence window. Thus, in order to utilize JP, the BSs need to perform partially independent beamformer design within the JP clusters. Still, we assume that the transmitted data can be shared among the serving BSs. The data is assumed to be queued at the RCP and the upper layer priority weights are distributed by the RCP to the serving BSs as beamformer weighting. Furthermore, we assume non-orthogonal and noisy pilots, which is expected in dense deployments. The beamformer estimation is done by direct estimation (DE), where the beamformers are directly estimated from the contaminated pilot training sequences. This is an alternative to stream specific estimation (SSE), where all interference sources are estimated individually and the transmit covariance matrices are constructed from the individual estimates.

A. Contributions

JP CoMP methods with data sharing are presented for the weighted sum rate maximization (WSRMax). The transmit beamformers are locally designed in each BS, which reduces the overall backhaul load. Our focus is on practically realizable pilot aided beamformer designs with low signaling overhead

Jarkko Kaleva, Antti Tölli and Markku Juntti are with Centre for Wireless Communications, University of Oulu P.O. Box 4500, 90014 University of Oulu, Finland. Email: jarkko.kaleva@oulu.fi; antti.tolli@oulu.fi; markku.juntti@oulu.fi. Randall Berry and Michael Honig are with Northwestern University, Dept. of Elect. Engineering and Comp. Science, Northwestern University, Evanston, IL, 60208, U.S. Email: rberry@eecs.northwestern.edu; mh@eecs.northwestern.edu.

Parts of this work has been published in IEEE International Conference on Acoustics, Speech, and Signal Processing, Florence, Italy, 2014 [1] and IEEE Global Communications Conference, San Diego, U.S, 2015 [2].

that are robust in the presence of channel fading. We propose over-the-air (OTA) beamformer training by DE, where full (effective) CSI estimation is not required at the BSs. This reduces the number of estimated parameters as opposed to estimating the CSI for each channel directly. We employ best response (BR) and gradient descent (GD) based schemes to provide decentralized algorithms with different performance properties, reduced computational complexity and signaling overhead. The signaling depends on the number of strong interferers rather than the number of antennas, which greatly reduces the overhead. We also consider imperfect pilot estimation and non-orthogonal pilot resources. The performance of the proposed method is evaluated in a cellular multi-user network with a time-correlated channel model.

Contributions of this paper are summarized as follows:

- Channel uncertainty and decentralized precoder design for JP CoMP are jointly considered, while taking into account the needed overhead.
- DE is shown to provide significant gain over SSE on limited pilot training sequence lengths.
- The signaling overhead of the proposed methods depends on the number of strong interferers rather than the number of antennas.
- Decentralized JP is demonstrated to be feasible with limited backhaul signaling, noisy non-orthogonal pilots and high user equipment (UE) mobility.

B. *Prior Work*

Unlike JP CoMP, CB has been extensively studied for decentralized inter-cell interference coordination. In CB, the interfering BSs cooperatively coordinate the beamformers in such a way that the interference conditions are not detrimental for the neighboring cells. CB can be efficiently performed with limited signaling overhead as shown, e.g., in [4]–[6]. WSRMax with CB has been studied, e.g., in [16]–[18]. The backhaul delay and capacity limitations have motivated proposals that go beyond the centralized processing architectures [6], [19]. A popular CB approach is the weighted minimum mean-squared error (WMMSE) method, where the rate objective is replaced with an equivalent weighted mean-squared error (MSE) minimization objective [4], [18]. The WMMSE has been shown to have a convenient structure for decentralized processing in cellular time division duplex (TDD) MIMO systems [4], [5]. Similar approaches to WSRMax have also been considered in [7], [20], [21]. More generalized parallel design frameworks have also been proposed. For example, a general BR framework, which allows straightforward parallel processing with different performance objectives, was proposed in [20]. JP inherently couples the beamformer processing among cooperating BSs. The authors of [22] propose heuristic JP CoMP schemes that make use of only local CSI. These schemes cannot achieve all available DoF due to the lack of the global CSI. Unlike [22], we present decentralized beamformer designs that asymptotically approach the performance of fully centralized processing as the number of transceiver iterations tend to infinity.

In centralized JP, the fronthaul/backhaul information can be handled in one of two ways: (*i*) In *data sharing*, the RCP exchanges the specific messages with the cooperating BSs explicitly and the joint beamformers are sent over the backhaul to the BSs separately from the data [23]. (*ii*) Using *compression*, the messages are precoded beforehand at the RCP and only the compressed versions of the analog beamformers are sent over the fronthaul to the RRHs [24], [25], thus avoiding the separate data exchange. Sparse JP designs are among the most common approaches for fronthaul/backhaul limited CoMP [23], [26]–[28]. These designs try to limit the JP cluster sizes and, thus, implicitly reduce the data sharing overhead. The compression approach has recently gained more attention [24], [25], [29], [30], where different aspects and benefits of this approach are studied. The data sharing and compression strategies for energy efficient communication and backhaul power consumption are compared in [25]. Our focus is on decentralized JP beamformer designs with data sharing that have inherently low signaling overhead. In addition, we could also readily apply the sparsity imposing techniques without major technical changes. However, to simplify the presentation, we limit ourselves to decentralized processing. We have chosen the data sharing approach as the compression based techniques require centralized knowledge of beamformers, which would invalidate our aim for decentralized beamformer processing.

Pilot non-orthogonality and contamination have been widely studied, albeit, not so much in the context of JP CoMP. The impact of partial or imperfect CSI feedback for CB and JP has been studied, e.g., [10], [31], [32]. Pilot contamination in TDD based transceiver training for CB has been considered, e.g., in [33]–[35]. In [34], [35], direct least squares (LS) beamformer estimation from the contaminated uplink (UL)/downlink (DL) pilots was shown to provide good performance as opposed to the estimation of the individual channels separately. Here, we show that similar conclusions hold for JP CoMP.

Various alternative performance objectives to the WSRMax have been proposed, e.g., sum MSE minimization in [1], [34] and traffic-aware sum queue minimization in [36]. These objectives are also directly applicable to the proposed JP CoMP framework. They impose more fairness among the users than WSRMax, which tends to prefer the strong or prioritized users. For guaranteed minimal UE rates, the decentralized rate balancing methods from [21] can also be straightforwardly applied to the proposed designs.

C. *Organization and Notation*

The rest of the paper is organized as follows. The system model is given in Section II. In Section III, the WSRMax problem is described along with the WMMSE successive convex approximation (SCA) approach. Decentralized beamformer designs are considered in Section IV. Finally, the numerical examples and concluding remarks are given in Sections V and VI, respectively.

Notation: Matrices and vectors are presented by boldface upper and lower case letters, respectively. The transpose of matrix $\mathbf{A}$ is denoted as $\mathbf{A}^{\mathrm{T}}$ and, similarly, the conjugate transpose is denoted as $\mathbf{A}^{\mathrm{H}}$. Conventional matrix inversion is written as $\mathbf{A}^{-1}$. Cardinality of a discrete set $\mathcal{A}$ is given by

$|A|$. The expected value of a random variable is denoted by $\mathrm{E}[\cdot]$.

II. System Model

We consider a multi-cell system with B BSs each equipped with N_{T} transmit antennas. There are, in total, K 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 joint processing cluster (set of phase-coherent 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 $k \in \mathcal{K}$ is denoted by $L_k \leq \min(|\mathcal{B}_k| N_{\mathrm{T}}, N_{\mathrm{R}})$. To simplify the notation in various places, we use the following set abbreviations: $\forall(k,l) \triangleq \{(k,l) | k \in \mathcal{K}, l = 1, \ldots, L_k\}$ and $\forall(b,k,l) \triangleq \{(b,k,l) | k \in \mathcal{K}, l = 1, \ldots, L_k, b \in \mathcal{B}_k\}$. The system model is illustrated in Fig. 1.

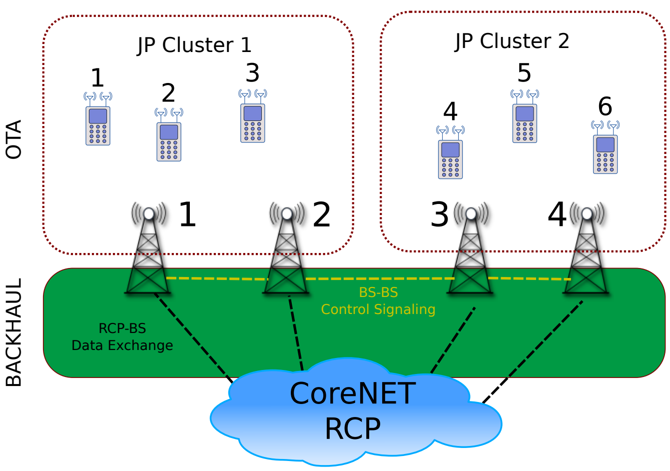

Fig. 1. Simplified system model with $B = 4$ cooperating BSs and $K = 6$ UEs that are split into two JP clusters. Here, for instance, $\mathcal{B}_2 = \{1, 2\}$ and $\mathcal{C}_4 = \{4, 5, 6\}$.

Without loss of generality, the downlink 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 served by $\mathcal{B}_k, k = 1, \ldots, K$. Only the local CSI knowledge is assumed, that is, each BS $b = 1, \ldots, B$ is only aware of the channel matrix $\mathbf{H}_{b,k} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}} \forall k = 1, \ldots, K$, while 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. Both, the UL and DL chains are assumed to be be fully calibrated in the sense that UL/DL channels are fully reciprocal. The channel fading process can have time-correlated behaviour and it is not assumed to be static. In Section V, the channel properties used to verify the algorithms are described in further detail.

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

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

where $\mathbf{m}_{b,i,j} \in \mathbb{C}^{N_{\mathrm{T}}}$ is the beamformer vector for the j^{th} 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 $\mathrm{E}\{|d_{k,l}|^2\} = 1$.

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

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

and the corresponding MSE is

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

Note that (3) is a convex function in terms of either the transmit or 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

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

where $\alpha_k, k = 1, \ldots, K$ are the user priority weights. The problem is non-convex and known to be NP-hard [37]. The optimal, i.e., rate maximizing, receive beamformers for (4) are the linear minimum mean-squared error (MMSE) receivers

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

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

It is well-known that, when the MMSE receive beamformers are applied, there is an inverse relation between the SINR and the corresponding MSE [4] given by

$$\varepsilon_{k,l}^{-1} = 1 + \gamma_{k,l}. \tag{6}$$

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

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

Since (7) is not jointly convex for the transmit and receive beamformers, we alternate between solving for the transmit and receive beamformers. This separation is convenient for TDD processing as the DL and UL transmissions are temporally separated. With fixed transmit beamformers $\mathbf{m}_{b,k,l}\ \forall\ (b,k,l)$, the optimal receive beamformer can be obtained from (5) [18]. This can be easily verified from the first-order Karush-Kuhn-Tucker (KKT) conditions. As (7) is still non-convex, even for fixed receive beamformers, we apply an iterative convex approximation algorithm based on the WMMSE criterion to determine the transmit beamformers [4], [18].

We use the WMMSE criterion to approximate the non-convex objective in (7) (minimization of the sum of concave functions). 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 follows:

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

Equation (8) provides the first-order linear approximation of the logarithmic function. The estimation error (with similar update distances) is higher at low MSE values, i.e., high signal-to-noise ratio (SNR). This approximation is tight at the point of approximation. In the centralized case, the approximation becomes tight after a few iterations [4].

Using (8) and ignoring the constant terms, the approximated transmit beamformer design subproblem can be restated as a WMMSE problem

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

where

$$w_{k,l}^{(n)} = \frac{\mu_k}{\log(2)\,\epsilon_{k,l}^{(n)}}\ \ \forall\ (k,l). \quad (10)$$

Since the MSE terms (3) are convex for fixed receive beamformers, (9) is a convex problem and can be formulated as a second order cone program (SOCP) for efficient computation as in [5]. The centralized algorithm with perfect CSI is outlined in Algorithm 1. As shown in [4], the successive approximation algorithm provides monotonic convergence of the objective function and convergence to a local stationary point of the original problem (4). Alternatively, we could also apply extended WSRMax techniques such as the ones proposed in [21], where an SCA approach with improved rate of convergence is proposed. For more general information on similar SCA formulations, see [38], [39].

We now formulate the centralized beamformer design problem with pilot estimation noise and interference from non-orthogonal pilot sequences. After which we propose two decentralized beamformer processing schemes which trade off computational complexity with faster convergence.

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 the MSE $\epsilon_{k,l}^{(n)}\ \forall\ (k,l)$ from (3).
5: Set the weights $w_{k,l}^{(n)}\ \forall\ (k,l)$ from (10).
6: Solve the precoders $\mathbf{m}_{b,k,l}\ \forall\ (b,k,l)$ from (9).
7: Set $n = n + 1$.
8: **until** Desired level of convergence has been reached.

Downlink beamformer estimation

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

$$\mathbf{R}_b = \sum_{k=1}^{K}\sum_{l=1}^{L_k} \mathbf{H}_{b,k}^{\mathrm{H}}\mathbf{u}_{k,l}\sqrt{w_{k,l}}\,\mathbf{b}_{k,l}^{\mathrm{H}} + \mathbf{N}_b, \quad (11)$$

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 UL training signal (11) at the BSs, we can rewrite the WSRMax problem (9) as

$$\begin{aligned} \min_{\mathbf{m}_{b,k,l}} \quad & \sum_{(k,l)} \Bigg(1 - 2\mathrm{Re}\Big\{\sum_{b\in\mathcal{B}_k}\sqrt{w_{k,l}}\,\mathbf{m}_{b,k,l}^{\mathrm{H}}\mathbf{R}_b\mathbf{b}_{k,l}\Big\} + \\ & \Big(\sum_{b\in\mathcal{B}_k}\mathbf{m}_{b,k,l}^{\mathrm{H}}\mathbf{R}_b\Big)\Big(\sum_{b\in\mathcal{B}_k}\mathbf{R}_b^{\mathrm{H}}\mathbf{m}_{b,k,l}\Big)\Bigg) \\ \text{s.t.} \quad & \sum_{k\in\mathcal{C}_b}\sum_{l=1}^{L_k}\|\mathbf{m}_{b,k,l}\|^2 \leq P_b, \ \ b = 1,\ldots,B. \end{aligned} \quad (12)$$

When we substitute (11) into (12), we have

$$\begin{aligned} \min_{\mathbf{m}_{b,k,l}} \quad & \frac{1}{S}\sum_{(k,l)} \Bigg(1 - 2\mathrm{Re}\Big\{w_{k,l}\mathbf{u}_{k,l}^{\mathrm{H}}\sum_{b\in\mathcal{B}_k}\mathbf{H}_{b,k}\mathbf{m}_{b,k,l}\Big\} + \\ & \nu_{k,l} + \sum_{b\in\mathcal{B}_k}\mathbf{m}_{b,k,l}^{\mathrm{H}}\Big(\sum_{i=1}^{K}\sum_{z=1}^{L_i}\mathbf{H}_{b,i}^{\mathrm{H}}\mathbf{u}_{i,z}w_{i,z}\mathbf{u}_{i,z}^{\mathrm{H}}\mathbf{H}_{b,i}\Big)\mathbf{m}_{b,k,l}\Bigg), \end{aligned} \quad (13)$$

where $\nu_{k,l} \geq 0$ indicates the weighted pilot cross interference. It is easy to see that (13) 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 (12) as the main problem.

The transmit beamformers can be written in a closed form, by evaluating the first-order optimality conditions of (12), as

$$\mathbf{m}_{b,k,l} = \left(\mathbf{R}_b\mathbf{R}_b^{\mathrm{H}} + \mathbf{I}\lambda_b\right)^{-1}\mathbf{R}_b\left(\mathbf{b}_{k,l}\sqrt{w_{k,l}} - \mathbf{c}_{b,k,l}^{\mathrm{H}}\right), \quad (14)$$

where

$$\mathbf{c}_{b,k,l} = \sum_{j\in\mathcal{B}_k\setminus\{b\}} \mathbf{R}_j^{\mathrm{H}}\mathbf{m}_{j,k,l}. \quad (15)$$

From (13) and (14), it is easy to see how the beamformer generation is coupled within the cooperating clusters. Problem (12) requires knowledge of the received training matrices

$\mathbf{R}_b$, training sequences $\mathbf{b}_{k,l}$ and the weights $w_{k,l}$. All of this can be gathered with carefully designed TDD pilots and feedback for the weights [5].

Since the interference channels are not explicitly estimated, the beamformer training becomes more robust to pilot contamination, i.e, the cross pilot interference and estimation noise are explicitly considered in the optimization problem. Furthermore, by having multiple individual estimation steps (for each interfering signal), the estimation noise accumulates with the number of interfering streams. This is numerically studied in Section V. In the following section, we will exploit this relation to derive an efficient decentralized JP beamforming algorithm.

An alternative approach to formulation (12), is to separately estimate the *effective channels* (channels with embedded beamformers) from (11) locally in each BS b as

$$\mathbf{R}_b \mathbf{b}_{i,j} = \mathbf{H}_{b,i}^{\mathrm{H}} \mathbf{u}_{i,j} \sqrt{w_{i,j}} + \boldsymbol{\epsilon}_{b,i,j} \ \forall (i,j), \tag{16}$$

where $\boldsymbol{\epsilon}_{b,i,j}$ denotes the estimation and pilot contamination noise. Here, $\boldsymbol{\epsilon}_{b,i,j} \to 0$ when the pilots are made sufficiently long and orthogonal. These estimates can be used to formulate the transmit beamforming problem in (9) [40]. We call this approach SSE and refer to the prior approach as DE. SSE works well, when the pilot sequences of the dominant interference sources are orthogonal and pilot noise levels are manageable. 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

$$\mathbf{T}_k = \sum_{i=1}^{K} \sum_{l=1}^{L_i} \left(\sum_{b \in \mathcal{B}_i} \mathbf{H}_{b,k} \mathbf{m}_{b,i,l} \right) \mathbf{b}_{i,l} + \mathbf{N}_k. \tag{17}$$

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

$$\mathbf{u}_{k,l} = \left(\mathbf{T}_k \mathbf{T}_k^{\mathrm{H}} + \mathbf{I} N_0 \right)^{-1} \mathbf{T}_k \mathbf{b}_{k,l}^{\mathrm{H}}. \tag{18}$$

Note that 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 (12). MMSE receive beamformer estimation is readily decentralized and, thus, we will focus on DL transmit beamformer estimation.

The outline of the pilot aided beamformer training with centralized knowledge of the received pilot training sequences, at the BSs, is given in Algorithm 2. In Section IV, decentralized algorithms are proposed for the training. Since we assume TDD based signaling, the DL and UL beamformer training must alternate in time.

Bi-directional signaling schemes for TDD can be used with the pilot aided beamformer design [34]. This allows direct exchange of the effective UL and DL channels from the corresponding precoded UL/DL pilot signals. The signaling sequence occupies a fraction α of the DL frame. The remaining portion $(1-\alpha)$ of the frame is reserved for the transmitted data. The frame structure is illustrated in Fig. 2, where D and U denote DL and UL pilots, respectively.

Algorithm 2 Centralized WMMSE with pilot aided beamformer training.

BS: initialize feasible $\mathbf{m}_{b,k,l}\ \forall (b,k,l)$ and $n = 1$.
repeat
  BS: transmit $\mathbf{m}_{b,k,l}\mathbf{b}_{k,l}^{\mathrm{H}}$ DL pilots for all $k \in \mathcal{C}_b, l = 1, \ldots, L_k$.
  UE: estimate MMSE receive beamformers from (18).
  UE: compute the MSE $\epsilon_{k,l}^{(n)}\ \forall (k,l)$ from (3).
  UE: set the weights $w_{k,l}^{(n)}\ \forall (k,l)$ from (10).
  UE: transmit UL pilots $\mathbf{u}_{k,l}\sqrt{w_{k,l}}\mathbf{b}_{k,l}^{\mathrm{H}}\ \forall (k,l)$.
  BS: estimate $\mathbf{R}_b\mathbf{R}_b^{\mathrm{H}}$ and $\mathbf{c}_{b,k,l}$ for all (b,k,l).
  BS: solve the precoders $\mathbf{m}_{b,k,l}\ \forall (b,k,l)$ from (14).
  Set $n = n + 1$.
until Desired level of convergence has been reached.

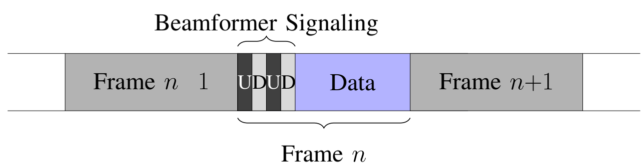

Fig. 2. TDD frame structure with two bi-directional training iterations.

IV. Decentralized Beamformer Design

In this section, we consider decentralized JP beamformer design when the effective channels (16) cannot be accurately estimated using SSE. The beamformer signaling relies crucially on the channel reciprocity of TDD. For further discussion of precoded pilot signaling see [5].

In [4] and [5], 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 (12) is coupled among the BSs due to the coherent signal reception, which prevents us from directly applying the same decentralized processing method. In the sequel, we propose two different approaches for decentralized JP.

A. Iterated Best Response

The BR design employs the parallel optimization scheme proposed in [20] to decentralize the beamformer design. This parallel framework is based on solving for the beamformers locally at 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 [20] 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. From (14) we have

$$\mathbf{c}_{b,k,l}^{(n)} = [\mathbf{R}_b^{(n)}]^{\mathrm{H}}\mathbf{m}_{b,k,l}^{(n)} \in \mathbb{C}^S, \tag{19}$$

where $\mathbf{R}_b^{(n)}$ can either be estimated indirectly using estimated effective channels (SSE) or directly from (11) (DE). Here (n) denotes the iteration index. Again as in (14), 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 coupling variables in (14), the optimal transmit beamformers can be determined from

$$\mathbf{m}_{b,k,l} = \left(\mathbf{R}_b\mathbf{R}_b^{\mathrm{H}} + \mathbf{I}\lambda_b\right)^{-1}\mathbf{R}_b\left(\mathbf{b}_{k,l}\sqrt{w_{k,l}} - [\bar{\mathbf{c}}_{b,k,l}^{(n)}]^{\mathrm{H}}\right) \tag{20}$$

by bisection search over λ_b to satisfy the transmit power constraints $\sum_{k \in \mathcal{C}_b}\sum_{l=1}^{L_k}\|\mathbf{m}_{b,k,l}\|^2 \le P_b$. Note that, if $\sum_{k \in \mathcal{C}_b}\sum_{l=1}^{L_k}\|\mathbf{m}_{b,k,l}\|^2 < P_b$ for $\lambda_b = 0$, then this is the optimal solution. Furthermore, the dimensions of $(\mathbf{R}_b\mathbf{R}_b^{\mathrm{H}} + \mathbf{I}\lambda_b)$ 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 [5].

If the number of transmit antennas is drastically increased, e.g., in massive MIMO systems, solving (20) becomes computationally cumbersome. In [4], a computationally more efficient method was provided that requires only taking the singular value decomposition of $\mathbf{R}_b\mathbf{R}_b^{\mathrm{H}}$. Still, the complexity of the singular value decomposition is significant at higher dimensions [41]. To this end, in Section IV-B, lower computational complexity, gradient-based methods, are proposed.

The iterated BR algorithm is summarized in Algorithm 3.

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

$$\mathbf{m}_{b,k,l}^{(n+1)} = \mathbf{m}_{b,k,l}^{(n)} + \alpha\left(\mathbf{m}_{b,k,l} - \mathbf{m}_{b,k,l}^{(n)}\right) \quad \forall\, (b,k,l), \tag{21}$$

where α is a sufficiently small step-size and $\mathbf{m}_{b,k,l}$ is the optimal solution for (20). Here, 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 [20]. [1]

Signaling requirements: The signaling requirements are apparent from (20). 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 accumulates into $\sum_{k \in \mathcal{C}_b} L_k S$ complex terms per BS. Note that vector $\mathbf{c}_{j,k,l}^{(n)}$ has length S and, thus, there is a trade-off between signaling overhead and performance. Note that the signaling overhead does not depend on the number of transmit antennas. This is significant signaling reduction when compared to full CSI exchange.

For S larger than the number of streams in the system, it is possible to design globally orthogonal pilot sequences, i.e., utilizing the SSE approach along with taking into account the pilot estimation noise due to limited training. This allows more control over interference management and training overhead, as weak interference sources can be neglected and not exchanged between the BSs. This approach has been further investigated in an extended technical report [42]. Basically, the SSE methods requires KL_k complex symbols to be exchanged per UE. Note that global pilot coordination may not be feasible with dense heterogeneous systems, in which case it is always beneficial to take the pilot interference into account.

[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 [20].

Algorithm 3 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 (18).
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 (10) and use $\sqrt{w_{k,l}}\mathbf{u}_{k,l}$ as the UL pilot.
6: *BS*: Solve the precoders $\mathbf{m}_{b,k,l}\ \forall\, (b,k,l)$ from (20).
7: *BS*: Exchange the fixed terms (19) among the cooperating BSs.
8: *BS*: Update the next iteration precoders from (21).
9: Set $n = n + 1$.
10: **until** Desired level of convergence has been reached.

B. Gradient Descent Methods

The BR based decentralized JP techniques have attractive convergence properties. However, the computational complexity, particularly, with large numbers of transmit antennas can be excessive for small cell systems. As a low complexity alternative to the aforementioned approaches, we propose gradient-based methods. These methods are based on updating the transmit beamformers, in each iteration, in the direction of the objective gradient, which greatly simplifies the transceiver processing.

GD updates the beamformers in the direction of the last iteration gradient. We can derive the gradient of (12) in terms of $\mathbf{m}_{b,k,l}$ to be

$$\mathbf{g}_{b,k,l} = -2\mathbf{R}_b\left(\mathbf{b}_{k,l}\sqrt{w_{k,l}} - \mathbf{R}_b^{\mathrm{H}}\mathbf{m}_{b,k,l} - \bar{\mathbf{c}}_{b,k,l}\right). \tag{22}$$

The gradients (22) are coupled. However, only the local composites $\mathbf{m}_{b,k,l}^{\mathrm{H}}\mathbf{R}_b$ need to be shared among the cooperating BSs. This gives us the following beamformer update rule

$$\mathbf{m}_{b,k,l}^{(n+1)} = \mathbf{m}_{b,k,l}^{(n)} + \alpha_b\mathbf{g}_{b,k,l}, \tag{23}$$

where α_b is the gradient update step-size and $\mathbf{g}_{b,k,l}$ denotes the part of (22) corresponding to BS b. The outline of the GD algorithm is given in Algorithm 4.

The beamformer update (23) is not sufficient for accurate beam coordination with JP as it does not take into account the power constraint. That is, (23) may lead to a solution, where the available power budget (P_b) is exceeded. To address this problem, we propose two approaches for power control.

1) Feasible projection: A straightforward approach for power control is to simply scale the beamformers to meet the power constraints. That is, if $\sum_{k\in\mathcal{C}_b}^{K}\sum_{l=1}^{L_k}\|\mathbf{m}_{b,k,l}\|^2 > P_b$ for some $b = 1,\ldots,B$, the corresponding BS scales the beamformers by $\frac{\sqrt{P_b}}{\sum_{k\in\mathcal{C}_b}^{K}\sum_{l=1}^{L_k}\|\mathbf{m}_{b,k,l}\|}$. The problem here is that the scaling is not global in the sense that each BS uses a different scaling. This also changes the direction of the beamformer, which may have a detrimental impact on the performance.

2) Dual decomposition: More sophisticated and better performing power control can be achieved by employing the dual decomposition technique to steer the beamformer updates (23) towards the feasible set. First, the augmented Lagrangian for (12) has the form

$$\min_{\mathbf{m}_{b,k,l}} \sum_{(k,l)} \Bigg(1 - 2\mathrm{Re}\Big\{\sum_{b\in\mathcal{B}_k}\sqrt{w_{k,l}}\,\mathbf{m}_{b,k,l}^{\mathrm{H}}\mathbf{R}_b\mathbf{b}_{k,l}\Big\} + \Big(\sum_{b\in\mathcal{B}_k}\mathbf{m}_{b,k,l}^{\mathrm{H}}\mathbf{R}_b\Big)\Big(\sum_{b\in\mathcal{B}_k}\mathbf{R}_b^{\mathrm{H}}\mathbf{m}_{b,k,l}\Big)\Bigg) + \sum_{b=1}^{B}\lambda_b\Bigg(\sum_{k\in\mathcal{C}_b}^{K}\sum_{l=1}^{L_k}\|\mathbf{m}_{b,k,l}\|^2 - P_b\Bigg). \quad (24)$$

where $\lambda_b, b = 1,\ldots,B$ are the dual variables corresponding to the power constraints. Taking the gradient of (24), we get

$$\mathbf{g}_{b,k,l}^{(n)} = \mathbf{g}_{b,k,l}^{(n)} + \lambda_b\mathbf{m}_{b,k,l}, \quad (25)$$

Now, the GD beamformer update becomes

$$\mathbf{m}_{b,k,l}^{(n+1)} = \mathbf{m}_{b,k,l}^{(n)} - \alpha\,\mathbf{g}_{b,k,l}^{(n)}. \quad (26)$$

To steer the beamformer updates towards feasible power levels, after each update (26), we update the dual variables as

$$\lambda_b^{(n+1)} = \max\Bigg(0, \lambda_b^{(n)} + \alpha\Bigg(P_b - \sum_{k\in\mathcal{C}_b}^{K}\sum_{l=1}^{L_k}\|\mathbf{m}_{b,k,l}^{(n)}\|^2\Bigg)\Bigg), \quad (27)$$

where α is a sufficiently small step size.

Regularized updates: As the GD updates are based solely on the currently available gradient, these updates can be, in some cases, overly aggressive. Step-size normalization is the most straightforward way to regularize the absolute step-size given as

$$\tilde{\alpha}_{k,l} = \frac{\alpha}{\|\mathbf{g}_{k,l}^{(n)}\|^2}, \quad (28)$$

where $\mathbf{g}_{k,l}^{(n)}$ is the full gradient vector for (k,l).

Another way to regularize the stochastic gradient (SG) updates, is to make the gradient update more dependent on the previous update direction. In other words this adds momentum for the general update direction. The momentum is adaptively updated as

$$\mathbf{m}_{b,k,l}^{(n+1)} = \mathbf{g}_{b,k,l} + \omega\,\mathbf{m}_{b,k,l}^{(n)}, \quad (29)$$

where $\omega \geq 0$ denotes the momentum magnitude. In principle this is close to the regularized BR update procedure in (21). Finally, the beamformer update becomes

$$\mathbf{m}_{b,k,l}^{(n+1)} = \mathbf{m}_{b,k,l}^{(n)} - \alpha\,\mathbf{m}_{b,k,l}^{(n+1)}. \quad (30)$$

The regularized update routines are particularly helpful in fading channels, where the gradient of the instantaneous channel realization may not fully represent the overall fading conditions. This is demonstrated by numerical examples in Section V.

The computational complexity of the GD design is significantly lower than the BR method or centralized SOCP formulation. Note that the GD approach does not involve an iterative matrix inversion.

Signaling Requirements: The signaling requirements are the same as for the BR design.

Algorithm 4 GD algorithm.

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 (5).
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 (10).
6: *BS*: Update the precoders $\mathbf{m}_{b,k,l}(i)\ \forall\,(b,k,l)$ from (26).
7: *BS*: Update the duals from (27)
8: **until** Desired level of convergence has been reached.

C. Stochastic Gradient

Instead of trying to estimate the complete gradient and update the beamformers only once per pilot *sequence*, they can be updated on each received pilot *symbol*. Since (22) is a linear relation, the complete training matrices $\mathbf{R}_b$ do not need to be available at the BSs before the backhaul signaling can start. That is, (22) can be split into training symbol level updates

$$\mathbf{g}_{b,k,l}(i) = -2\mathbf{R}_b(i)\Big(\mathbf{b}_{k,l}(i)\sqrt{w_{k,l}} - \mathbf{R}_b(i)^{\mathrm{H}}\mathbf{m}_{b,k,l}(i)\Big) + 2\mathbf{R}_b(i)\mathbf{c}_{b,k,l}(i), \quad (31)$$

where $\mathbf{R}_j(i)$ denotes the i^{th} column vector of $\mathbf{R}_j$, $\mathbf{b}_{k,l}(i)$ is the i^{th} element of vector $\mathbf{b}_{k,l}$ and $\mathbf{c}_{j,k,l}(i) = [\mathbf{R}_j(i)]^{\mathrm{H}}\mathbf{m}_{j,k,l}(i)$. Per each training sample (symbol), the beamformers are updated as

$$\mathbf{m}_{b,k,l}^{(n)}(i+1) = \mathbf{m}_{b,k,l}^{(n)}(i) - \alpha\,\mathbf{g}_{b,k,l}^{(n)}(i). \quad (32)$$

This, along with the reduced computational complexity (no matrix inversion required), can be used to reduce the signaling delays even with limited computational resources.

Signaling Requirements: The total signaling requirements are somewhat increased when compared to the GD design. For each pilot symbol, the $\mathbf{c}_{b,k,l}(i)$ terms need to be exchanged among the BSs. To reduce the signaling overhead, (31) can be exploited by averaging over multiple iterations i and signaling over the averaged values, thus, not sharing all S symbols, but an averaged subset $\sum_{i=a}^{b}\frac{\mathbf{c}_{b,k,l}(i)}{b-a}$ for some interval $[a,b]$.

D. Feedback Quantization

In all proposed decentralized schemes, the feedback signaling information has to be quantized before it is exchanged over a feedback channel or the backhaul. This is equivalent to separately quantizing the I/Q branches of the $d_{b,k,l}$ terms for the proposed methods. Thus, robustness to the quantization errors is crucial for any design realizable in practice. In addition, quantization reduces the backhaul utilization. In Section V, we study the performance of the proposed beamformer design algorithms with q-bit quantization of the feedback information.

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. If not otherwise mentioned, the number of transmit and receive antennas are set to $N_T = 4$ and $N_R = 2$, 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 interference will limit the number of active users per-BS. The number of active spatial streams per users is limited to one. The simulation environment is illustrated in Fig. 3. The SNR is defined on

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

the cell edge from the closest BS b, i.e, $\mathrm{SNR} = \frac{g_{b,k} P_b}{\sigma_k^2}$, where $g_{b,k}$ denotes the corresponding path loss. The SNR is fixed to 15dB. The pilot training sequences are random binary sequences with 10dB pilot power gain over the SNR. This reflects the scenario, where there is no pilot planning over the BSs and none of the pilot resources are made orthogonal. Joint transmission is allowed over the all BSs and, as such, there are no distinct joint transmission clusters ($\mathcal{C}_b = \mathcal{K}; b = 1, \ldots, B$). 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. For specific conditions the results could be somewhat improved by fine tuning the parameters. The default for the BR design is 0.25. The GD beamformers are generated with normalization, 0.25 momentum, 0.005 dual step-size and 0.25 beamformer step-size. Similarly, the SG beamformers are generated without normalization, $= 0.25$ momentum, $= 0.005$ dual step-size and $= 0.25$ beamformer step-size.

The channels are generated with Jakes' Doppler spectrum model. The channel coherence time is defined by normalized user terminal velocity $t_s f_d$, 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 lengths shorter than 0.25ms [43]. Assuming backhaul info exchange may happen once per frame, the normalized velocity $t_s f_d = 0.01$ equals to 20km/h at 5 GHz carrier frequency. 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 is a stream specific estimation method such as proposed in [5]. As a general performance upper bound we use SSE with fully orthogonal pilot allocation without pilot estimation noise. This is denoted by SSE (ideal). A summary of the simulation parameters is given in Table I. The SSE reference designs are available in a separate report [42], which contains distributed JP CoMP designs for perfect effective CSI and an additional alternating direction method of multipliers (ADMM) based approach for the proposed system model as well.

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	10dB/20dB
Distance between adjacent BSs	600m
The path loss exponent	3

The bi-directional signaling overhead is considered using coefficient $= 0.05$, so that the achievable rate is defined as $(1 \quad)R$. The number of UL/DL signaling iterations is set to $T_{bf} = 3$. We assume that the stream specific weights $w_{k,l}\ \forall (k,l)$ can be exchanged only once per frame and the backhaul information exchange ($_{b,k,l}\ \forall (b,k,l)$) can be done T_{bf} times per frame. That is, the bidirectional iteration, within a frame, only involves TDD based beamformer signaling.

Fig. 4 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 all pilots can be made fully orthogonal, which reduces the

Fig. 4. Behavior for varying training sequence lengths in constant channel.

Fig. 5. Comparison of the proposed decentralized methods with varying training sequence lengths with constant channels.

performance gap with large pilot lengths. Even in that case, the pilot estimation noise should be included in the beamformer estimates.

An overview of the impact of the pilot sequence lengths on the proposed decentralized schemes is shown in Fig. 5. The BR quickly approaches the orthogonal upper bound as the pilot sequences are made longer. The GD methods do not achieve the orthogonal SSE rate even at very large sequence lengths. This is due to constant step-size, which make the algorithm oscillate around a stationary point. Since the pilot training vector power is fixed, there is always a cap between the estimation techniques and ideal case due to the estimation noise.

Similarly, in Fig. 6, the achievable rate for the proposed schemes is shown when the number of transmit antennas is increased to $N_T = 8$. It can be seen that the overall performance in all schemes is improved. In particular, the performance with long training sequence lengths is now close to optimal. The most noticeable improvement can be seen with the GD and SG that now out-perform the BR design at small training sequence lengths. As the number of antennas in the system is increased with respect to the number of active streams, the interference management becomes easier and performance is improved. Also, shorter training sequences are sufficient as there is less interference.

Fig. 6. Comparison of the proposed decentralized methods with varying training sequence lengths with constant channels with $N_T = 8$ transmit antennas.

The convergence behavior in constant and time correlated fading ($t_s f_d = 0.01$) channels is shown in Fig. 7. Reducing the training sequence length shifts the performance curve, but does not significantly change the shape of the convergence curve. In time correlated fading channels, the trend is that the performance saturates after 20 iterations. The GD and SG designs have similar, but somewhat slower, convergence behavior. Thus, the corresponding convergence figures are neglected.

Robustness to UE mobility can be seen from Fig. 8. Here, the JP CoMP performance is compared to the CB beamformer design. It can be seen that BR designs are robust to UE mobility. However, the gradient based designs have quickly deteriorating performance as the UE velocities increase. It should be noted that the SG design has comparable performance to the BR method in slow channel fading conditions. The performance of the GD methods could be improved with velocity dependent step-size selection. Also, the gain from bi-direction signaling is shown here with the SSE and $T_{bi} = 1$. It is evident that, as the mobility grows, improved beamformer training is highly beneficial.

Fig. 9 shows the impact of feedback information quantization to the convergence behavior with varying levels of quantization. The I/Q branches of each backhaul data symbol in $c_{b;k;l}$ $\forall(b;k;l)$ are separately quantized with Q-bit quantization as discussed in Section IV-D. Symbol-by-symbol beamformer iteration of the SG method provides significant gain at lower quantization levels. From here, we can also observe

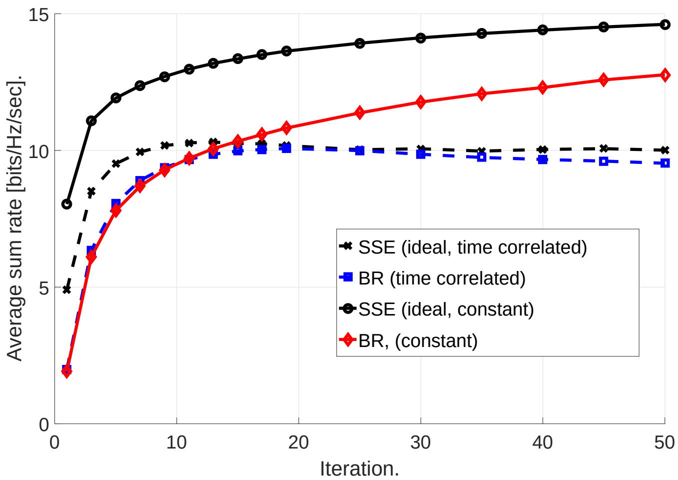

Fig. 7. BR performance for varying training sequence lengths in constant and time correlated (normalized UE velocity $t_s f_d = 0{:}01$) channels.

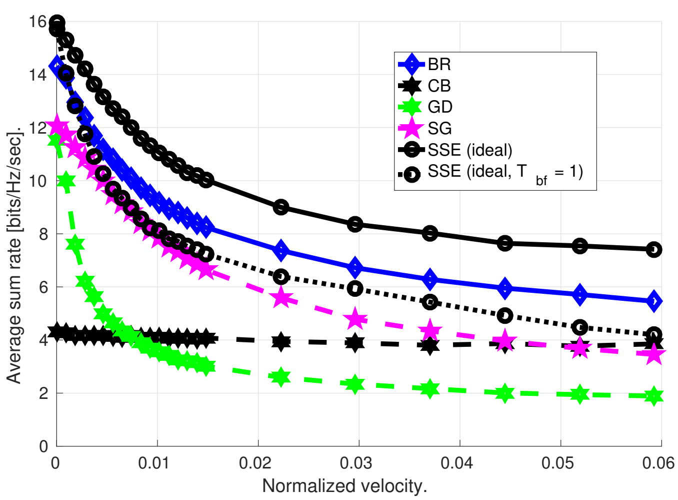

Fig. 8. Saturated performance comparison with normalized UE velocities.

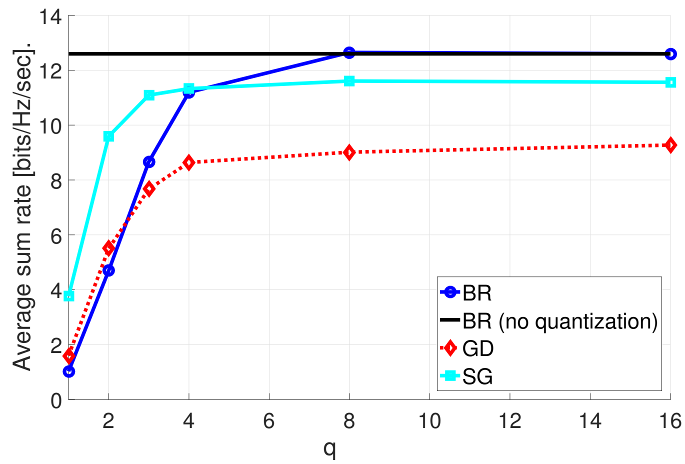

Fig. 9. Feedback signaling with q-bit quantization.

that from $q = 4$ bit quantization already achieves maximum performance. Note that the upper bound is the BR performance with 41 symbol training pilots and no quantization.

VI. Conclusions

We have proposed decentralized transceiver designs for JP CoMP WSRMax in the presence of non-orthogonal pilot resources and pilot estimation noise. Emphasis was given to designs that enable 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. The BR approach was used to provide a JP algorithm with attractive convergence and performance properties. As a less complex alternative, GD and SG based transmit beamformer design were also proposed. The implementation complexity and performance trade-off was studied by numerical evaluation. The numerical results indicated that BR and SG designs provide good performance and stability even with moderately fast fading channel conditions. The GD approach provided reduced computational complexity w.r.t. the BR method, but was not as robust to UE mobility as the SG design.

References

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

[2] J. Kaleva, A. Tölli, M. J. Juntti, R. Berry, and M. L. Honig, “Decentralized joint precoding with pilot-aided beamformer estimation,” in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2015, pp. 1–6.

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

[4] 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.

[5] 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.

[6] A. Tölli, H. Pennanen, and P. Komulainen, “Decentralized minimum power multi-cell beamforming with limited backhaul signaling,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 570–580, Feb. 2011.

[7] T. Bogale and L. Vandendorpe, “Weighted sum rate optimization for downlink multiuser MIMO coordinated base station systems: Centralized and distributed algorithms,” *IEEE Trans. Signal Processing*, Dec. 2011.

[8] A. Tölli, M. Codreanu, and M. Juntti, “Cooperative MIMO-OFDM cellular system with soft handover between distributed base station antennas,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1428–1440, Apr. 2008.

[9] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, “Multi-Cell MIMO Cooperative Networks: A New Look at Interference,” *IEEE J. Select. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, 2010.

[10] S. Zhou, J. Gong, and Z. Niu, “Distributed Adaptation of Quantized Feedback for Downlink Network MIMO Systems,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 61–67, Jan. 2011.

[11] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana, “Coordinated multipoint transmission and reception in LTE-Advanced: deployment scenarios and operational challenges,” *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 148–155, Feb. 2012.

[12] 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.

[13] J. Zhang and J. Andrews, “Adaptive spatial intercell interference cancellation in multicell wireless networks,” *IEEE J. Select. Areas Commun.*, vol. 28, no. 9, pp. 1455–1468, 2010.

[14] S. Han, C. Yang, G. Wang, D. Zhu, and M. Lei, “Coordinated Multi-Point Transmission Strategies for TDD Systems with Non-Ideal Channel Reciprocity,” *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4256–4270, Oct. 2013.

[15] T. M. Kim, F. Sun, and A. Paulraj, "Low-Complexity MMSE Precoding for Coordinated Multipoint With Per-Antenna Power Constraint," *IEEE Signal Processing Lett.*, vol. 20, no. 4, pp. 395–398, 2013.

[16] S. Shi, M. Schubert, and H. Boche, "Rate optimization for multiuser mimo systems with linear processing," *IEEE Trans. Signal Processing*, vol. 56, no. 8, pp. 4020–4030, Aug. 2008.

[17] M. Codreanu, A. Tölli, M. Juntti, and M. Latva-aho, "Joint design of Tx-Rx beamformers in MIMO downlink channel," *IEEE Trans. Signal Processing*, vol. 55, no. 9, pp. 4639–4655, Sep. 2007.

[18] S. S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[19] C. Shen, T.-H. Chang, K.-Y. Wang, Z. Qiu, and C.-Y. Chi, "Distributed robust multicell coordinated beamforming with imperfect CSI: An ADMM approach," *IEEE Trans. Signal Processing*, vol. 60, no. 6, pp. 2988–3003, Jun. 2012.

[20] 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.

[21] J. Kaleva, A. Tölli, and M. Juntti, "Decentralized sum rate maximization with QoS constraints for interfering broadcast channel via successive convex approximation," *IEEE Trans. Signal Processing*, vol. 64, no. 11, pp. 2788–2802, Jun. 2016.

[22] 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.

[23] 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.

[24] 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.

[25] B. Dai and W. Yu, "Energy efficiency of downlink transmission strategies for cloud radio access networks," *IEEE J. Select. Areas Commun.*, vol. 34, no. 4, pp. 1037–1050, Apr. 2016.

[26] F. Zhuang and V. Lau, "Backhaul limited asymmetric cooperation for MIMO cellular networks via semidefinite relaxation," *IEEE Trans. Signal Processing*, vol. 62, no. 3, pp. 684–693, Feb. 2014.

[27] 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.

[28] J. Kaleva, M. Bande, A. Tölli, M. Juntti, and V. V. Veeravalli, "Sum Rate Maximizing Joint Processing with Limited Backhaul and Tree Topology Constraints," in *Proc. IEEE Works. on Sign. Proc. Adv. in Wirel. Comms.*, Edinburgh, UK, Jul. 2016.

[29] S. H. Park, O. Simeone, O. Sahin, and S. Shamai, "Inter-cluster design of precoding and fronthaul compression for cloud radio access networks," *IEEE Commun. Lett.*, vol. 3, no. 4, pp. 369–372, Aug. 2014.

[30] S. H. Park, O. Simeone, O. Sahin, and S. S. Shitz, "Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory," *IEEE Signal Processing Mag.*, vol. 31, no. 6, pp. 69–79, Nov. 2014.

[31] D. Kim, O.-S. Shin, I. Sohn, and K. B. Lee, "Channel Feedback Optimization for Network MIMO Systems," *IEEE Trans. Veh. Technol.*, vol. 61, no. 7, pp. 3315–3321, Sep. 2012.

[32] D. Jaramillo-Ramrez, 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.

[33] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell tdd systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[34] 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.

[35] M. Xu, D. Guo, and M. L. Honig, "Distributed bi-directional training of nonlinear precoders and receivers in cellular networks," *IEEE Trans. Signal Processing*, vol. 63, no. 21, pp. 5597–5608, Nov. 2015.

[36] G. Venkatraman, A. Tölli, M. Juntti, and L. N. Tran, "Traffic aware resource allocation schemes for multi-cell MIMO-OFDM systems," *IEEE Trans. Signal Processing*, vol. 64, no. 11, pp. 2730–2745, Jun. 2016.

[37] 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.

[38] N. Vucic, S. Shi, and M. Schubert, "DC programming approach for resource allocation in wireless networks," in *Proc. Int. Workshop Res. Alloc. Wireless Netw.*, Avignon, France, Jun. 4 2010, pp. 360–366.

[39] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Journal of the Operations Research Society of America*, vol. 26, no. 4, pp. 681–683, Jul. –Aug. 1978.

[40] 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.

[41] G. H. Golub and C. F. V. Loan, *Matrix Computations, 3rd ed.* Baltimore: The Johns Hopkins University Press, 1996.

[42] J. Kaleva, A. Tölli, M. J. Juntti, R. Berry, and M. L. Honig, "Joint transmission with limited backhaul connectivity," Tech. Rep., https://arxiv.org/abs/1705.05252 2016.

[43] 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.

Jarkko Kaleva (S'11) received his M.Sc. (Tech.) degree in information engineering from University of Oulu, Oulu, Finland in 2011. In 2010, he joined Centre for Wireless Communications (CWC) at University of Oulu, Finland, where he is currently working as a research scientist. From June 2011, he has been working towards the Dr.Sc. (Tech.) degree. His main research interests are in radio resource management for wireless communication and nonlinear programming.

Antti Tölli (M'08, SM'14) received the Dr.Sc. (Tech.) degree in electrical engineering from the University of Oulu, Oulu, Finland, in 2008. Before joining the Centre for Wireless Communications (CWC) at the University of Oulu, he worked for 5 years with Nokia Networks as a Research Engineer and Project Manager both in Finland and Spain. In May 2014, he was granted a five year (2014-2019) Academy Research Fellow post by the Academy of Finland. He also holds an Adjunct Professor position with University of Oulu. During the academic year 2015-2016, he visited at EURECOM, Sophia Antipolis, France. He has authored more than 150 papers in peer-reviewed international journals and conferences and several patents all in the area of signal processing and wireless communications. His research interests include radio resource management and transceiver design for broadband wireless communications with a special emphasis on distributed interference management in heterogeneous wireless networks. He is currently serving as an associate editor for IEEE Transactions on Signal Processing.