

Gradient Based Decentralized Joint Beamforming

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

[†]Centre for Wireless Communications, Univ. of Oulu, P.O. Box 4500, 90014, Univ. of Oulu, Finland

[‡]Dept. of Elect. Engineering and Comp. Science, Northwestern University, Evanston, IL, 60208, U.S.

Email: jarkko.kaleva@oulu.fi

Abstract—Gradient based downlink beamforming with low computational complexity and training overhead is proposed for joint processing (JP) coordinated multi-point transmission (CoMP). Pilot contamination and estimation noise are taken into account in the pilot based transceiver training process. The proposed designs enable decentralized JP when the backhaul and computational limitations do not allow centralized processing. The impact of backhaul quantization is also considered. The stochastic gradient based designs are shown to be more robust to feedback quantization when compared to more complex methods. The trade-off between the implementation complexity and performance is established for the proposed algorithms. The results show that low complexity decentralized JP CoMP is feasible even with limited backhaul capacity.

I. INTRODUCTION

The ever increasing need for spectral efficiency, imposes the demand for effective interference management and transmission coordination. A lot of research efforts have been invested in coordinated beamforming (CB) for multi-cell systems. However, Multi-cell CB is still in the initial stages with respect to the current Long Term Evolution Advanced (LTE-A) standard [1]. Much of this research has focused on decentralized coordination strategies [2]–[4]. Joint processing (JP) coordinated multi-point transmission (CoMP) allows joint transmission from cooperating base stations (BSs), which greatly increases the available degrees of freedom (DoF) [5]. The practical limitations in BS connectivity are still hindering implementation of the JP CoMP transmission.

In this paper, low complexity JP CoMP methods are presented for weighted sum rate maximization (WSR-Max). We focus on a scenario where the limited backhaul prevents channel state information (CSI) sharing among the cooperating BSs. The limited backhaul connectivity only supports centralized data sharing and limited control signaling. Furthermore, we assume non-orthogonal and noisy pilots, which is expected in dense deployments. The transmit beamformers are locally designed in each BS. This reduces the overall backhaul load. We propose gradient descent (GD) and stochastic gradient (SG) schemes for decentralized JP CoMP beamforming. These

methods are shown to have comparable performance to more complex best response (BR) design.

In centralized JP, the fronthaul/backhaul information can be handled in one of two ways: (i) In *data sharing*, the remote central processor (RCP) exchanges the messages with the cooperating BSs explicitly and the joint beamformers are exchanged over the backhaul separately from the data [6]. (ii) Using *compression*, the messages are precoded beforehand at the RCP and only the compressed versions of the analog beamformers are sent over the backhaul [7], thus avoiding the separate data exchange. Sparse JP designs are among the most common approaches for fronthaul/backhaul limited CoMP [6]. These designs try to limit the JP cluster sizes and, thus, implicitly reduce the data sharing overhead. Different aspects and benefits of compression approach are studied in [7]. Our focus is on decentralized data sharing JP with low signaling overhead.

Unlike JP CoMP, WSRMax for CB has been extensively studied for decentralized inter-cell interference coordination [2], [3], [8], [9]. Particularly, the weighted minimum mean-squared error (WMMSE) has been shown to have a convenient structure for decentralized processing in time division duplex (TDD) [2], [3]. JP transmission is inherently coupled. Thus, the CB methods are not directly applicable to this scenario. In [10] heuristic JP CoMP schemes that make use of only local CSI are proposed. Without global CSI, these schemes cannot achieve all available DoF.

Pilot non-orthogonality and contamination have been widely studied, albeit, not for JP CoMP. Pilot contamination in TDD based transceiver training for CB has been considered, e.g., in [11]. In [11], 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 channels individually. Here, we show that similar conclusions hold for decentralized JP CoMP.

II. SYSTEM MODEL

We consider a downlink system with B BSs each equipped with N_T transmit antennas. There are, in to-

tal, K user equipments (UEs) each with N_R receive antennas. Each UE $k = 1, \dots, K$ is coherently served by $|\mathcal{B}_k|$ BSs, where the set \mathcal{B}_k defines the JP cluster (set of phase-coherent serving BSs) for UE k . Similarly, the set of UE indices served by BS $b = 1, \dots, B$ is denoted by $\mathcal{C}_b = \{k | b \in \mathcal{B}_k\}$. The number of spatial data streams allocated to UE $k = 1, \dots, K$ is denoted by L_k . To simplify the notation, we use the following set abbreviations: $(k, l) \triangleq \{(k, l) | k = 1, \dots, K, l = 1, \dots, L_k\}$ and $(b, k, l) \triangleq \{(b, k, l) | k = 1, \dots, K, l = 1, \dots, L_k, b \in \mathcal{B}_k\}$. The downlink transmission within the JP set is considered to be symbol synchronous in the sense that the transmitted symbols from each $\mathcal{B}_k, k = 1, \dots, K$ are coherently combined at each UE. Each BS $b = 1, \dots, B$ is only aware of the local channel matrix $\mathbf{H}_{b,k} \in \mathbb{C}^{N_R \times N_T} \forall k = 1, \dots, K$, while data sharing is assumed within each serving set \mathcal{B}_k . Furthermore, we assume TDD, which is used to share the effective UL/DL CSI.

The received signal at UE $k = 1, \dots, K$ 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 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 $d_{k,l} \forall (k, l)$ are assumed to be independent and identically distributed (i.i.d.) with $\mathbb{E}\{|d_{k,l}|^2\} = 1$.

The estimated symbol at UE k over stream l , after applying the 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

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

and the corresponding mean-squared error (MSE) is

$$\epsilon_{k,l} \triangleq \left| \sum_{b \in \mathcal{B}_k} \mathbf{u}_{k,l}^H \mathbf{H}_{b,k} \mathbf{m}_{b,k,l} - 1 \right|^2 + \|\mathbf{u}_{k,l}\|^2 \sigma_k^2 + \sum_{i=1}^K \sum_{\substack{j=1, \\ (i,j) \neq (k,l)}}^{L_i} \left| \sum_{b \in \mathcal{B}_i} \mathbf{u}_{k,l}^H \mathbf{H}_{b,k} \mathbf{m}_{b,i,j} \right|^2. \quad (3)$$

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

III. CENTRALIZED WSRMAX SOLUTION

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

stated as

$$\max_{\mathbf{u}_{k,l}, \mathbf{m}_{b,k,l}} \sum_{k=1}^K \sum_{l=1}^{L_k} \mu_k \log_2 (1 + \Gamma_{k,l}) \quad (4)$$

$$\text{s. t.} \quad \sum_{\substack{l=1, \\ k \in \mathcal{C}_b}}^{L_k} \|\mathbf{m}_{b,k,l}\|^2 \leq P_b, \quad b = 1, \dots, B, \quad (5)$$

where $\mu_k, k = 1, \dots, K$ are the user priority weights. The optimal rate maximizing minimum mean-squared error (MMSE) receive beamformers are given as

$$\mathbf{u}_{k,l} = \mathbf{K}_k^{-1} \left(\sum_{b \in \mathcal{B}_k} \mathbf{H}_{b,k} \mathbf{m}_{b,k,l} \right), \quad (6)$$

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

It is well-known that, when the MMSE receivers are used, there is an inverse relation between the SINR and the corresponding MSE [2] given by $\epsilon_{k,l}^{-1} = 1 + \Gamma_{k,l}$. Now, we can equivalently formulate (4) as

$$\min_{\mathbf{u}_{k,l}, \mathbf{m}_{b,k,l}} \sum_{k=1}^K \sum_{l=1}^{L_k} \mu_k \log_2 (\epsilon_{k,l}) \quad \text{s. t. (5)}. \quad (7)$$

Since (7) is not jointly convex for the transmit and receive beamformers, we alternate between solving for the transmit and receive beamformers. This separation is convenient for TDD processing as the DL/UL are temporally separated. With fixed transmit beamformers $\mathbf{m}_{b,k,l} \forall (b, k, l)$, the optimal receive beamformer can be obtained from (6). 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 [2].

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 $\log_2(\epsilon_{k,l}) \approx (\epsilon_{k,l} - \epsilon_{k,l}^{(n)}) / (\log(2) \epsilon_{k,l}^{(n)}) + \log_2(\epsilon_{k,l}^{(n)})$. Ignoring the constant terms, the approximated transmit beamformer design subproblem can be restated as a WMMSE problem

$$\min_{\mathbf{m}_{b,k,l}} \sum_{k=1}^K \sum_{l=1}^{L_k} w_{k,l}^{(n)} \epsilon_{k,l} \quad \text{s. t. (5)}, \quad (8)$$

where $w_{k,l}^{(n)} = \frac{\mu_k}{\log(2) \epsilon_{k,l}^{(n)}} \forall (k, l)$. As shown in [2], the successive approximation algorithm provides monotonic convergence of the objective function and convergence to a local stationary point of the original problem (4).

We now formulate 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 l^{th} 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 at BS b is

$$\mathbf{R}_b = \sum_{k=1}^K \sum_{l=1}^{L_k} \mathbf{H}_{b,k}^H \mathbf{u}_{k,l} \sqrt{w_{k,l}} \mathbf{b}_{k,l}^H + \mathbf{N}_b, \quad (9)$$

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

Assuming only the UL training signal (9) at the BSs, we can rewrite the WSRMax problem (8) by substituting the effective channels in (3) by the corresponding estimates as

$$\begin{aligned} \min_{\mathbf{m}_{b,k,l}} & \sum_{(k,l)} \left(1 - 2\text{Re}\left\{ \sum_{b \in \mathcal{B}_k} \sqrt{w_{k,l}} \mathbf{m}_{b,k,l}^H \mathbf{R}_b \mathbf{b}_{k,l} \right\} + \right. \\ & \left. \left(\sum_{b \in \mathcal{B}_k} \mathbf{m}_{b,k,l}^H \mathbf{R}_b \right) \left(\sum_{b \in \mathcal{B}_k} \mathbf{R}_b^H \mathbf{m}_{b,k,l} \right) \right) \\ \text{s. t.} & \sum_{k \in \mathcal{C}_b} \sum_{l=1}^{L_k} \|\mathbf{m}_{b,k,l}\|^2 \leq P_b, \quad b = 1, \dots, B. \end{aligned} \quad (10)$$

Problem (10) 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 [3]. In the following section, we will exploit this relation to derive an efficient decentralized JP beamforming algorithm.

Uplink beamformer estimation

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

$$\mathbf{T}_k = \sum_{i=1}^K \sum_{l=1}^{L_i} \left(\sum_{b \in \mathcal{B}_i} \mathbf{H}_{b,k} \mathbf{m}_{b,i,l} \right) \mathbf{b}_{i,l} + \mathbf{N}_k. \quad (11)$$

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

$$\mathbf{u}_{k,l} = (\mathbf{T}_k \mathbf{T}_k^H + \mathbf{I} \sigma_k^2)^{-1} \mathbf{T}_k \mathbf{b}_{k,l}^H. \quad (12)$$

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 (10). MMSE receive beamformer estimation (12) 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. The beamformer signaling relies crucially on the channel reciprocity of TDD. For further discussion of precoded pilot signaling see [3]. In [2], [3], it was shown that, in CB, the WMMSE algorithm is inherently decoupled. As such, it can be easily decentralized

with low signaling overhead. However, the JP transmit beamformer design in (10) is coupled among the BSs due to the coherent signal reception, which prevents us from directly applying the same approach.

A. Gradient Descent

We propose GD for low computational complexity decentralized JP CoMP. This method is based on updating the transmit beamformers in the direction of the objective gradient, which greatly simplifies the transceiver processing. The GD methods are known to have slower rate of convergence than the more advanced algorithms. However, for practical scenarios with limited number of iterations, most of the gain is achieved with few first iterations. The convergence behavior is studied in more detail in extended technical report [12].

GD algorithm updates the beamformers in the direction of the last iteration gradient. The gradient of (10) can be derived in terms of $\mathbf{m}_{b,k,l}$ to be

$$\mathbf{g}_{b,k,l} = -2\mathbf{R}_b (\mathbf{b}_{k,l} \sqrt{w_{k,l}} - \mathbf{R}_b^H \mathbf{m}_{b,k,l} - \bar{\mathbf{c}}_{b,k,l}), \quad (13)$$

where $\bar{\mathbf{c}}_{b,k,l} = \sum_{j \in \mathcal{B}_k \setminus \{b\}} \mathbf{R}_j^H \mathbf{m}_{j,k,l}$. Note that the gradients (13) are still coupled among the serving BSs. However, only the local composites $\mathbf{m}_{b,k,l}^H \mathbf{R}_b$ need to be shared among the cooperating BSs. This gives us the following beamformer update rule

$$\mathbf{m}_{b,k,l}^{(n+1)} = \mathbf{m}_{b,k,l}^{(n)} - \alpha_b \mathbf{g}_{b,k,l}, \quad (14)$$

where α_b is the gradient update step-size and $\mathbf{g}_{b,k,l}$ denotes the part of (13) corresponding to BS b . The outline of the GD algorithm is given in Algorithm 1.

The beamformer update (14) is not sufficient for accurate beam coordination with JP as it does not take into account the power constraint. That is, (14) may lead to a solution, where the available power budget (5) is exceeded. To address this problem, we propose dual decomposition based power control to steer the beamformer updates (14) towards the feasible set. First, the augmented Lagrangian for (10) has the form

$$\begin{aligned} \min_{\mathbf{m}_{b,k,l}} & \sum_{(k,l)} \left(1 - 2\text{Re}\left\{ \sum_{b \in \mathcal{B}_k} \sqrt{w_{k,l}} \mathbf{m}_{b,k,l}^H \mathbf{R}_b \mathbf{b}_{k,l} \right\} + \right. \\ & \left. \left(\sum_{b \in \mathcal{B}_k} \mathbf{m}_{b,k,l}^H \mathbf{R}_b \right) \left(\sum_{b \in \mathcal{B}_k} \mathbf{R}_b^H \mathbf{m}_{b,k,l} \right) \right) + \\ & \sum_{b=1}^B \nu_b \left(\sum_{k \in \mathcal{C}_b} \sum_{l=1}^{L_k} \|\mathbf{m}_{b,k,l}\|^2 - P_b \right). \end{aligned} \quad (15)$$

where $\nu_b, b = 1, \dots, B$ are the dual variables corresponding to the power constraints. Taking the gradient of (15), we get

$$\bar{\mathbf{g}}_{b,k,l}^{(n)} = \mathbf{g}_{b,k,l}^{(n)} + \nu_b \mathbf{m}_{b,k,l}, \quad (16)$$

Now, the GD beamformer update becomes

$$\mathbf{m}_{b,k,l}^{(n+1)} = \mathbf{m}_{b,k,l}^{(n)} - \alpha \bar{\mathbf{g}}_{b,k,l}^{(n)}. \quad (17)$$

The beamformers are steered towards feasible power levels, after each update (17), by updating the duals as

$$\nu_b^{(n+1)} = \max(0, \nu_b^{(n)} + \beta(P_b - \sum_{k \in \mathcal{C}_b} \sum_{l=1}^{L_k} \|\mathbf{m}_{b,k,l}^{(n)}\|^2)), \quad (18)$$

where β is a sufficiently small step-size.

As the GD is based solely on the currently available gradient, these updates can be, in some cases, overly aggressive. Step-size normalization $\tilde{\alpha}_{k,l} = \alpha / \|\mathbf{g}_{k,l}^{(n)}\|$, where $\mathbf{g}_{k,l}^{(n)}$ the full gradient vector for (k, l) can be used to regularize the updates.

Another way to regularize the GD 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

$$\tilde{\mathbf{m}}_{b,k,l}^{(n+1)} = \mathbf{g}_{b,k,l} + \omega \tilde{\mathbf{m}}_{b,k,l}^{(n)}, \quad (19)$$

where $\omega \geq 0$ denotes the momentum magnitude. Finally, the beamformer update becomes

$$\mathbf{m}_{b,k,l}^{(n+1)} = \mathbf{m}_{b,k,l}^{(n)} - \alpha \tilde{\mathbf{m}}_{b,k,l}^{(n+1)}. \quad (20)$$

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. The computational complexity of the GD design is significantly lower than the centralized second order cone program (SOCP) formulation [3].

The signaling requirements are apparent from (13). Each BS b requires the knowledge of $\mathbf{c}_{j,k,l} = \mathbf{R}_j^H \mathbf{m}_{j,k,l}$ 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 $\mathbf{c}_{j,k,l}$ has length S and, thus, there is a tradeoff between signaling overhead and performance. Note also that the signaling overhead does not depend on the number of transmit antennas.

Algorithm 1 GD algorithm.

- 1: Initialize feasible $\mathbf{m}_{b,k,l} \forall (b, k, l)$ and $n = 1$.
 - 2: **repeat**
 - 3: Solve the MMSE receivers $\mathbf{u}_{k,l} \forall (k, l)$ from (6).
 - 4: Compute the MSE $\epsilon_{k,l}^{(n)}$ and $w_{k,l}^{(n)} \forall (k, l)$ from (3).
 - 5: Update the precoders $\mathbf{m}_{b,k,l} \forall (b, k, l)$ from (17).
 - 6: Update the duals from (18)
 - 7: **until** Desired level of convergence has been reached.
-

B. 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 (13) 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, (13) can be split into symbol level updates

$$\begin{aligned} \mathbf{g}_{b,k,l}(i) = & -2\mathbf{R}_b(i) (\mathbf{b}_{k,l}(i)\sqrt{w_{k,l}} - \mathbf{R}_b(i)^H \mathbf{m}_{b,k,l}(i)) \\ & + 2\mathbf{R}_b(i) \tilde{\mathbf{c}}_{b,k,l}(i) + \nu_b \mathbf{m}_{b,k,l}(i), \end{aligned} \quad (21)$$

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 $\tilde{\mathbf{c}}_{j,k,l}(i) = [\mathbf{R}_j(i)]^H \mathbf{m}_{j,k,l}(i)$. Per each training sample (symbol), the beamformers are updated as

$$\mathbf{m}_{b,k,l}^{(n)}(i+1) = \mathbf{m}_{b,k,l}^{(n)}(i) - \alpha \mathbf{g}_{b,k,l}^{(n)}(i). \quad (22)$$

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

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, (21) 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 $[a, b]$.

V. FEEDBACK QUANTIZATION

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 $\mathbf{c}_{b,k,l}$ terms for the proposed methods. Thus, robustness to the quantization errors is crucial for any, in practise, realizable design. In addition, quantization reduces the backhaul utilization. In Section VI, we study the performance of the proposed beamformer design algorithms with q -bit quantization of the feedback information.

VI. NUMERICAL EXAMPLES

The simulations are carried in a 7-cell wrap around model ($B = 7$). The are $N_T = 4$ transmit and $N_R = 2$ receive antennas. Each cell has $K_b = 3$ cell edge users. In total, there are $K = BK_b = 21$ users. We assume full cooperation, i.e., all users are coherently served by every BS in the system. In practice, 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 cell edge signal-to-noise ratio (SNR) from the closest BS is fixed to 15dB. The pilot training sequences are random binary sequences with 10dB pilot power gain over the SNR. This reflects a worse-case scenario, where none of the pilot resources are made orthogonal. As a reference scenario, we use a computationally more complex BR design [12]. In the BR design, the beamformers are solved locally, while assuming that the cooperating BSs use fixed transmitters. Since the BSs rely only on the previous iteration information, the

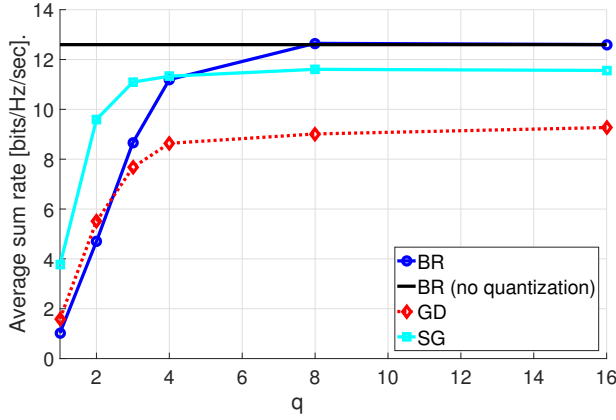


Fig. 2. Feedback signaling with q -bit quantization.

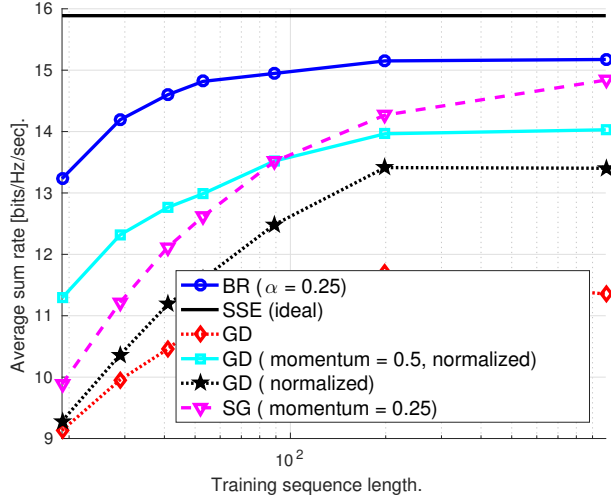


Fig. 1. Comparison of the proposed decentralized methods with varying training sequence lengths with constant channels.

beamforming is decoupled. In the figures, SSE (ideal) presents the performance upper bound with orthogonal pilots and no estimation noise. The default α for the BR design is 0.25. The GD and SG beamformers are generated using $\omega = 0.25$ momentum, $\beta = 0.005$ dual step-size and $\alpha = 0.25$ beamformer step-size.

The impact of the pilot sequence length is shown in Fig. 1. The GD methods do not achieve the orthogonal stream specific estimation (SSE) rate even at very large sequence lengths. This is due to constant step-size, which make the algorithm oscillate around a stationary point. On the other hand, the SG approaches the BR performance as the training sequence length grows. The SG approach has more fine grained beamformer updates, i.e., once per sample, which provides better convergence properties for fixed step-size. Since the pilot training vector power is fixed, there is a gap between the estimation techniques and ideal case due to the estimation noise.

Fig. 2 shows the impact of feedback quantization. In this case, the pilot training sequence length is fixed to 41. The I/Q branches of each backhaul data symbol in $\mathbf{c}_{b,k,l} \forall (b,k,l)$ are separately quantized with q -bit quantization as discussed in Section V. Symbol-by-

symbol beamformer iteration of the SG method provides significant gain at lower quantization levels. From here, we can also observe that from $q = 4$ bit quantization already achieves maximum performance. Note that the upper bound is the BR performance with the same pilot training sequences and no quantization.

VII. CONCLUSIONS

We have proposed low complexity gradient based decentralized transceiver designs for JP CoMP WSR-Max in the presence of non-orthogonal pilot resources and pilot estimation noise. GD and SG based transmit beamformer designs were proposed as low complexity alternatives to more complex BR design. The implementation complexity and performance trade-off was studied by numerical evaluation. The numerical results indicated that the SG based design is more robust to feedback quantization than the GD and BR designs with comparable performance.

REFERENCES

- [1] E. Dahlman, S. Parkvall, and J. Sködl, *4G LTE / LTE-Advanced for Mobile Broadband*. Academic Press, 2011.
- [2] 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.
- [3] 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.
- [4] A. Tölli, H. Pannanen, 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.
- [5] 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.
- [6] M. Hong, R. Sun, H. Baligh, and Z.-Q. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks," *IEEE J. Select. Areas Commun.*, vol. 31, no. 2, pp. 226–240, Feb. 2013.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [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] J. Kaleva, A. Tölli, M. J. Juntti, R. Berry, and M. L. Honig, "Joint transmission with limited backhaul connectivity," Tech. Rep., 2016. [Online]. Available: <https://arxiv.org/abs/1705.05252>