DECENTRALIZED SUM MSE MINIMIZATION FOR COORDINATED MULTI-POINT TRANSMISSION

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

[†]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.

ABSTRACT

Two decentralized minimum mean-squared error downlink beamformer designs are proposed for multiple-input single-output coherent coordinated multi-point transmission. We propose a parallel beamformer design with a fast initial rate of convergence for systems with relatively few cooperative base stations (BSs). An alternating direction method of multipliers based design is provided for more complex systems with a large number of cooperating BSs. Support for data sharing among the serving BSs is assumed over limited backhaul connectivity. Channel state information (CSI) is not shared among the cooperating transmitters, and, thus, only local CSI is available at each BS via uplink pilot signaling.

Index Terms— Cellular networks, coordinated beamforming, mean-squared error minimization, multi-user beamforming, non-linear optimization.

1. INTRODUCTION

Wireless cellular networks are becoming increasingly crowded as the cells are becoming smaller and more dense. In systems, with high cross channel gains, cooperation of simultaneously transmitting base stations (BSs) can be exploited to provide significant gains in terms of the achievable system performance. Different forms of coordinated multi-point (CoMP) transmission schemes have been proposed to leverage these gains in dense networks, where conventional non-cooperative methods fail due to highly complex interference management requirements and limited spatial resources [1].

In this paper, we provide two decentralized minimum sum mean-squared error (MSE) beamformer designs for CoMP transmission with data sharing over the backhaul. The proposed beam coordination is based on an assumption that the channel state information (CSI) cannot be shared over the backhaul and, thus, each BS has only knowledge of the locally measured CSI. The lack of global CSI requires iterative beamformer design, where limited beam coordination information is exchanged in each iteration. CoMP without complete CSI exchange has been not been widely studied in the past. Most of the research focus has been more towards variations of network multiple-input multiple-output (MIMO) type systems, where the CSI is exchanged among the cooperating BSs along with data [2]. While CSI sharing allows centralized processing, it becomes difficult in practice for highly dense systems or systems with fast fading and time correlated channel conditions, where the backhaul signaling requirements increase rapidly with the system complexity. Due to queueing and buffering, transmitted data is more tolerant to delay. Furthermore, hierarchically, the data can be considered to be readily available at the central processing units, while the CSI has to be collected from the transmitters.

MSE has been used as a performance objective for various types of systems. In terms of achievable throughput, it was shown in [3] and [4] that MSE minimization provides a lower bound for the mutual information. CoMP beamformer designs have been considered for various performance objectives and degrees of decentralization. The decentralized design of CoMP beamformers for minimizing the total power was considered for multiple-input single-output (MISO) systems using dual decomposition in [5], while in [6], this problem was considered using the alternating direction method of multipliers (ADMM) [7] to separate the coupling interference. In [8], low-complexity MIMO MSE minimization with per-antenna power constraints was proposed with per-BS Gauss-Seidel type iterative beamformer updates. Weighted MSE minimization with imperfect CSI for network MIMO systems was considered in [9]. All of the aforementioned schemes assume CSI exchange among the cooperating BSs.

We propose two minimum sum MSE algorithms. For relatively few cooperating BSs, we provide a parallel beamformer design, where BSs assume fixed beamformers from the collaborating BSs in each iteration. This method provides fast initial performance improvement. However, the convergence cannot be guaranteed and, particularly, for high complexity systems, the method can become unstable. For a more stable approach, we propose an ADMM design with similar signaling overhead and comparable rate of convergence.

2. SYSTEM MODEL

We consider a multi-cell system with B BSs each equipped with N_{T} transmit antennas. There are, in total, K single-antenna user terminals (UEs). Each UE $k = 1, \ldots, K$ is served coherently by $|\mathcal{B}_k|$ BSs, where $\mathcal{B}_k$ contains the serving set of BS indices. Similarly, the set of user indices served by BS $b = 1, \ldots, B$ is denoted by $\mathcal{C}_b = \{k | b \in \mathcal{B}_k, k = 1, \ldots, K\}$. For notational convenience, we will denote the set of all user indices as $\mathcal{K} = \{1, \ldots, K\}$.

The downlink transmission is considered to be symbol synchronous in the sense that each transmitted symbol from $\mathcal{B}_k, k = 1, \ldots, K$ is coherently combined at all user terminals. Data sharing is assumed within each serving set of BSs $\mathcal{B}_k$. Only local CSI knowledge is assumed, that is, each BS $b = 1, \ldots, B$ is only aware of the channel vectors $\mathbf{h}_{b,k} \in \mathbb{C}^{N_\mathrm{T}}\ \forall\, k = 1, \ldots, K$. Furthermore, we consider time division duplexing (TDD), which imposes strong correlation between the uplink and downlink channels.

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

$$y_k = \sum_{b \in \mathcal{B}_k} \mathbf{h}_{b,k}^\mathrm{T} \mathbf{m}_{b,k} d_k + \sum_{i=1, i \neq k}^{K} \sum_{j \in \mathcal{B}_i} \mathbf{h}_{j,k}^\mathrm{T} \mathbf{m}_{i,j} d_i + n_k, \quad (1)$$

where $\mathbf{m}_{b,k} \in \mathbb{C}^{N_\mathrm{T}}$ is the beamformer of BS b to user k and $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma_k^2 \mathbf{I})$ denotes the receiver noise. The complex data symbols $d_k, k = 1, \ldots, K$ are assumed to be independent and identically distributed (i.i.d.) with $\mathbb{E}\{|d_k|^2\} = 1$.

The received symbol at UE $k = 1, \ldots, K$ is given as $\hat{d}_k = g_k y_k$, where g_k is the gain compensating receive filter of the corresponding user. Finally, the MSE at UE k is given as

$$\begin{aligned} \epsilon_k \triangleq\ & \mathbb{E}\{|\hat{d}_k - d_k|^2\} = |\sum_{b \in \mathcal{B}_k} g_k \mathbf{h}_{b,k}^\mathrm{T} \mathbf{m}_{b,k} - 1|^2 + \\ & \sum_{i=1, i \neq k}^{K} |\sum_{b \in \mathcal{B}_i} g_k \mathbf{h}_{b,k}^\mathrm{T} \mathbf{m}_{b,i}|^2 + |g_k|^2 \sigma_k^2. \end{aligned} \quad (2)$$

3. PROBLEM FORMULATION

By regrouping the corresponding MSE expressions (2), the sum MSE minimization problem subject to total power constraints per BS can be written as

$$\begin{aligned} \min_{g_k, \mathbf{m}_{b,k}}.\ & \sum_{k=1}^{K} (-2\mathrm{Re}\{\sum_{b \in \mathcal{B}_k} g_k \mathbf{h}_{b,k}^\mathrm{T} \mathbf{m}_{b,k}\} + \\ & \sum_{i=1}^{K} |\sum_{b \in \mathcal{B}_i} g_k \mathbf{h}_{b,k}^\mathrm{T} \mathbf{m}_{b,i}|^2 + |g_k|^2 \sigma_k^2) \\ \text{s. t.}\ & \sum_{k \in \mathcal{C}_b} \|\mathbf{m}_{b,k}\|^2 \leq P_b,\ b = 1, \ldots, B, \end{aligned} \quad (3)$$

where $P_b, b = 1, \ldots, B$ are the sum power limits per BS. Problem (3), is not jointly convex in the receive filters and transmit beamformers. However, fixing either one results in a convex problem. For fixed transmit beamformers, the optimal gain normalizing receive filters can be found from the roots of the Lagrangian of (3) to be

$$g_k = \frac{\sum_{b \in \mathcal{B}_k} \mathbf{m}_{b,k}^\mathrm{H} \mathbf{h}_{b,k}^*}{\sum_{i=1}^{K} |\sum_{b \in \mathcal{B}_i} \mathbf{h}_{b,k}^\mathrm{T} \mathbf{m}_{b,i}|^2 + \sigma_k^2},\ k = 1, \ldots, K. \quad (4)$$

The receive filter design is trivially decoupled between the UEs and relies only on knowledge of the local CSI. In TDD systems, the effective local CSI can be efficiently distributed by using precoded downlink demodulation pilots.

For fixed receive filters (4), (3) does not decouple between the transmitting BSs due to coherent reception at the UEs. This is a major distinction between the coherent and non-coherent signal reception, when considering distributed solutions. Non-coherent MSE minimization is readily decoupled and is, as such, easily decentralized [10].

4. DECENTRALIZED BEAMFORMER DESIGN

In this section, we will derive two decentralized beamforming methods for (3). We focus on the beamformer design, while assuming that the receive filters $g_k, k = 1, \ldots, K$ are fixed and of the form given by (4). The alternating receive filter and transmit beamformer updates can be straightforwardly extended to MIMO systems (similarly to [10]).

4.1. Direct Parallel Design

A direct parallel design for (3) can be carried out by assuming that the beamformers from the cooperating BSs are fixed to match the received signal from the previous iteration. With this assumption, the beamformer design problem in the n^th iteration for each BS $b = 1, \ldots, B$ reduces to

$$\begin{aligned} \min_{\mathbf{m}_{b,k}}.\ & \sum_{k \in \mathcal{C}_b} (\sum_{i=1}^{K} |g_k \mathbf{h}_{b,k}^\mathrm{T} \mathbf{m}_{b,i} + c_{k,i}^n - g_k \mathbf{h}_{b,k}^\mathrm{T} \mathbf{m}_{b,i}^{n-1}|^2 - \\ & \quad 2\mathrm{Re}\{g_k \mathbf{h}_{b,k}^\mathrm{T} \mathbf{m}_{b,k}\}) \\ \text{s. t.}\ & \sum_{k \in \mathcal{C}_b} \|\mathbf{m}_{b,k}\|^2 \leq P_b,\ b = 1, \ldots, B. \end{aligned} \quad (5)$$

where

$$c_{k,i}^n = \sum_{j \in \mathcal{B}_i} g_k \mathbf{h}_{j,k}^\mathrm{T} \mathbf{m}_{j,i}^{n-1}\ \forall\, (k, i) \in \mathcal{K} \times \mathcal{K}. \quad (6)$$

It can be seen that the decoupled problem formulation requires each UE $k = 1, \ldots, K$, to provide the coherently received signals $c_{k,i}^n$ from the previous iteration for all active users $i = 1, \ldots, K$ to the BSs.

The direct parallel design achieves significant performance gains on first few iterations. However, the convergence cannot be guaranteed and the algorithm is unstable when the number of simultaneously transmitting BSs becomes large (see Sect. 5). The stability of the algorithm can be improved by imposing smoothing after each iteration to restrict the degree of change in the beamformers [11, Chap. 2.2]. With smoothing, the next iteration beamformers are given as

$$\mathbf{m}_{b,k}^{n+1} = \bar{\mathbf{m}}_{b,k} + \alpha(\mathbf{m}_{b,k}^{n} - \bar{\mathbf{m}}_{b,k}) \; \forall \, b = 1, \dots, B, k \in \mathcal{C}_b \tag{7}$$

for some $\alpha \in [0, 1]$, where $\bar{\mathbf{m}}_{b,k}$ is the solution of (5) and α determines the degree of smoothing. The selection of step size α is heuristic since at each step the BSs cannot evaluate the impact on the system performance. Convergence of best-response type algorithms have been further studied in [12], where convergence could be guaranteed for strongly convex subproblems with sufficient small step size. Note that since the total sum power constraint is convex, the smoothing step will preserve the feasibility of the power constraints [13].

If the beamformer updates (5) are performed in a sequential manner across the BSs, e.g., following a Gauss-Seidel type update process, the algorithm is guaranteed to converge as each update strictly improves the objective. However, without parallel updates, the rate of convergence is significantly reduced as shown in Sect. 5.

4.2. ADMM Design

A more stable alternative to the direct parallel design can be achieved by using the ADMM approach, which has provided an iterative procedure with efficient decomposition and good convergence properties for various types of problems [7].

First, we introduce new variables $s_{i,k,b}$ and constraints

$$s_{k,i,b} = g_k \mathbf{h}_{b,k}^{\mathrm{T}} \mathbf{m}_{b,i} \; \forall \, (k,i) \in \mathcal{K} \times \mathcal{K}, b \in \mathcal{B}_i. \tag{8}$$

The dual variables (Lagrangian multipliers) of (8) are denoted as $\lambda_{k,i,b}$. The principal idea in ADMM is to alternate the updates of variables $s_{k,i,b}$ and $\mathbf{m}_{b,i}$ along with the dual variables $\lambda_{k,i,b}$ of (8) while keeping the others fixed. Now, to separate the updates, we use Lagrangian relaxation of the constraints (8). Additionally, we impose penalty norm terms for the constraint violation, which are used to enforce the constraints and improve the rate of convergence. Finally, to reduce the signaling requirements, we introduce average received signal constraints [7, Sect. 7.3] for each UE corresponding to the reception from each collaborating set of BSs $\mathcal{B}_k, k \in \mathcal{K}$ as $\bar{s}_{k,i} = \frac{1}{|\mathcal{B}_i|} \sum_{b \in \mathcal{B}_i} s_{k,i,b} \; \forall \, (k,i) \in \mathcal{K} \times \mathcal{K}$.

Now, in n^{th} iteration, we need to solve the following optimization problem

$$\begin{aligned}
\min_{\substack{s_{k,i,b}, \bar{s}_{k,i}, \\ \mathbf{m}_{b,k}}} . \quad & \sum_{k=1}^{K} (\sum_{i=1}^{K} ||\mathcal{B}_i| \bar{s}_{k,i}|^2 - 2\mathrm{Re}\{ \sum_{b \in \mathcal{B}_k} g_k \mathbf{h}_{b,k}^{\mathrm{T}} \mathbf{m}_{b,k} \}) \\
& \sum_{k=1}^{K} \sum_{i=1}^{K} \sum_{b \in \mathcal{B}_i} \rho | g_k \mathbf{h}_{b,k}^{\mathrm{T}} \mathbf{m}_{b,i} - s_{k,i,b} + \lambda_{k,i,b}^{n} |^2 \\
\text{s. t.} \quad & \bar{s}_{k,i} = \frac{1}{|\mathcal{B}_i|} \sum_{b \in \mathcal{B}_i} s_{k,i,b} \; \forall \, (k,i) \in \mathcal{K} \times \mathcal{K}, \\
& \sum_{k \in \mathcal{C}_b} \|\mathbf{m}_{b,k}\|^2 \leq P_b, \; b = 1, \dots, B,
\end{aligned} \tag{9}$$

where $|\mathcal{B}_i|$ denotes the cardinality of set $\mathcal{B}_i$, parameter ρ is adjusted to determine the degree of enforcement for constraints (8). The dual variables $\lambda_{k,i,b}^{n}$ in (9) are scaled so that they can be incorporated into the penalty norms. For a detailed discussion on ρ balancing and scaled dual variables, see [7]. Note that $s_{k,i,b} \; \forall \, (k,i,b)$ and $\bar{s}_{k,i} \; \forall \, (k,i)$ are complex variables in (9).

Decentralized solution for (9) would still require exchanging all $s_{k,i,b} \; \forall \, (k,i,b)$ within the serving set $\mathcal{B}_i$. Also, each UE k would need to be able to separate individual effective channels $g_k \mathbf{h}_{b,k}^{\mathrm{T}} \mathbf{m}_{b,i} \; \forall \, i \in \mathcal{K}$, which is practically intractable as it would require orthogonal pilot signaling within each cooperating set of BSs $\mathcal{B}_k, k \in \mathcal{K}$.

To further simplify the problem formulation, we can eliminate the auxiliary variables $s_{k,i,b} \; \forall \, (k,i,b)$ [7]. To this end, we solve for the individual $s_{k,i,b}$ from (9), while keeping the other variables fixed. This results in

$$s_{k,i,b} = a_{k,i,b}^{n} + \bar{s}_{k,i} - \bar{\lambda}_{k,i}^{n} - \frac{1}{|\mathcal{B}_i|} \sum_{j \in \mathcal{B}_i} g_k \mathbf{h}_{j,k}^{\mathrm{T}} \mathbf{m}_{j,i}, \tag{10}$$

where $\bar{\lambda}_{k,i}^{n} = \frac{1}{|\mathcal{B}_i|} \sum_{j \in \mathcal{B}_i} \lambda_{k,i,j}^{n}$ and $a_{k,i,b}^{n} = \lambda_{k,i,b}^{n} + g_k \mathbf{h}_{b,k}^{\mathrm{T}} \mathbf{m}_{b,i}$. When we substitute each $s_{k,i,b}$ in (9) with (10), the dual variables are equal for all $b \in \mathcal{B}_i$ [7]. Thus, we can combine all dual variables for each pair $(k,i) \in \mathcal{K} \times \mathcal{K}$ as

$$\bar{\lambda}_{k,i}^{n+1} = \bar{\lambda}_{k,i}^{n} + \frac{1}{|\mathcal{B}_i|} \sum_{j \in \mathcal{B}_i} g_k \mathbf{h}_{j,k}^{\mathrm{T}} \mathbf{m}_{j,i}^{n+1} - \bar{s}_{k,i}^{n+1}. \tag{11}$$

Now, with (10) and dual update (11), (9) can be formulated as

$$\begin{aligned}
\min_{\substack{\bar{s}_{k,i}, \\ \mathbf{m}_{b,k}}} . \quad & \sum_{k=1}^{K} (\sum_{i=1}^{K} ||\mathcal{B}_i| \bar{s}_{k,i}|^2 - 2\mathrm{Re}\{ \sum_{b \in \mathcal{B}_k} g_k \mathbf{h}_{b,k}^{\mathrm{T}} \mathbf{m}_{b,k} \}) + \\
& \sum_{i=1}^{K} \sum_{k=1}^{K} \rho \sum_{j \in \mathcal{B}_i} | g_k \mathbf{h}_{j,k}^{\mathrm{T}} \mathbf{m}_{j,i} - g_k \mathbf{h}_{j,k}^{\mathrm{T}} \mathbf{m}_{j,i}^{n} + q_{k,i}^{n} |^2 \\
\text{s. t.} \quad & \sum_{k \in \mathcal{C}_b} \|\mathbf{m}_{b,k}\|^2 \leq P_b, \; b = 1, \dots, B,
\end{aligned} \tag{12}$$

where $q_{k,i}^{n} = \frac{1}{|\mathcal{B}_i|} \sum_{j \in \mathcal{B}_i} g_k \mathbf{h}_{j,k}^{\mathrm{T}} \mathbf{m}_{j,i}^{n} - \bar{s}_{k,i}^{n} + \bar{\lambda}_{k,i}^{n}$. Finally, we can solve $\bar{s}_{k,i} \; \forall \, (k,i) \in \mathcal{K} \times \mathcal{K}$ from (12) for fixed beamformers $\mathbf{m}_{b,k}^{n+1} \; \forall \, (b,k)$ and given dual variables $\bar{\lambda}_{k,i}^{n}$ as

$$\bar{s}_{k,i}^{n+1} = \frac{\rho}{1+\rho} \left(\frac{1}{|\mathcal{B}_i|} \sum_{b \in \mathcal{B}_i} g_k \mathbf{h}_{b,k}^{\mathrm{T}} \mathbf{m}_{b,i}^{n+1} + \bar{\lambda}_{k,i}^{n} \right). \tag{13}$$

The decentralized solution is obtained by observing that, both, the update (13) and dual update (11) can be managed locally at each BS b, if the averaged coherently received signals $\frac{1}{|\mathcal{B}_i|} \sum_{j \in \mathcal{B}_i} g_k \mathbf{h}_{j,k}^{\mathrm{T}} \mathbf{m}_{j,i}^{n}$ are known from each UE $k \in \mathcal{C}_b$. Assuming that the UEs communicate the averaged coherently received signals from the previous iteration (as in Sect. 4.1), each BS can locally keep track of the primal and dual variables. This leads to the ability to solve for the local beamformers at each BS from (12) for fixed $\bar{s}_{k,i} \; \forall \, (k,i) \in \mathcal{K} \times \mathcal{K}$.

Each UE $k \in \mathcal{K}$ has to communicate K complex symbols to all BSs. This results in total K^2 complex numbers per iteration. However, in practice, only the dominant interfering signals should be considered, in order to reduce the signaling overhead. The trade-off between the required amount of signaling and system performance is a topic for future work. The outline of the ADMM algorithm is given in Algorithm 1.

Due to lack of space, detailed convergence analysis is neglected. Proof of convergence for the ADMM method for convex problems can be found from [7]. These results can be applied to the beamformer convergence for fixed receive filters. However, as the sum MSE minimization problem is, in general, non-convex (see [14] for convexity conditions), the receive filter update (4) requires extended analysis. Rough convergence conditions can be derived by noting that the receive filter update strictly improves the objective value. Now, such conditions for ρ can derived that, after each full iteration, Algorithm 1 moves towards a stationary point of (3).

Algorithm 1 ADMM algorithm for MSE minimization

UE: Initialize receive filters $g_k = 1 \; \forall \, k \in \mathcal{K}$.
BS: Initialize the variables $\bar{s}^n_{k,i} = 0$ and dual variables $\bar{\lambda}^n_{k,i} = 0$ for all $(k, i) \in \mathcal{K} \times \mathcal{K}$.
repeat
    UE: Inform each BS of the downlink effective channels by uplink pilots.
    BS: Update the local beamformers from (12).
    UE: Obtain local effective CSI from the downlink demodulation pilots.
    UE: Report back the average combined signals $\frac{1}{|\mathcal{B}_i|} \sum_{b \in \mathcal{B}_i} g_k \mathbf{h}^{\mathrm{T}}_{b,k} \mathbf{m}_{b,i} \; \forall \, (k, i) \in \mathcal{K} \times \mathcal{K}$ for the serving set of BSs (K complex symbols per UE).
    BS: Locally update the variables $\bar{s}_{k,i}$ and dual variables $\bar{\lambda}_{k,i}$ from (13) and (11) according to the average signal feedback from UEs.
    UE: Update the receive filters $g_k \; \forall \, k \in \mathcal{K}$ from (4).
until Desired level of convergence has been reached.

5. NUMERICAL EXAMPLES

Simulations are carried out with $B = 3$ cooperating BSs each equipped with $N_{\mathrm{T}} = 2$ transmit antennas. There are $K = 8$ single-antenna users who are served simultaneously by all BSs, that is, $\mathcal{B}_1 = \mathcal{B}_2 = \mathcal{B}_3 = \mathcal{K}$. This is a spatially overloaded scenario with more initially active beamformers than degrees-of-freedom. The signal-to-noise ratio (SNR) for all users is 15dB. Channel realizations are drawn from circularly symmetric Gaussian distribution with unit variance. We consider cell edge scenario, where all users have are located equally far from all serving BSs. The results are averaged over 100 independent channel realizations.

The non-coherent beamformer optimization provides a lower bound for the achievable MSE, in which, each BS optimizes the beamformers according to the MSE by treating all signals from other BSs as noise. Joint optimization gives a reference point, where beamformers are optimized with global CSI knowledge (still alternating the receive filter and transmit beamformers).

In Fig. 1, the convergence in terms of the achieved average sum MSE is illustrated on logarithmic scale. The direct parallel design provides the best performance for the first few iterations. However, the ADMM method provides improved convergence properties and intersects with the parallel methods after the first few iterations. The parallel design without smoothing fails to converge in some cases, which results in a gap to the achievable performance. From the non-coherent beamforming results, it can be seen that coherent beamforming can achieve significantly improved system performance.

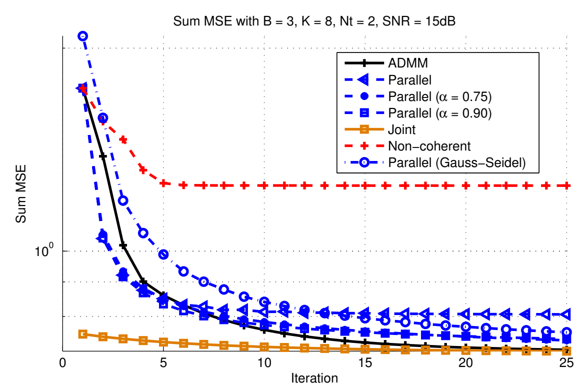

Fig. 1. Sum MSE convergence at SNR = 15dB, $B = 3$, $N_{\mathrm{T}} = 2$, $K = 8$ and $\rho = 2$.

6. CONCLUSIONS

We proposed decentralized beamforming methods with limited backhaul signaling for downlink CoMP MISO transmission. The system performance was optimized in terms of sum MSE. Low complexity direct parallel updates were observed to provide improved initial rate of convergence. For improved stability, a smoothing step was proposed after each beamformer update. To overcome the stability issues, the ADMM technique was proposed with identical signaling requirements and improved convergence properties for systems with more than two cooperating BSs. Trade-off between the signaling overhead and system performance is left for future study.

Acknowledgment

This work has been supported by the Finnish Funding Agency for Technology and Innovation (Tekes), Nokia Siemens Networks, Renesas Mobile Europe, Xilinx, Academy of Finland and Infotech Oulu Doctoral Program.

7. REFERENCES

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

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

[3] Petre Stoica, Yi Jiang, and Jian Li, "On MIMO channel capacity: an intuitive discussion," *IEEE Signal Processing Mag.*, vol. 22, no. 3, pp. 83–84, 2005.

[4] A. Vosoughi and A. Scaglione, "On the effect of receiver estimation error upon channel mutual information," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 459–472, 2006.

[5] 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.

[6] Chao Shen, Tsung-Hui Chang, Kun-Yu Wang, Zhengding Qiu, and Chong-Yung Chi, "Distributed robust multicell coordinated beamforming with imperfect CSI: An ADMM approach," *IEEE Trans. Signal Processing*, vol. 60, no. 6, pp. 2988–3003, June 2012.

[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[8] Tae Min Kim, Fan Sun, and A.J. 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.

[9] Haewook Park, Seok-Hwan Park, Han-Bae Kong, and Inkyu Lee, "Weighted Sum MSE Minimization under Per-BS Power Constraint for Network MIMO Systems," *IEEE Commun. Lett.*, vol. 16, no. 3, pp. 360–363, 2012.

[10] 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, Sept. 2011.

[11] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, 2nd edition, 1999.

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

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

[14] Z. Q. Luo, T. N. Davidson, G. Giannakis, and K. M. Wong, "Transceiver optimization for block-based multiple access through ISI channels," *IEEE Trans. Signal Processing*, vol. 52, no. 4, pp. 1037–1052, Apr. 2004.