

Robust Linear Precoder for Coordinated Multipoint Joint Transmission under Limited Backhaul with Imperfect CSI

Rikke Apelfröjd and Mikael Sternad

Signals and Systems, Uppsala University

Box 534, 751 21 Uppsala, Sweden, {rikke.apelfrojdmikael.sternad}@signal.uu.se

Abstract—Coordinated Multipoint (CoMP) transmission provides high theoretic gains in spectral efficiency, in particular with coherent linear Joint Transmission (JT) to multiple users. However, this requires high backhaul capacity. If the backhaul requirement cannot be met by the system, then CoMP gains decrease as the linear precoder matrix must be adjusted to include zeros. To minimize the loss of CoMP gains, all elements in the precoder should be adjusted as zeros are added to the precoder. We here propose a low complexity method for adjusting a precoder matrix when some elements are required to be zero, with respect to a robust MSE criterion. This is done by introducing penalties on specific precoder matrix elements. This generalized MSE criterion can then be used as a low complexity tool for optimizing e.g. with respect to sum-rate. Results show that this does indeed provide a better solution than if zeros are added separately. It is especially beneficial for cell edge users, i.e. for the same users that can gain the most from JT CoMP.

I. INTRODUCTION AND ASSUMPTIONS

Shadowed areas and interference at cell borders pose challenges for future wireless broadband systems. A potentially powerful remedy would be Coordinated Multipoint (CoMP) transmission. It can overcome interference limitations in cellular radio networks and also provide coverage gains. First steps towards support for CoMP have recently been added to the 3GPP LTE standard in Release 11 [1].

Linear coherent Joint Transmission (JT), also referred to as network Multiple-Input Multiple-Output (MIMO), can provide high potential gains in spectral efficiency at full load, see e.g. [2], [3], by converting harmful interference power into useful signal power, in particular for users at cell edges [4].

A drawback with coherent JT is that it requires high backhaul capacity, as Channel State Information (CSI) reports, precoder information and payload data must be shared amongst all base stations, see Figure 1. We here assume that users (terminals) in a Frequency Division Duplex (FDD) system, measure the Orthogonal Frequency-Division Multiplexing (OFDM) downlink channels from all base stations within the cooperation cluster, based on pilot symbols in the downlink transmission. All users feed back this CSI and estimates of its uncertainty to their strongest base station. The base stations then send this to a Control Unit (CU) which calculates the precoder.

The backhaul network may place structural constraints, capacity constraints and/or delay constraints on the control traffic and on the distribution of payload data. Such constraints

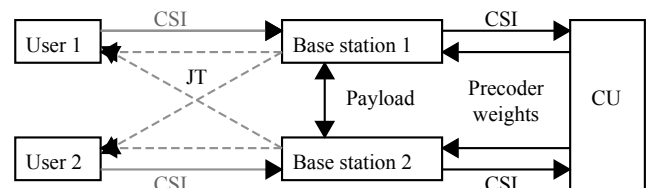


Figure 1. Feedback traffic (gray arrows), backhaul traffic (black arrows) and JT of payload (dashed arrows) for a simple downlink FDD setup with two base stations and two users.

need to be taken into accounts in the precoder design. One option is to enforce zeros in the precoding matrix. For each zero in the precoding matrix, one less precoding weight and one less data symbol needs to be shared over backhaul links.

The problem of limited feedback in combination with linear precoding for coherent JT CoMP has been previously studied in e.g. [5] where zeros are enforced in the precoder without adjusting the non-zero elements. In [6], [7], multi-dimensional searches are instead used to adjust all elements of the precoder under backhaul constraints. However, multi-dimensional search generates high computational complexity. Moreover, the contributions of [5], [6], [7] do not consider CSI uncertainty in the precoder design.

The robust linear precoding solution of [8], [9] is here extended to also handle backhaul limitations without losing optimality with respect to a robust MSE criterion, which averages over the CSI uncertainty. The extension is obtained by introducing penalties on specific elements of the precoder matrix. This generalized MSE criterion can then be used as a low complexity tool for optimization with respect to other criteria, e.g. the sum-rate.

Near accurate CSI at the transmitter side is important for coherent JT CoMP gains [10]. System delays, such as feedback and backhaul delay, will translate into inaccuracy in the CSI. In evaluations, we will here assume that a channel predictor, e.g. a Kalman predictor, is used to ensure that the outdated of CSI is partly counteracted [11].

Notations: Below, $(\cdot)^T$, $(\cdot)^*$, $\text{Re}(\cdot)$, and $\text{tr}(\cdot)$ denote the transpose, conjugate transpose, the real part and the trace of a matrix respectively, \odot denotes element wise multiplication and $\|\cdot\|$ denotes the euclidean norm of a vector. Expectation over channel uncertainty is denoted $\bar{E}[\cdot]$ while $E[\cdot]$ is the

expectation over message and noise statistics. A_{ij} denotes the submatrix of A associated with user i and base station j .

II. SYSTEM MODEL

We assume OFDM downlinks in a CoMP cluster with N base stations, where M users have been scheduled to be served within a particular time-frequency resource block. Let $n_{r,i}$ and $n_{t,j}$ denote the number of receive antennas at user i and transmit antennas at base station j respectively. The total numbers of receive and transmit antennas are denoted \mathcal{M} and \mathcal{N} respectively. For each subcarrier and OFDM symbol within the resource block, the received signal vector $y \in \mathbb{C}^{\mathcal{M} \times 1}$ after OFDM receiver processing is

$$y = Hu + v. \quad (1)$$

Here $H \in \mathbb{C}^{\mathcal{M} \times \mathcal{N}}$ is the complex channel matrix and $v \in \mathbb{C}^{\mathcal{M} \times 1}$ is the sum of noise and intercluster interference. It is assumed to be zero mean and will be referred to as noise. Let $R \in \mathbb{C}^{\mathcal{N} \times \mathcal{M}}$ and $s \in \mathbb{C}^{\mathcal{M} \times 1}$ be the linear precoding matrix and the message vector, respectively. Then, the transmit signal vector $u \in \mathbb{C}^{\mathcal{N} \times 1}$ is given by

$$u = \frac{1}{c}Rs. \quad (2)$$

Here, c is a real-valued scaling constant used to fulfill a per base station power constraint. We assume that all messages are i.i.d. zero mean with unit variance.

The base station with the strongest channel gain to a users, on average over all multipaths and all resource blocks, will be denoted the master base station for the specific user.

A. Representation of channel uncertainty

The complex channel matrix H is represented by the sum of the known channel estimate \hat{H} and some uncertainty ΔH , including prediction errors and quantization errors,

$$H = \hat{H} + \Delta H. \quad (3)$$

We will assume that the channel uncertainty is zero mean with known covariances $\bar{E}[\Delta h_{m,n} \Delta h_{l,k}^*]$ where $\Delta h_{m,n}$ is the uncertainty in the channel gain between transmit antenna $n \in [1, \mathcal{N}]$ and receive antenna $m \in [1, \mathcal{M}]$. These second order moments of the channel uncertainty are represented by a covariance matrix $C_H \in \mathbb{C}^{\mathcal{M} \cdot \mathcal{N} \times \mathcal{M} \cdot \mathcal{N}}$ where each $\mathcal{M} \times \mathcal{M}$ block element (n, k) is given by

$$(C_H)_{nk} = \bar{E}[\Delta H_n \Delta H_k^*]. \quad (4)$$

Here, ΔH_n is the column of ΔH associated with transmit antenna n . In contexts where \hat{H} is seen as (or generated as) a random variable, such as in the simulation study of Section IV, we assume that $\bar{E}[\hat{H} \Delta H^*] = 0$. This assumption is valid e.g. when the channels are predicted through a Kalman filter, see [13], and then quantized. In general, and when applying channel estimation and precoder design to real-world data, we use a Bayesian approach. We then regard the estimate as given and known, while the estimation error (ΔH in (3)) is unknown so the true channel (H in (3)) is partly unknown.

Under backhaul constraints, feedback constraints or both, all base stations may not serve all users. Let $U \in \mathbb{C}^{\mathcal{M} \times \mathcal{N}}$ be a matrix of blocks U_{ij} , which are all ones if base station j will serve user i and all zeros if not. Feedback and backhaul demands can be lowered by providing the CU only with estimates of the channels that will be used for payload transmission. The CSI known at the CU is then restricted to

$$\hat{H}_{CU} = \hat{H} \odot U. \quad (5)$$

The channel matrix in (3) can now be expressed as

$$H = \hat{H}_{CU} + \underbrace{\hat{H}_{\perp CU} + \Delta H}_{\Delta H_{CU}}, \quad (6)$$

where $\hat{H}_{\perp CU} = \hat{H} - \hat{H}_{CU}$ represents all channel component estimates that are not made available at the CU. The corresponding covariance matrix of ΔH_{CU} in (6) has $\mathcal{M} \times \mathcal{M}$ blocks

$$(C_{H,CU})_{nk} = (C_H)_{nk} + \bar{E}\left[\left(\hat{H}_{\perp CU}\right)_n \left(\hat{H}_{\perp CU}\right)_k^*\right]. \quad (7)$$

The covariance matrices (4) and (7) only need to be fed back on a slow time scale, which relates to the shadow fading, and therefore only cause small overhead in feedback and backhaul. We therefore assume that the CU has full access to them.

III. PRECODER DESIGN

Often the desired optimization metric poses a non-convex problem which requires a $2 \cdot \mathcal{M} \cdot \mathcal{N}$ dimensional search over all real and imaginary parts of the elements of the precoding matrix R . To simplify the design, we here utilize an auxiliary robust weighted scalar MSE criterion

$$J_0 = \bar{E}\left[E\|V\varepsilon\|^2 + E\|Su\|^2\right], \quad (8)$$

where

$$\varepsilon = \frac{1}{c}(HR - D)s \quad (9)$$

is the difference between the received signals and a desired set of target signals $(1/c)Ds$ in a noise free scenario. V is a diagonal positive definite penalty matrix and S is positive semidefinite penalty matrix. These penalty matrices can and will be used as design parameters to optimize an arbitrary criterion, see Section III-A. The desired target matrix, D , can be designed in various ways, see [8].

Theorem 1. For a transmission system (1)-(2), with a channel defined as (6), assume that c in (2) is given, that \hat{H}_{CU} in (6) is known, that $\bar{E}[\Delta H_{CU}] = 0$, that $S \in \mathbb{R}^{\mathcal{N} \times \mathcal{N}}$ in (8) has full rank and that s in (2) is white with zero mean. Define

$$\beta^* \beta = \hat{H}_{CU}^* V^* V \hat{H}_{CU} + S^* S + \bar{E}[\Delta H_{CU}^* V^* V \Delta H_{CU}]. \quad (10)$$

Then the precoding matrix R minimizing J_0 by (8) exists and is given uniquely by

$$R_0 = (\beta^* \beta)^{-1} \hat{H}_{CU}^* V^* V D. \quad (11)$$

For a proof see [8].

As V in (8) is diagonal, the elements (k, n) of $\bar{E}[\Delta H_{CU}^* V^* V \Delta H_{CU}]$ in (10) can be calculated from (7) by

$$\left(\bar{E}[\Delta H_{CU}^* V^* V \Delta H_{CU}]\right)_{kn} = \text{tr}(V^* V (C_{H,CU})_{nk}). \quad (12)$$

Note that the special case when the channel matrix H in (1) is invertible and perfectly known, i.e. $\Delta H_{CU} = 0$ in (6), the target matrix $D = I$ in (9) and the penalty matrices in (8) are set to $V = I$ and $S = \epsilon I$ where $\epsilon \rightarrow 0$ then the precoder in (11) coincides with the zero forcing solution (\hat{H}^{-1}). In [8] the proposed precoder was compared with the zero forcing solution where the power balancing was set through $\hat{H}^{-1}D$ in the case of imperfect CSI. The proposed precoder then performed similar to zero forcing for "good" user groups and significantly outperformed zero forcing for spatially difficult user groups.

A. Optimization with respect to an arbitrary criterion

The robust solution of Theorem 1 can be used as a tool for adjusting the precoder matrix R with respect to a general criterion

$$f \left\{ \bar{E} [E [P_{S,i}]], \bar{E} [E [P_{I,i}]], \bar{E} [E [P_{N,i}]], i = 1, \dots, \mathcal{M} \right\}. \quad (13)$$

Here, $P_{S,i}$, $P_{I,i}$ and $P_{N,i}$, are the powers of the signals, the interference, and the noise at receiver antenna i .

Diagonal penalty matrices V and S in (8) provide significant flexibility. Optimization of their elements with respect to (13) provides a flexible tool for adjusting the precoder matrix by a low-dimensional numerical search. The elements of V mainly affect the weighting (fairness) between users, while the elements of S affect the power balance between transmit antennas.

One particular case is when (13) is set to approximate an unweighted sum-rate criterion. Then the use of a fixed $V = I$ is appropriate. The use of a small regularization term $S = \epsilon I$, with ϵ being a very small scalar, would then approximately minimize the intracluster interference, but not maximize the sum-rate. This is because the noise in (1) is not taken into account in (8) and its impact might be enhanced by the scaling by c to meet the power constraint through (2). The performance with respect to (13) is then for most cases improved significantly by iteratively adjusting a few real valued diagonal elements of the transmit power penalty matrix S , to rebalance the received powers, interference and noise, see [9]. This procedure is outlined in Appendix A.

B. Constraining the structure of the precoder

The precoder design in (11) can handle limited CSI due to feedback and backhaul constraints through (7). However, it is so far designed without any constraints on its elements.

Assume that our objective is to minimize (13) under the constraint that the submatrix R_{ji} should be an all zero matrix if base station j is not going to serve user i . One option is then to calculate the precoding matrix according to (11) and to simply enforce zeros in R through

$$R = R_0 \odot U^T. \quad (14)$$

However, this solution destroys optimality with respect to (8). If instead, all elements of R are readjusted, then optimality with respect to (8) may be preserved. In order to do this we

extend the auxiliary criterion to include penalty terms on R_{ji} . We introduce

$$J = \bar{E} \left[E \|V\varepsilon\|^2 + E \|Su\|^2 + E \sum_{k=1}^{\mathcal{N}} \sum_{m=1}^{\mathcal{M}} \frac{1}{c^2} |w_{km} r_{km} s_m|^2 \right], \quad (15)$$

where V , S , and ε are as in (8) and where r_{km} is the gain in R from message m to transmit antenna k , s_m is the message intended for receiver m and w_{km} is a real-valued weight that is set to zero if data for receiver m is to be transmitted from transmitter k . Otherwise, w_{km} is set to a large number, forcing r_{km} towards zero.

Theorem 2. For a transmission system (1)-(2) with a channel defined as (6), assume that c in (2) is given, that \hat{H}_{CU} in (6) is known, that $\bar{E} [\Delta H_{CU}] = 0$, that $S \in \mathbb{R}^{\mathcal{N} \times \mathcal{N}}$ has full rank and that s in (2) is white with zero mean. Then the precoding matrix R minimizing J by (15) exists and is obtained by

$$R = R_0 - (\beta^* \beta)^{-1} (W \odot W \odot R), \quad (16)$$

where $W = \{w_{km}\}_{k=1, \dots, \mathcal{N}, m=1, \dots, \mathcal{M}}$, R_0 is given by (11) and $\beta^* \beta$ is given by (10).

For proof, see Appendix B.

After solving (16) with respect to R , which for a given R_0 and $\beta^* \beta$ represents a linear system of equations, the scale factor c in (2) is adjusted to fulfill the transmit power constraint. This scales the criterion (15) by $1/c^2$, but does not affect the minimizing precoder R .

IV. SIMULATIONS

In this section we investigate the effect of calculating the precoder through (16) instead of through (14).

Limiting the number of cooperating cells per cluster might potentially lower the CoMP gains. Investigations in [14], [15] show that a cluster size above 7-9 cells will not provide large additional CoMP gains. With this in mind, we here focus our investigations on a $N = 9$ cell cluster with base stations located at three different sites, with a distance of 500 m, as shown in Figure 2. We assume two transmit antennas per base station and single-antenna users, i.e. $n_{t,j} = 2$, $n_{r,i} = 1$.

In [8] it was shown that CoMP gains increased if the users that are scheduled on the same resource have different master base stations. Users are therefore here randomly distributed with one user per cell ($M = 9$)¹. Assuming different users to be in different cells also increases the chance that the estimated channel matrix (5) has full rank. The precoders are evaluated over a total of 10 000 user drops, each with 9 users.

The channels are assumed to be i.i.d. zero mean complex Gaussian with variance $\sigma_{h,i,j}^2$, using the frequently used path loss PL^{dB} and antenna power gain G^{dB} model [16]

$$PL^{dB} = 128.1 + 37.6 \log_{10}(d),$$

$$G^{dB} = 14 - \min \left\{ 12 \left(\frac{\phi}{70} \right)^2, 20 \right\}.$$

¹Using multi-user MIMO, up to two users per cell can be accommodated. The total sum rate would be maximized by serving an average number of users per cell between 1 and 2. Here, we focus on the effects of restricted backhaul, and restrict the number of users per cell to 1, to simplify the interpretation of the results.

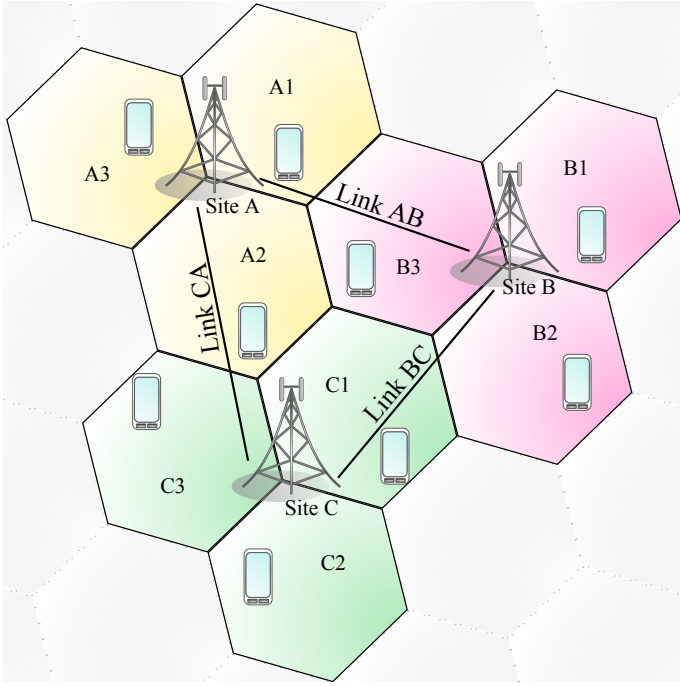


Figure 2. Set-up of the considered CoMP cluster. Nine base stations, each with 2 antennas, are located at three sites. The corresponding cells are A1-3, B1-3, C1-3. Base stations at different sites exchange information over the backhaul links (AB, BC or CA). Nine single antenna users are served jointly on each resource. We assume that a scheduler ensures that the users all have different master base stations, i.e. are located in different cells.

Here d is the distance in km between the user and base station and $\phi \in [-180^\circ, 180^\circ]$ is the horizontal (azimuth) angle.

We here include no shadow fading. This makes it easier to study the characteristics of users who benefit from different solutions by displaying their locations. The channel uncertainties are modelled as i.i.d. zero mean complex Gaussian with variance $\sigma_{h,i,j}^2/100$, i.e. 20 dB below average channel powers. This could be achieved with low prediction errors (e.g. with short system delays and low user mobility), see [8], [13], and a quantization of 4-5 bits per complex part and channel component. The value is set low to ensure that there will be useful CoMP gains compared with single cell transmission. The covariance matrix (4) will then be diagonal

$$C_H = \text{diag} \left\{ \text{diag} \left\{ \sigma_{h,i,k}^2/100 \right\}_{i=1,\dots,\mathcal{M}} \right\}_{k=1,\dots,\mathcal{N}}.$$

To ensure that the uncertainty is uncorrelated with the channel estimate we model and generate the quantized channel estimate as i.i.d. zero mean complex Gaussian with variance $99\sigma_{h,i,j}^2/100$ which is uncorrelated to ΔH , so $\bar{E}(\hat{H}\Delta H) = 0$. We then calculate the corresponding channel H through (3). The noise power in (1) is set such that the average Signal to Noise Ratio (SNR) between a user and its master base station is 20 dB. This level of intercluster interference could be achieved with e.g. the intercluster interference limiting scheme of [15].

The target matrix D is set diagonal with attainable gains

$$D = \text{diag} \left\{ \max_{j,n} \left| \left(\hat{H}_{i,j} \right)_n \right| \right\}_i, \quad (17)$$

where $\left(\hat{H}_{i,j} \right)_n$ is the estimated complex channel gain from antenna n at base station j to user i .

The optimization criterion (13) is selected as an estimate of the sum-rate per unit bandwidth based on \hat{H}_{CU}

$$\sum_{i=1}^{\mathcal{M}} \log_2 \left(1 + \frac{\left(\hat{H}_{CU} R \right)_{ii}^2}{\sigma_v^2 + \sum_{m \neq i}^{\mathcal{M}} \left(\hat{H}_{CU} R \right)_{im}^2} \right). \quad (18)$$

Here $(\cdot)_{im}$ denotes element i, m of the matrix.

The precoders (16), (14) are tested for three scenarios:

- 1) As a baseline we assume that the amount of information that is exchanged over backhaul is equally limited between all base stations. Each user is here allocated its K strongest base stations to be served by. In addition, we assume that the feedback capacity is also limited, proportionally to the backhaul, such that each user only feeds back the CSI of its K strongest base stations.
- 2) In scenario 2 we assume that the base stations located at the same site, e.g. base stations A1, A2 and A3 in Figure 2, have no backhaul restrictions. In this scenario, each user is then always served by its master base station and the two other base stations at the same site. In addition, it is served by the $K-3$ strongest base stations from other sites. As in scenario 1, we assume that the feedback capacity is also limited, proportionally to the backhaul, so only the used channels are fed back. This scenario is more realistic than scenario 1, as the backhaul constraints will be mainly limited to the intersite links.
- 3) Generally, the network structure may not be symmetric. For example link AB in Figure 2 could be a fiberoptic cable while link BC might not even be a direct link but require several routing hops through other links in the network. Then the base stations at sites A and B can share a lot of data, while the base stations at sites B and C can only share very little data. To represent this, we, in scenario 3 assume that specific intersite links are closed for sharing payload data. However, in this particular scenario we assume full feedback. As the CSI sharing over backhaul is often low compared to the sharing of payload data, we assume that the CSI can be shared over backhaul network such that the CU have access to the full CSI \hat{H} . Each user will be served by all base stations at its master base station site and from all base stations at a site that its master base station site have usable backhaul connections to. For this scenario we study three cases. First, the links AB and CA in Figure 2 are fully open for sharing of payload data while link BC is closed. This could e.g. occur if sites B and C only can communicate via site A and that the routing via site A is very slow. Second, only the link AB is open for sharing of payload data. Third, all intersite links are closed for sharing of payload data. The third case coincides with scenario 2 when $K = 3$, with the exception that in scenario 2 the CU only has access to the limited CSI of (5) and (7). In the two first cases in scenario 3, different users are served by different numbers of base stations.

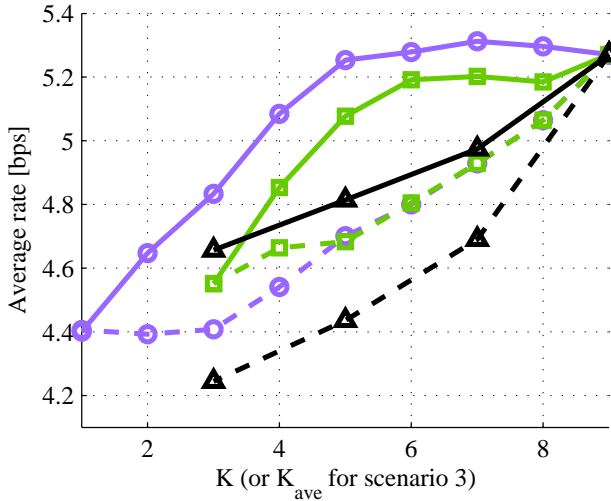


Figure 3. The average user rate per resource when the precoder is calculated through (16) (solid) and through (14) (dashed) for scenario 1 (circles), 2 (squares) and 3 (triangles).

The average number of serving base stations per user is denoted K_{ave} .

A. Results

The results of the simulations are shown in Figure 3. For all scenarios, the average rate per user increases when all precoder elements are adjusted through (16). The largest increase is observed for Scenario 1 and for $K = 4$, with a gain of 12%.

Some users experience a very large gain when the precoder in (16) is used. In Figure 4 the positions of the users with a gain $\geq 50\%$ are plotted for scenario 1 with $K = 4$ base stations serving each user. These users are more densely located at the cell borders. In other words, the users that need CoMP most are those that loose most performance if zeros are forced into the precoder as in (14).

For scenario 1 with $K = 1$ (single cell transmission), and for scenario 2 with $K = 3$ (intrasite CoMP), the channel matrix in (5) becomes block diagonal. In such cases, the corresponding precoder R_0 in (11) will automatically have zeros in the correct positions and the two precoding solutions (14) and (16) become the same. Therefore, if the zeros are not handled in the precoder design, a user grouping scheme that selects users to ensure that the channel matrix (5) is block diagonal is necessary. This is counterproductive as it will in turn increase demands on backhaul. With the proposed precoder (16) a local user grouping scheme, see e.g. [8], that places no additional demand on backhaul can be used.

Comparing scenarios 1 and 3, we see that there is a loss from having structural backhaul constraints, where users cannot always be served by their strongest base stations. However, structural constraints will occur in real systems. Therefore it is important that the precoder design can handle them.

V. CONCLUSIONS

We have here suggested a low complexity approach to include backhaul constraints in the design of the precoder,

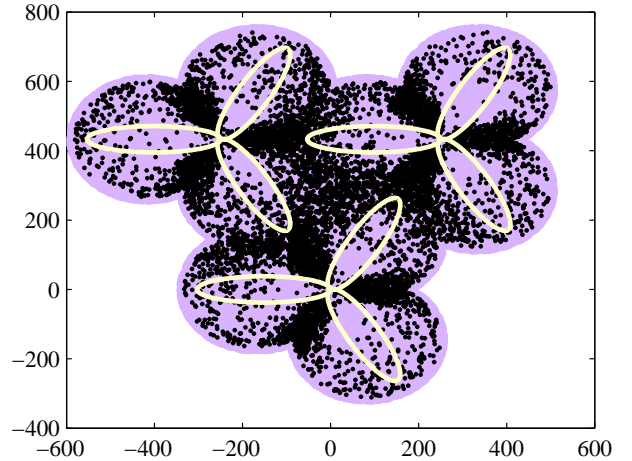


Figure 4. The users that benefit by more than 50% through the use of (16), in scenario 1 with $K = 4$, are shown as black dots. These make up 9.6% of the total number of users. The base station antenna directions are provided by the pale yellow ellipses and all other users are marked as purple circles.

as opposed to forcing zeros into the precoder as a final step of the design. Results show that, given a sum-rate criterion, this approach increases the average user rate. Gains are highest for cell edge users. This is of importance, as these are the users that need JT CoMP most.

Open issues: A way to prioritize users that need JT CoMP most could be to base the decision of how many base stations that would serve a certain user on a relative SNR threshold value. This would allow users that have equally strong channels to many base stations to be served by more base stations than users with one channel much stronger than the others.

Another aspect is that with limited backhaul, we might need to pre-select only a subset of the users as candidates for using CoMP. The question then arises on what grounds such a pre-selection can be made. A promising approach here is to base the pre-selection on the expected utility, from an application point of view, of the increased user data rate that can be obtained by allowing the user to use CoMP transmission. This research direction is explored in [17].

APPENDIX A

ITERATIVE ADJUSTMENT OF THE PENALTY MATRIX S

The criterion (18) can be optimized by adjusting the transmit powers with a step-by-step, greedy algorithm. First, calculate the optimal precoder from (11) with $V = I$ and $S = \epsilon I$. Here ϵ is a small real valued number ensuring that S is positive definite. The resulting precoder, which minimizes the intracluster interference but might not be optimal with respect to (13), is then used as the initial value for a sequence of iterative, one dimensional, searches where we sequentially adjust the penalties on the transmit powers used by each base station.

Now, calculate the transmit powers at the different antennas, with this constellation, using (2) under the per base station power constraint and set

$$S_{p,1} = \epsilon I + \text{diag} \{ \rho_1 \cdot \mathbf{1}_{j_{max}} \} \quad (19)$$

Here $\mathbf{1}_{j_{max}}$ denotes a vector with ones if the corresponding antennas are located at the base station j_{max} with highest transmit power and zeros otherwise. For example assume a system with $N = 3$, $n_{t,j} = 2$, $M = 2$, $n_{r,i} = 1$ and

$$\frac{1}{c}R = \begin{bmatrix} 1 & 2 & 3 & 1 & 1 & 0 \\ 0 & 1 & 4 & 1 & 1 & 1 \end{bmatrix}^T.$$

Then the transmit powers at base station 1, 2 and 3 are 6, 27 and 3 respectively, so $j_{max} = 2$ and $\mathbf{1}_{j_{max}} = [0 \ 0 \ 1 \ 1 \ 0 \ 0]$. The scalar parameter ρ_1 is iteratively optimized with respect to (13) over an interval $]0, \rho_{1,\max}[$ where $\rho_{1,\max}$ is the smallest value that will cause j_{max} to change. The procedure is repeated for all remaining base station in order of decreasing transmit power. Each parameter ρ_j is optimized such that the order of the base stations transmit power does not change. The final penalty matrix is

$$S_\rho = \epsilon I + \sum_{j=1}^N \text{diag} \{ \rho_j \cdot \mathbf{1}_j \}. \quad (20)$$

The solution will be suboptimal but, in a comparative study in [12] we showed that with this step by step approach, the precoder of (11) performed close to a near optimal linear precoder found through a high dimensional search of all the complex elements of R with respect to the sum-rate criterion.

APPENDIX B PROOF OF THEOREM 2

Since S in (8), (15) is assumed positive definite, $(\beta^* \beta)^{-1}$ and R_0 by (10), (11) exists. Assume that the precoder R defined by (16) does not minimize (15). Since the symbol vector s in (2) is assumed white, any alternative potentially superior linear precoder can then be expressed as a linear function of s at time t only. Then there exists a matrix T s.t. $u = \frac{1}{c}(R + T)s$ will decrease (15). Then (15) can be expressed as $J = J_1 + 2\text{Re}(J_2) + J_3$ where

$$\begin{aligned} J_1 &= \bar{E} \left[E \left[\text{tr} (V \varepsilon_0 \varepsilon_0^* V^* + S u_0 u_0^* S^*) + \sum_{i,j} \frac{w_{ij}^2}{c^2} |r_{ij}|^2 |s_j|^2 \right] \right], \\ J_2 &= \bar{E} \left[E \left[\text{tr} (V \varepsilon_0 \varepsilon_\delta^* V^* + S u_0 u_\delta^* S^*) + \sum_{i,j} \frac{w_{ij}^2}{c^2} r_{ij} \tau_{ij}^* |s_j|^2 \right] \right], \\ J_3 &= \bar{E} \left[E \left[\text{tr} (V \varepsilon_\delta \varepsilon_\delta^* V^* + S u_\delta u_\delta^* S^*) + \sum_{i,j} \frac{w_{ij}^2}{c^2} |\tau_{ij}|^2 |s_j|^2 \right] \right], \end{aligned} \quad (21)$$

with $\varepsilon_0 = \frac{1}{c}(HR - D)s$, $\varepsilon_\delta = \frac{1}{c}HTs$, $u_0 = \frac{1}{c}Rs$ and $u_\delta = \frac{1}{c}Ts$. The parameters r_{ij} and τ_{ij} are the elements of R and T respectively. Note that J_1 does not depend on T and that $J_3 \geq 0$. Hence, the only way T can decrease the criterion is through J_2 . Set

$$\alpha = \frac{1}{c^2} \sum_{i,j} w_{ij}^2 r_{ij} \tau_{ij}^* = \frac{1}{c^2} \text{tr} ([W \odot W \odot R] T^*),$$

and recall $E[ss^*] = I$. Then

$$J_2 = \frac{1}{c^2} \bar{E} [\text{tr} (V (HR - D) T^* H^* V^* + S R T^* S^*)] + \alpha. \quad (22)$$

The trace rotation rule, $\text{tr}(AB) = \text{tr}(BA)$ gives

$$J_2 = \frac{1}{c^2} \text{tr} (((\bar{E}[H^* V^* V H] + S^* S) R - \bar{E}[H^* V^* V D]) T^*) + \alpha. \quad (23)$$

Assuming that $\bar{E}[\Delta H] = 0$ and therefore also $\bar{E}[\Delta H \hat{H}^*] = 0$ since \hat{H}^* is assumed deterministic and known, and inserting (3), (10), (11) and (16) into (23) we get $J_2 = 0$ for all T . Hence, we cannot chose a matrix T that will decrease the cost function, $J = J_1 + J_3$. The minimum $J = J_1$ is attained only by setting $T = 0$, so R by (16) minimizes the cost function.

REFERENCES

- [1] J. Lee, Y. Kim, H. Lee, B. L. Ng, D. Mazzaresse, J. Liu, W. Xiao and Y. Zhou, "Coordinated multipoint transmission and reception in LTE-Advanced systems," *IEEE Com. Mag.* **50**, pp. 44-50, 2012.
- [2] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzaresse, S. Nagata, and K. Sayana, "Coordinated multipoint transmission and reception in LTE-advanced deployment: scenarios and operational challenges," *IEEE Wireless Com. Mag.* **50**, pp. 148-155, 2012.
- [3] W. Lee, I. Lee, J.S. Kwak, B. Ihm and S. Han, "Multi-BS MIMO cooperation: challenges and practical solutions in 4G systems," *IEEE Wireless Com.* **19**, pp. 89-96, 2012.
- [4] H. Zhang and H. Dai, "Cochannel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks," *EURASIP Jour. on Wirel. Commun. and Networking*, 2004, doi:10.1155/S1687147204406148.
- [5] P. Marsch, G. Fettweis, "On multicell cooperative transmission in backhaul-constrained cellular systems," *Ann Telecommun.* **63**(5), pp. 253-269, 2008, doi:10.1007/s12243-008-0028-3.
- [6] C.T.K. Ng, H. Huang, "Linear precoding in cooperative MIMO cellular networks with limited coordination clusters," *IEEE Jour. on Sel Areas in Com.* **28**(9), pp 1446-1454, 2010.
- [7] T.R. Lakshmana, C. Botella, and T. Svensson, "Partial joint processing with efficient backhauling using particle swarm optimization," *EURASIP J. on Wireless Communications and Networking*, May, 2012, doi:10.1186/1687-1499-2012-182.
- [8] R. Apelfröjd and M. Sternad, "Design and measurement based evaluations of coherent JT CoMP - a study of precoding, user grouping and resource allocation using predicted CSI," *EURASIP J. on Wireless Communications and Networking*, June, 2014, doi: 10.1186/1687-1499-2014-100.
- [9] R. Apelfröjd, M. Sternad and D. Aronsson, "Measurement-based evaluation of robust linear precoding for downlink CoMP," in *Proc. of IEEE ICC 2012*, Ottawa, Canada, June 2012.
- [10] H. Thiele, M. Olbrich, M. Kurras and B. Matthiesen, "Channel aging effects in CoMP transmission: Gains from linear channel prediction," *45th Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, USA, Nov., 2011.
- [11] D. Aronsson, *Channel Estimation and Prediction for MIMO OFDM Systems - Key Design and Performance Aspects of Kalman-based Algorithms*, Dissertation, Uppsala University, 2011. <http://www.signal.uu.se/Publications/ptheses.html>. Accessed 7 Nov. 2013.
- [12] ARTIST4G D1.4, "Interference avoidance techniques and system design," Artist4G technical deliverable, <https://ict-artist4g.eu/projet/deliverables>. Accessed 7 Nov. 2013.
- [13] R. Apelfröjd, "Kalman predictions for multipoint OFDM downlink channels", Technical report, Uppsala University, May 2014, <http://www.signal.uu.se/Publications/preports.html> Accessed 24 June 2014.
- [14] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, "Network MIMO with clustered linear precoding," *IEEE Trans. Wireless Com.* **8**, pp. 1910-1921 2009.
- [15] W. Mennerich and W. Zirwas, "Reporting effort for cooperative systems applying interference floor shaping," in *Proc. of IEEE PIMRC 2011*, Toronto, Canada, Sept. 2011.
- [16] 3GPP TR 36.814 v9.0.0, "Evolved Universal Terrestrial Radio Access (E-UTRA): Further advancements for E-UTRA physical layer aspects (Release 9)," 2010.
- [17] A. Klockar, M. Sternad, A. Brunstrom and R. Apelfröjd, "User-centric Pre-selection and scheduling for coordinated multipoint systems," *IEEE ISWCS14*, Barcelona, Spain, Aug., 2014.