Coordinated Multipoint Transmission with Limited Backhaul Data Transfer

Jian Zhao, *Member, IEEE,* Tony Q. S. Quek, *Senior Member, IEEE,* and Zhongding Lei, *Senior Member, IEEE*

Abstract

When the joint processing technique is applied in the coordinated multipoint (CoMP) downlink transmission, the user data for each mobile station needs to be shared among multiple base stations (BSs) via backhaul. If the number of users is large, this data exchange can lead to a huge backhaul signaling overhead. In this paper, we consider a multi-cell CoMP network with multi-antenna BSs and single antenna users. The problem that involves the joint design of transmit beamformers and user data allocation at BSs to minimize the backhaul user data transfer is addressed, which is subject to given quality-of-service and per-BS power constraints. We show that this problem can be cast into an ℓ_0-norm minimization problem, which is NP-hard. Inspired by recent results in compressive sensing, we propose two algorithms to tackle it. The first algorithm is based on reweighted ℓ_1-norm minimization, which solves a series of convex ℓ_1-norm minimization problems. In the second algorithm, we first solve the ℓ_2-norm relaxation of the joint clustering and beamforming problem and then iteratively remove the links that correspond to the smallest transmit power. The second algorithm enjoys a faster solution speed and can also be implemented in a semi-distributed manner under certain assumptions. Simulations show that both algorithms can significantly reduce the user data transfer in the backhaul.

Index Terms

Coordinated Multipoint (CoMP) Transmission; Optimization; Uplink-Downlink Duality; Clustering Methods; Array Signal Processing

J. Zhao and Z. Lei are with the Institute for Infocomm Research, A*STAR, 1 Fusionopolis Way, #21-01 Connexis, Singapore 138632 (e-mail: {jzhao, leizd}@i2r.a-star.edu.sg).

T. Q. S. Quek is with Singapore University of Technology and Design (e-mail: qsquek@ieee.org). He is also with the Institute for Infocomm Research, A*STAR, Singapore.

I. Introduction

Coordinated multipoint (CoMP) transmission has been incorporated into the latest 3GPP LTE-Advanced Releases for combating the inter-cell interference (ICI) in cellular wireless systems [1]. The idea is to aggregate neighboring base stations (BSs) and form the *CoMP cooperating set*. Transmission strategies for the BSs within this set are coordinated to effectively mitigate the interference at the mobile stations (MSs). Two major approaches are available in CoMP: coordinated beamforming (CB) and joint processing (JP) [2]. The difference is that the data stream for each MS is only available at and transmitted from a single BS in CB, while the data for each MS in JP is distributed among multiple cooperating BSs and can be transmitted simultaneously from them.

In terms of system throughput improvement, the JP approach is attractive since it simply forms a virtual large antenna array for each MS at the expense of high signaling overhead through the backhaul. This high signaling overhead is incurred by distributing each user's data to multiple BSs in the cooperating set. In a low-mobility environment, the amount of user data to be transported in the backhaul significantly outweighs that of the channel state information (CSI). In order to alleviate the backhaul requirement, it is desirable to distribute the user data only to a subset of the cooperating BSs. This is termed *BS clustering*.

Different distributed beamformer design methods for the CB approach under various performance criteria have been proposed in [3]–[6]. Ref. [3] discusses coordinated beamformer design to minimize the maximum BS antenna power without user data exchange between BSs. A decentralized method to minimize the sum transmit power under given quality-of-service (QoS) requirements has been proposed in [4] for both CB and JP approaches. In [5], each BS maximizes the rate for intra-cell users and cancels inter-cell interference in a decentralized manner. Fast iterative algorithms to maximize the minimum user rate have been proposed in [6]. The optimal assignment of one BS for each user has been addressed in [7]. The JP approach has been considered in [8]–[11]. Ref. [8] considers fully synchronous transmission from cooperating BSs. Forming clusters in multi-cell networks and performing inter-cluster coordination have been considered in [9] to mitigate interference to cluster edge users. In [10], a zero-forcing scheme has been proposed, which requires all the users' data to be distributed on all the BSs. BS clustering has been considered in [11] using heuristic methods. All those works assign the BSs to each MS in a predefined manner.

Weighted sum-rate maximization (WSRM) in multi-cell CoMP networks has been discussed in [12]–[14]. Ref. [12] has designed distributed WSRM algorithms that iteratively update the transmit power and the user assignment at each BS using local information and limited BS data exchange. Algorithms with co-channel user selection and power allocation across tones have been presented in [13] for WSRM in multi-cell multi-carrier networks. Both works [12], [13] assume no user data exchange between BSs. Centralized and distributed WSRM algorithms can be found in [14] for multi-cell networks, where all the users data are available at all the transmitting BSs.

A number of recent research efforts have considered CoMP networks with finite-capacity backhaul [15]–[19]. The information-theoretic analysis for uplink CoMP cellular networks using a linear Wyner-type cellular model with no fading has been presented in [15], where cooperative decoding at the BSs is enabled by local and finite-capacity backhaul links. This work has been extended in [16] to consider Rayleigh-fading channels, where a joint decoding scheme based on distributed Wyner-Ziv compression [20] is compared with a partial local decoding scheme at the BSs. An analysis of various uplink CoMP schemes under a constrained backhaul infrastructure and imperfect CSI has been performed in [17]. The optimal training length impacted by the finite backhaul capacity in an uplink CoMP network has been studied in [18] using large system analysis. A two-cell downlink CoMP transmission scenario with a finite-capacity backhaul has been considered in [19], where a rate-splitting approach has been proposed, i.e., the data symbols transmitted at each BS are composed of private messages that are only available to the BS itself and shared messages that are available at both BSs. A corresponding achievable rate region with respect to the backhaul constraints has been characterized. Those works [15]–[19] focus on the derivation and analysis of achievable rate regions for cellular CoMP networks under finite-capacity backhaul constraints.

In this paper, we address the following problem for the JP approach: which BSs should be assigned with which MS's data and how to design the corresponding beamformers in order to minimize the user data transfer from the data center to the BSs in the backhaul? Specifically, we formulate the problem of minimizing backhaul user data transfer, which jointly determines the optimal BS clustering and the transmit beamformers, into an ℓ_0-norm minimization problem. Unfortunately, this problem is NP-hard. Inspired by recent developments in compressive sensing [21], [22], we propose two algorithms to tackle the problem. The first algorithm is based on the ℓ_1-relaxation and the reweighted ℓ_1-norm minimization, which requires to solve a series of convex ℓ_1-norm minimization problems using, e.g., the interior-point method. The second algorithm is based on iterative link removal. The idea is that we first solve the ℓ_2-norm relaxation of the joint clustering and beamforming problem and then iteratively remove the links that correspond to the smallest link transmit power. This algorithm utilizes the uplink-downlink duality [3], [6], [23] in multi-cell systems and it employs projected subgradient algorithm with fast iterative function evaluation to find the results in each intermediate step. Furthermore, this algorithm can also be implemented in a semi-distributed manner under certain assumptions. Simulation results show that the proposed algorithms can significantly reduce the user data transfer in the backhaul, as well as design the transmit beamformers to satisfy the per-BS power constraints at the BSs and the QoS requirements at the MSs. To the best of our knowledge, formalized discussions and solutions on how to minimize the backhaul data exchange between BSs under given QoS constraints cannot be found in existing literature. Our main contributions are therefore two-fold: one is the generic formulation of the problem and the other is the proposal of analytical near-optimum solution algorithms.

The remainder of the paper is organized as follows: in Section II, we introduce the system model and formulate our joint clustering and beamforming problem into an ℓ_0-norm minimization problem. The reweighted ℓ_1-norm minimization based method is presented in Section III. The iterative link removal based method is presented in Section IV. Section V provides the convergence behavior and the simulation results in a multi-cell scenario. Finally, conclusions are drawn in Section VI.

Notation: we use bold uppercase letters to denote matrices and bold lowercase letters to denote vectors. $\mathcal{CN}(0, \sigma^2)$ denotes a circularly symmetric complex normal zero mean random variable with variance σ^2. $(\cdot)^T$ and $(\cdot)^H$ stand for the transpose and conjugate transpose, respectively. $\mathrm{vec}(\mathbf{A})$ is a column vector composed of the entries of $\mathbf{A}$ taken column-wise. $\mathbb{C}$, $\mathbb{R}$ and $\mathbb{Z}$ stand for the complex, the real and the integer numbers, respectively. $\{\mathbf{w}_{ij}\}$ denotes the set made of $\mathbf{w}_{ij}, \forall i, j$. $\lfloor x \rfloor$ is the largest integer not greater than x. $[\mathbf{X}]_{ij}$ denotes the (i, j)-th element of $\mathbf{X}$. $\mathrm{diag}(\{\mathbf{A}_i\})$ denotes a block diagonal matrix formed by $\mathbf{A}_i, \forall i$.

II. System Model and Problem Formulation

A. System Model

We consider the CoMP downlink transmission as shown in Fig. 1. There are B BSs cooperating to transmit to K MSs using the same time and frequency channel. Here $x_i \in \mathbb{C}$ denotes the complex data symbol for the ith MS, where $i \in \{1, \cdots, K\}$, and $\mathrm{E}\left\{|x_i|^2\right\} = 1$. The data center has access to the data of all the MSs. The backhaul is the channel connecting the data center to the BSs, which can be optical fiber or out-of-band microwave links. The control information and user data at each BS is obtained directly from the data center via the backhaul [24], [25].

We assume that each BS is equipped with M transmit antennas and each MS has a single receive antenna, where $\mathbf{w}_{ij} \in \mathbb{C}^M$ denotes the transmit beamformer for the ith MS on the jth BS, such that $j \in \{1, \cdots, B\}$ and the transmit power constraint at the jth BS is $\sum_{i=1}^{K} \mathbf{w}_{ij}^H \mathbf{w}_{ij} \leq P_j$. The channel from the jth BS to the ith MS is denoted as $\mathbf{h}_{ij}^H$, where $\mathbf{h}_{ij} \in \mathbb{C}^M$. The received signal y_i at the ith MS can be expressed as

$$y_i = \sum_{j=1}^{B} \mathbf{h}_{ij}^H \mathbf{w}_{ij} x_i + \sum_{k \neq i}^{K} \sum_{j=1}^{B} \mathbf{h}_{ij}^H \mathbf{w}_{kj} x_k + n_i. \tag{1}$$

The first term on the right-hand side of (1) represents the received useful signal, the second term represents the interference, and $n_i \sim \mathcal{CN}(0, \sigma^2)$ is the additive white Gaussian noise (AWGN) at the ith MS. The signal-to-interference-plus-noise ratio (SINR) at the ith MS can then be written as

$$\mathsf{SINR}_i = \frac{\left|\sum_{j=1}^{B} \mathbf{h}_{ij}^H \mathbf{w}_{ij}\right|^2}{\sigma^2 + \sum_{k \neq i}^{K} \left|\sum_{j=1}^{B} \mathbf{h}_{ij}^H \mathbf{w}_{kj}\right|^2}. \tag{2}$$

The SINR is an important QoS metric. SINR requirements have to be satisfied for the successful transmission of data. We denote the SINR requirement for the ith MS as γ_i. For ease of discussion, we only consider fixed and feasible target SINR requirements. In a commercial cellular network, the actual transmission rate will usually be adjusted with dynamic link adaptation.

To minimize the signaling overhead in the backhaul, we want to distribute the user data only to the minimum number of cooperating BSs, while satisfying the SINR constraint of each MS. We denote the set of BS indices that the user data x_i is distributed to as $\mathcal{B}_i$, where $\mathcal{B}_i \subseteq \{1, \cdots, B\}$. We assume the data rate to transfer the data x_i, $\forall i \in \{1, \cdots, K\}$, through the backhaul to each of the BS in $\mathcal{B}_i$ to be equal and denote it as R. Thus, the sum of user data rate R_{sum} transferred from the data center through the backhaul is given by

$$R_{\text{sum}} = R \cdot \sum_{i=1}^{K} |\mathcal{B}_i| \tag{3}$$

where $|\mathcal{B}_i|$ denotes the cardinality of the set $\mathcal{B}_i$. For the given SINR requirement, we obtain that minimizing R_{sum} is equivalent to minimizing $\sum_{i=1}^{K} |\mathcal{B}_i|$.

B. Problem Formulation

We define a data routing matrix $\mathbf{P} \in \mathbb{R}_+^{K \times B}$ to capture the data transfer in the backhaul, where the (i, j)-th element of $\mathbf{P}$ is given by

$$[\mathbf{P}]_{ij} = \mathbf{w}_{ij}^H \mathbf{w}_{ij} \tag{4}$$

and it represents the power allocated at the jth BS for the data x_i. The matrix $\mathbf{P}$ also shows which MS's data are routed from the data center to each of the BSs. When $[\mathbf{P}]_{ij} = 0$, the data x_i is not routed to BS j. When $[\mathbf{P}]_{ij} > 0$, we say there exists a *link* between the jth BS and the ith MS, and the data center needs to distribute x_i to the BSs with $[\mathbf{P}]_{ij} > 0$ before the BSs transmit to the MSs. Therefore, $\mathcal{B}_i$ corresponds to the indices of the nonzero elements of the ith row in $\mathbf{P}$ and $\sum_{i=1}^{K} |\mathcal{B}_i|$ is equivalent to counting the number of nonzero elements in $\mathbf{P}$. Since the size of $\mathbf{P}$ is fixed and equals $K \times B$, minimizing $\sum_{i=1}^{K} |\mathcal{B}_i|$ is equivalent to maximizing the number of zero elements, i.e., the *sparsity*, in $\mathbf{P}$. We can mathematically formulate the R_{sum} minimization problem in (3) as

$$\begin{aligned} \min_{\{\mathbf{w}_{ij}\}} \quad & \|\text{vec}(\mathbf{P})\|_0 \\ \text{subject to} \quad & \mathsf{SINR}_i \geq \gamma_i, \ \forall i \in \{1, \ldots, K\} \\ & \textstyle\sum_{i=1}^{K} \mathbf{w}_{ij}^H \mathbf{w}_{ij} \leq P_j, \ \forall j \in \{1, \ldots, B\} \end{aligned} \tag{5}$$

where $\|\cdot\|_0$ represents the ℓ_0-norm, which denotes the number of nonzero elements of a vector. Note that the optimization problem (5) needs to jointly determine the BS subsets for the routing of each user's data, as well as to design the transmit beamformers at the BSs.

Remark 1: The problem (5) is a combinatorial optimization problem and is NP-hard.

The proof for a related problem where all the constraints are linear can be found in [26]. Using similar arguments as in [26], [27], the problem (5) is NP-hard. A brute-force solution to a combinatorial optimization problem like (5) is by exhaustive search. That is, for each number $S \in \mathbb{Z}$, where $0 \leq S \leq KB$, we must check all the $\binom{KB}{S}$ possible assignments of zero elements in $\mathbf{P}$. For each assignment, we must search for the $\{\mathbf{w}_{ij}\}$ satisfying the constraints of (5). In the end, we pick out the maximum S with feasible $\{\mathbf{w}_{ij}\}$. However, the complexity of exhaustive search grows exponentially with the size of $\mathbf{P}$, which is not applicable in real-world applications.

The difficulty to solve (5) lies in the combinatorial non-convex objective function involving the ℓ_0-norm. One approach to obtain approximate solutions for non-convex optimization problems is convex relaxation [28], i.e., replacing the non-convex objective function with its convex envelop. The convex envelope of the ℓ_0-norm is the ℓ_1-norm. Optimization based on the ℓ_1-norm has been shown to exactly recover sparse signals in ℓ_0-norm problems and closely approximate compressible signals with high probability using relatively few measurements if certain property is satisfied [21], [29]. This is known as *basis pursuit* [21], [29] that conveniently reduces a sparse estimation problem to a linear program in signal detection, which is the essence of compressive sensing. However, simply substituting the objective function of (5) by $\|\mathrm{vec}\,(\mathbf{P})\|_1$ does not produce sparse solutions in general and such a relaxation is not useful at first glance. Therefore, we provide in the next section an equivalent formulation of (5), where sparse solutions can be obtained.

III. Reweighted ℓ_1-Norm Minimization Based Method

We observe that $[\mathbf{P}]_{ij} > 0$ or $= 0$ is equivalent to $\|\mathbf{w}_{ij}\|_2 > 0$ or $= 0$, respectively. By introducing slack variables s_{ij}, the problem (5) can be equivalently reformulated as

$$\begin{aligned} \min_{\{\mathbf{w}_{ij}\},\{s_{ij}\}} \quad & \|\mathrm{vec}\,(\mathbf{S})\|_0 \\ \text{subject to} \quad & \mathsf{SINR}_i \geq \gamma_i, \ \forall i \in \{1,\ldots,K\} \\ & \|\mathbf{w}_{ij}\|_2 \leq s_{ij}, \ \forall i,j \\ & \textstyle\sum_{i=1}^{K} s_{ij}^2 \leq P_j, \ \forall j \in \{1,\ldots,B\} \end{aligned} \tag{6}$$

where $[\mathbf{S}]_{ij} = s_{ij}$. The first and second constraints are second-order cone (SOC) constraints, the third constraint is a sum-of-squares constraint, and all those constraints are convex. The problem (6) is equivalent to (5). Applying the ℓ_1-norm relaxation to (6), we obtain

$$\begin{aligned} \min_{\{\mathbf{w}_{ij}\},\{s_{ij}\}} \quad & \textstyle\sum_{i,j} s_{ij} \\ \text{subject to} \quad & \text{Constraints of (6).} \end{aligned} \tag{7}$$

Now the objective function becomes linear and the problem (7) is a convex optimization problem. Furthermore, sparse solutions of $\{s_{ij}\}$ can be obtained. Note only this formulation of the ℓ_1-relaxation directly produces sparse solutions and other types of convex relaxations do not produce sparse solutions in general [21], [22]. The following proposition states which s_{ij} is obtained with zero value solution in (7).

Proposition 1: The solutions of $\{s_{ij}\}$ in (7) is obtained at $s_{ij}^* = 0$ if $\rho_{ij} < 1$, where ρ_{ij} is the dual variable associated with the constraint "$\|\mathbf{w}_{ij}\|_2 \leq s_{ij}$" in (7), $\forall i \in \{1,\cdots,K\}$ and $j \in \{1,\cdots,B\}$.

Proof: See Appendix A. □

The solutions obtained from (7) are suboptimal solutions of (6) due to convex relaxation. Improved solutions can be obtained by applying a weighting factor on each s_{ij} in the objective function of (7). This is called *reweighted ℓ_1-norm minimization methods* [30], originally proposed to enhance the data acquisition in compressive sensing. The weighting factors should be chosen such that the solutions with large s_{ij} values are penalized. A good choice of the weighting factor β_{ij} in the $(n+1)$-th iteration is $\beta_{ij}^{(n+1)} = 1/\left(\delta + \left|s_{ij}^{(n)}\right|\right)$, where $s_{ij}^{(n)}$ is the solution of s_{ij} in the nth iteration, and δ is a small nonnegative parameter to ensure numerical stability. Applying this method to our problem, the resulting sparsity of $\{s_{ij}\}$ typically converges within 10 iterations and the largest improvement in sparsity is obtained in the first few iterations. Although still suboptimal, this algorithm produces a better approximate solution of the original problem (6) and it is summarized in Algorithm 1.

Algorithm 1 Reweighted ℓ_1-norm minimization based method

1: Set the iteration counter $n = 0$ and $\beta_{ij}^{(0)} = 1, \forall i \in \{1, \cdots, K\}$ and $j \in \{1, \cdots, B\}$.

2: Solve the weighted ℓ_1-norm minimization problem

$$\left\{\mathbf{w}_{ij}^{(n+1)}, s_{ij}^{(n+1)}\right\} = \arg\min \sum_{i,j} \beta_{ij}^{(n)} s_{ij}, \quad \text{subject to: Constraints of (6).}$$

3: Update the weighting factor: $\beta_{ij}^{(n+1)} = 1/\left(\delta + \left|s_{ij}^{(n)}\right|\right), \forall i, j.$

4: Calculate the relative error $r = \left\|\text{vec}\left(\mathbf{S}^{(n)} - \mathbf{S}^{(n-1)}\right)\right\|_2 / \left\|\text{vec}\left(\mathbf{S}^{(n)}\right)\right\|_2$. Terminate if $r < \varepsilon$ or $n \geq N_{\text{max}}$, where ε is a predefined threshold and N_{max} is the specified maximum number of iterations. Otherwise, set $n = n + 1$ and go to step 2.

Furthermore, we have the following proposition for Algorithm 1:

Proposition 2: Algorithm 1 converges to a local optimal solution.

Proof: See Appendix B. □

The reweighted ℓ_1-norm minimization based method achieves better sparsity results than (7) because although the ℓ_1-norm is the best *convex* relaxation of the ℓ_0-norm, some *concave* functions can offer tighter approximations [21], [28]. In the proof, the choice of the weighting factors is explained using the log-sum surrogate function of the ℓ_0-norm [30], [31]. Because the log-sum penalty function is concave, the global minimum solution may not always be attained. Choosing a suitable starting point is important for the final result. It is suggested in [30], [32] to initialize the algorithm by the solution of the unweighted ℓ_1-norm minimization.

IV. Iterative Link Removal Based Method

The reweighted ℓ_1-norm minimization based method needs to solve a series of convex optimization problems using, e.g., interior-point methods, which can be computationally intensive. In this section, we propose another approach. The basic idea is that we first solve the ℓ_2-norm relaxation of the joint clustering and beamforming problem (6) and then iteratively remove the links that correspond to the smallest link transmit power. Methods employing fixed-point iterations [6], [33] are utilized to find the solution in each intermediate step, which enjoy low complexity and are usually faster than the interior-point methods. We show that the iterative link removal based method can also be implemented in a semi-distributed manner with a lightweight central controller under certain assumptions. The computational complexity analysis will be presented at the end of this section.

A. ℓ_2-Norm Relaxation of Joint Clustering and Beamforming

We first consider the ℓ_2-norm relaxation of the joint clustering and beamforming problem (6) subject to SINR and per-BS power constraints, which can be expressed as

$$\begin{aligned} \min_{\{\mathbf{w}_{ij}\}} \quad & \textstyle\sum_{i,j} \mathbf{w}_{ij}^H \mathbf{w}_{ij} \\ \text{subject to} \quad & \mathsf{SINR}_i \geq \gamma_i, \ \forall i \in \{1, \ldots, K\} \\ & \textstyle\sum_i \mathbf{w}_{ij}^H \mathbf{w}_{ij} \leq P_j, \ \forall j \in \{1, \ldots, B\}. \end{aligned} \tag{8}$$

Note that the power constraints P_j, $\forall j \in \{1, \ldots, B\}$, in (8) are constants. The problem (8) is different from the min-max BS antenna power problem considered in [3], [23], where the per-BS transmit power constraints are variables to be optimized. The problem (8) is a SOC programming problem that is convex and it can be solved using the interior-point methods. However, that may be computationally intensive. Therefore, we exploit the special structure of (8) and propose a low-complexity approach that employs the fixed-point methods. Our proposed approach is based on the following theorem, which generalizes the uplink-downlink duality to multi-cell environment under *explicitly given* per-BS power constraints.

Theorem 1: The dual problem of the downlink transmit optimization problem (8) is

$$\begin{aligned} \max_{\{\nu_j\}} \min_{\{q_i\},\{\hat{\mathbf{w}}_{ij}\}} \quad & \textstyle\sum_{i=1}^{K} q_i\sigma^2 - \sum_{j=1}^{B} \nu_j P_j \\ \text{subject to} \quad & \mathsf{SINR}_i^{\mathrm{U}} \geq \gamma_i,\ q_i \geq 0, \quad \forall i \in \{1,\ldots,K\} \\ & \nu_j \geq 0, \quad \forall j \in \{1,\ldots,B\} \end{aligned} \tag{9}$$

where

$$\mathsf{SINR}_i^{\mathrm{U}} = \frac{q_i \left| \sum_{j=1}^{B} \hat{\mathbf{w}}_{ij}^H \mathbf{h}_{ij} \right|^2}{\sum_{k\neq i}^{K} q_k \left| \sum_{j=1}^{B} \hat{\mathbf{w}}_{ij}^H \mathbf{h}_{kj} \right|^2 + \sum_{j=1}^{B} (1+\nu_j) \left\| \hat{\mathbf{w}}_{ij} \right\|_2^2}. \tag{10}$$

This is an uplink multiuser receive beamforming problem in the dual uplink channel with the same SINR constraints and uncertain noise variances, where ν_j is the dual variable associated with the jth BS power constraint in (8), and $1+\nu_j$ can be interpreted as the dual uplink noise variance at the jth BS in (9). Here q_i is the dual variable associated with the ith SINR constraint in (8) and it can be interpreted as the dual uplink transmit power of the ith MS. Here $\hat{\mathbf{w}}_{ij}$ represents the uplink receive beamformer at the jth BS for the data of MS i. Note that the optimal $\hat{\mathbf{w}}_{ij}$ is a scaled version of the optimal $\mathbf{w}_{ij}$ for the downlink.

Proof: See Appendix C. □

The problem (9) consists of an outer maximization with regard to the dual uplink noise variances $\{\nu_j\}$ and an inner minimization with regard to the uplink power $\{q_i\}$ and beamformers $\{\hat{\mathbf{w}}_{ij}\}$. The optimal objective value of the inner minimization is a function of $\{\nu_j\}$, i.e.,

$$\begin{aligned} \varphi(\{\nu_j\}) = \min_{\{q_i\},\{\hat{\mathbf{w}}_{ij}\}} \quad & \textstyle\sum_i q_i\sigma^2 - \sum_j \nu_j P_j \\ \text{subject to} \quad & \mathsf{SINR}_i^{\mathrm{U}} \geq \gamma_i,\ q_i \geq 0, \quad \forall i \in \{1,\ldots,K\}. \end{aligned} \tag{11}$$

For given $\{\nu_j\}$, the term $\sum_j \nu_j P_j$ is a fixed constant. Evaluating $\varphi(\{\nu_j\})$ for given $\{\nu_j\}$ is a uplink joint sum power minimization and beamforming problem with fixed BS noise variances [34]. In the following, we apply a fixed-point algorithm to solve (11).

Denoting all the BSs as a BS group, we define $\mathbf{h}_i = \left[\mathbf{h}_{i1}^H, \mathbf{h}_{i2}^H, \cdots, \mathbf{h}_{iB}^H\right]^H$, where $i \in \{1,\ldots,K\}$. That is the dual uplink channel vector from the ith MS to the BS group. Concatenating the BS downlink transmit beamformers together, we obtain $\mathbf{w}_i = \left[\mathbf{w}_{i1}^H, \mathbf{w}_{i2}^H, \cdots, \mathbf{w}_{iB}^H\right]^H$ which is the BS group downlink beamformer to the ith MS. Likewise, we concatenate the BS receive beamformers together as $\hat{\mathbf{w}}_i = \left[\hat{\mathbf{w}}_{i1}^H, \hat{\mathbf{w}}_{i2}^H, \cdots, \hat{\mathbf{w}}_{iB}^H\right]^H$. Define $\mathbf{Q}_j$ as an all-zero $MB \times MB$ matrix, except that the $(jM-M+1)$-th to the jM-th diagonal elements are all 1's, and we can express $\mathsf{SINR}_i^{\mathrm{U}}$ as

$$\mathsf{SINR}_i^{\mathrm{U}} = \frac{q_i \hat{\mathbf{w}}_i^H \mathbf{h}_i \mathbf{h}_i^H \hat{\mathbf{w}}_i}{\hat{\mathbf{w}}_i^H \mathbf{\Sigma}_i \hat{\mathbf{w}}_i} \tag{12}$$

where

$$\mathbf{\Sigma}_i = \sum_{k \neq i}^{K} q_k \mathbf{h}_k \mathbf{h}_k^H + \left(\mathbf{I} + \sum_{j=1}^{B} \nu_j \mathbf{Q}_j \right). \tag{13}$$

Since the optimal receive beamformer $\hat{\mathbf{w}}_i$ that maximizes $\mathsf{SINR}_i^{\mathrm{U}}$ is the minimum mean square error (MMSE) receiver [33], [35], we have

$$\hat{\mathbf{w}}_i = \mathbf{\Sigma}_i^{-1} \mathbf{h}_i \tag{14}$$

and the corresponding SINR is then given by

$$\mathsf{SINR}_i^{\mathrm{U,\ max}} = q_i \mathbf{h}_i^H \mathbf{\Sigma}_i^{-1} \mathbf{h}_i. \tag{15}$$

At the optimal solution, it is necessary that the SINR constraints in (11) are satisfied with equality [36], i.e., $\mathsf{SINR}_i^{\mathrm{U,\ max}} = \gamma_i$. Therefore the optimal uplink transmit power satisfies the following set of equations

$$q_i^* = \frac{\gamma_i}{\mathbf{h}_i^H \mathbf{\Sigma}_i^{-1} \mathbf{h}_i} \triangleq f_i(\mathbf{q}^*), \quad \forall i \in \{1, \cdots, K\} \tag{16}$$

where $\mathbf{q}^* = [q_1^*, q_2^*, \cdots, q_K^*]^T$. Using fixed-point methods, we can determine the optimal uplink transmit power as follows:

$$q_i^{(l+1)} = f_i(\mathbf{q}^{(l)}), \quad \forall i \in \{1, \cdots, K\}. \tag{17}$$

Starting from some initial value of $\mathbf{q}^{(0)}$, the fixed-point equation (17) is guaranteed to converge to a unique fixed-point because $f_i(\mathbf{q})$ is a standard function [33], [37]. Note the $f_i(\mathbf{q})$ in (16) is different from that of [3, Eq. (29)] in the expressions of the covariance matrix and the coefficient, and a faster convergence speed by our formulation is observed in simulations.

After the optimal $\mathbf{q}^*$ is obtained, the optimal receive beamformer $\hat{\mathbf{w}}_i$ can be calculated according to (14). The optimal downlink beamformer $\mathbf{w}_i$ is a scaled version of $\hat{\mathbf{w}}_i$, i.e., $\hat{\mathbf{w}}_i = \sqrt{\rho_i}\mathbf{w}_i, \forall i \in \{1, \cdots, K\}$. The scaling factor ρ_i can be calculated using the fact that the downlink SINR constraints in (8) must be satisfied with equality for the optimal solution, i.e.,

$$\frac{1}{\gamma_i} \rho_i \left| \mathbf{h}_i^H \hat{\mathbf{w}}_i \right|^2 = \sigma^2 + \sum_{k \neq i}^{K} \rho_k \left| \mathbf{h}_i^H \hat{\mathbf{w}}_k \right|^2, \quad \forall i \in \{1, \ldots, K\}. \tag{18}$$

Denoting $\boldsymbol{\rho} = [\rho_1, \cdots, \rho_K]^T$, we can rewrite (18) as $\mathbf{F}\boldsymbol{\rho} = \mathbf{1}\sigma^2$, where $\mathbf{1}$ is a $K \times 1$ all-one vector and the elements of $\mathbf{F}$ are defined as

$$[\mathbf{F}]_{ik} = \begin{cases} \frac{1}{\gamma_i} \left|\mathbf{h}_i^H \hat{\mathbf{w}}_i\right|^2, & i = k; \\ -\left|\mathbf{h}_i^H \hat{\mathbf{w}}_k\right|^2, & i \neq k. \end{cases} \tag{19}$$

So the scaling vector $\boldsymbol{\rho}$ can be obtained as

$$\boldsymbol{\rho} = \mathbf{F}^{-1}\mathbf{1}\sigma^2. \tag{20}$$

The downlink transmit power at the jth BS for the ith MS data stream is

$$p_{ij} = \|\mathbf{w}_{ij}\|_2^2 = \rho_i \|\hat{\mathbf{w}}_{ij}\|_2^2 \tag{21}$$

where $\hat{\mathbf{w}}_{ij}$ is the $M \times 1$ vector formed by the $(jM - M + 1)$-th to the jM-th elements of the $\hat{\mathbf{w}}_i$ obtained from (14) and the downlink transmit power at the jth BS is

$$p_j = \sum_{i=1}^{K} p_{ij}. \tag{22}$$

Hence, this solves the inner minimization for given $\{\nu_i\}$.

The outer maximization can be solved using the projected subgradient method [38]–[40]. Projected subgradient methods following, e.g., the square summable but not summable step size rule, have been proved to converge to the optimal value [38]–[40]. Similar to [23], it can be shown that $\varphi(\{\nu_j\})$ is concave in $\{\nu_j\}$, and a subgradient of $\varphi(\{\nu_j\})$ with respect to ν_j is $p_j - P_j$. With a step size t_n, the dual uplink noise variance $\{\nu_j\}$ can be updated as

$$\nu_j^{(n+1)} = \max\left\{\nu_j^{(n)} + t_n(p_j - P_j), 0\right\}, \quad \forall j = \{1, \cdots, B\}. \tag{23}$$

Since (8) is convex and Slater's condition holds, the respective primal and dual objective values in (8) and (9) must be equal at the global optimum point. A stopping criterion for updating (23) can be that $\left|\varphi\left(\left\{\nu_j^{(n)}\right\}\right) - \sum_j p_j^{(n)}\right| \leq \epsilon$. If the stopping criterion cannot be satisfied after a predefined maximum number N_{m} iterations, the iteration can be aborted and it indicates that (8) is infeasible.

B. Iterative Link Removal

The solutions obtained from (8) is not sparse in general. Therefore, we propose to trim the links, which correspond to the entries of $\mathbf{P}$, to obtain sparse solutions. We call this process *iterative link* *removal*. After the problem (8) is solved, the data streams can be sorted according to the p_{ij} in (21) from the lowest to the highest. We use $(i,j)^*$ to denote the position of the (i,j)-th data stream in the sorted list, and we propose to gradually "remove" the data streams according to their p_{ij} from the lowest to the highest. That is, for each $0 \leq S \leq KB, S \in \mathbb{Z}$, we set the corresponding transmission power for data streams $(i,j)^* \leq S$ to zero, i.e., $\mathbf{w}_{ij} = \mathbf{0}, \forall\,(i,j)^* \leq S$. Here $S = 0$ corresponds to full BS cooperation, i.e., no removal. Meanwhile, we solve the following problem to see if the SINR and power constraints are still feasible.

$$\begin{array}{ll} \min\limits_{\{\mathbf{w}_{ij}\}} & \sum_{i,j} \mathbf{w}_{ij}^H \mathbf{w}_{ij} \\ \text{subject to} & \mathsf{SINR}_i \geq \gamma_i,\ \forall i \in \{1,\ldots,K\} \\ & \sum_i \mathbf{w}_{ij}^H \mathbf{w}_{ij} \leq P_j,,\ \forall j \in \{1,\ldots,B\} \\ & \mathbf{w}_{ij} = \mathbf{0},\ \forall\,(i,j)^* \leq S, \text{except that } (i,j) \text{ is the last link for MS } i. \end{array} \tag{24}$$

This problem is same to (8) except for the last constraint. Here, S corresponds to the sparsity of the routing matrix $\mathbf{P}$ in (4). By increasing S, we can find the largest S that satisfies the constraints of (24). For each user index i, we must keep at least one $\mathbf{w}_{ij} \neq \mathbf{0}$ within the removal process, i.e, we always skip the last transmission link to MS i, which corresponds to the link from its serving cell. The problem (24) is also a SOC programming problem and its feasibility can be determined using, e.g., the interior-point methods. However, by exploiting the similarity between (24) and (8), we can also determine the feasibility of (24) with slight modification of the algorithm in Section IV-A.

As mentioned previously, we denote the set of BS indices that have the data stream for MS i as $\mathcal{B}_i$ for the given value of S. Following the same discussion as in Section IV-A, the dual uplink SINR for the data stream of MS i can be expressed as

$$\mathsf{SINR}_i^{\mathrm{U}} = \frac{q_i \left|\sum_{j\in\mathcal{B}_i} \hat{\mathbf{w}}_{ij}^H \mathbf{h}_{ij}\right|^2}{\sum_{k\neq i}^K q_k \left|\sum_{j\in\mathcal{B}_i} \hat{\mathbf{w}}_{ij}^H \mathbf{h}_{kj}\right|^2 + \sum_{j\in\mathcal{B}_i} (1+\nu_j)\left\|\hat{\mathbf{w}}_{ij}\right\|_2^2}. \tag{25}$$

Note the dual uplink SINR for MS i only depends on the dual uplink channel to those BS $j \in \mathcal{B}_i$. For $m = 1, \cdots, K$, let $\tilde{\mathbf{h}}_{mj} = \mathbf{h}_{mj}, \forall j \in \mathcal{B}_i$, and $\tilde{\mathbf{h}}_{mj} = \mathbf{0}, \forall j \notin \mathcal{B}_i$. We introduce the equivalent channel $\tilde{\mathbf{h}}_m = [\tilde{\mathbf{h}}_{m1}^H, \cdots, \tilde{\mathbf{h}}_{mB}^H]^H$, and use $\tilde{\mathbf{h}}_m$ instead of $\mathbf{h}_m, \forall m \in \{1, \cdots, K\}$ when calculating $\hat{\mathbf{w}}_i$ and q_i according to (13), (14) and (16). It only affects the inner minimization, and the outer maximization process remains unchanged. Whether (24) is feasible is indicated by whether the algorithm converges, e.g., whether the stopping criterion can be satisfied within N_{m} iterations.

The iterative link removal based method is summarized in Algorithm 2. The bisection method [40], [41] is applied to speed up the link removal process.

Algorithm 2 Iterative link removal based method

Start: Solve the problem (8) using the fixed-point iteration (17) and subgradient update (23);

if not convergent **then**

  The problem (8) is infeasible. Exit to link adaptation;

else

  Sort the data streams according to their power p_{ij};

  Initialize: $S_{\min} = 0, S_{\max} = KB$, and $S = \lfloor (S_{\min} + S_{\max})/2 \rfloor$.

  while $S_{\max} - S_{\min} \geq 2$ **do**

    Solve problem (24) for S using the fixed-point iteration (17) and subgradient update (23) with equivalent channel $\tilde{\mathbf{h}}_m, \forall m = 1, \cdots, K$, for calculating q_i, $\hat{\mathbf{w}}_i$ and $\mathbf{w}_i$;

    if infeasible **then**

      $S_{\max} = S, S = \lfloor (S_{\min} + S_{\max})/2 \rfloor$;

    else

      $S_{\min} = S, S = \lfloor (S_{\min} + S_{\max})/2 \rfloor$;

    end if

  end while

  return S and its corresponding $\{\mathbf{w}_{ij}\}$.

end if

C. Implementation Issues

Similar to the reweighted ℓ_1-norm minimization based method, the iterative link removal based method can be implemented at a centralized controller that performs all the computations. However, if the cross-covariance terms between the channel vectors of different BSs are small for each user compared to the auto-covariance terms and thus can be neglected, i.e., $\mathbf{h}_{kj}\mathbf{h}_{kj'}^H \approx \mathbf{0}, \forall j \neq j'$, which is usually valid for multi-cell networks due to the distance between different BSs, the iterative link removal based method can also be implemented in a semi-distributed manner in time-division duplexing (TDD) mode, where the downlink channels can be estimated from the uplink transmission.

Consider the calculation of $\mathbf{\Sigma}_i$ in (13). Since the cross-covariance terms between the channels of different BSs are small for each user k, we have $\mathbf{h}_k\mathbf{h}_k^H \approx \mathrm{diag}\left(\left\{\mathbf{h}_{kj}\mathbf{h}_{kj}^H\right\}\right)$ and $\mathbf{\Sigma}_i \approx \mathrm{diag}\left(\left\{\sum_{k\neq i}^K q_k\mathbf{h}_{kj}\mathbf{h}_{kj}^H + (1+\nu_j)\mathbf{I}\right\}\right) \triangleq \mathrm{diag}\left(\{\mathbf{\Sigma}_{ij}\}\right)$, where the set of matrices inside $\mathrm{diag}(\cdot)$ are indexed by $j = 1, \cdots, B$. That means, the matrix $\mathbf{\Sigma}_i$ becomes *block-diagonal*. Moreover, the jth component block, i.e., the covariance matrix $\mathbf{\Sigma}_{ij}$, only involves the channel vector $\mathbf{h}_{kj}$ and thus it can be estimated locally at the jth BS. Therefore, the calculation of $\hat{\mathbf{w}}_i$ in (14) can be divided into the calculation of $\hat{\mathbf{w}}_{ij}$ at the jth BS, where

$$\hat{\mathbf{w}}_{ij} \approx \mathbf{\Sigma}_{ij}^{-1}\mathbf{h}_{ij}, \quad \forall j = 1, \cdots, B. \tag{26}$$

Furthermore, in the fixed-point iteration (16), we have

$$\frac{\gamma_i}{\mathbf{h}_i^H\mathbf{\Sigma}_i^{-1}\mathbf{h}_i} \approx \frac{\gamma_i}{\sum_{j=1}^B \mathbf{h}_{ij}^H\mathbf{\Sigma}_{ij}^{-1}\mathbf{h}_{ij}} \tag{27}$$

where the term $\mathbf{h}_{ij}^H\mathbf{\Sigma}_{ij}^{-1}\mathbf{h}_{ij}$ can be computed locally at the jth BS.

As a result, the semi-distributed implementation of the iterative link removal based method in TDD mode can be described as follows: i) Each MS-i chooses an initial uplink transmit power $q_i^{(0)}$. ii) Each BS-j estimates the channels $\mathbf{h}_{ij}$ from all the MSs and calculates $\mathbf{\Sigma}_{ij}$ locally. iii) Each BS-j calculates and sends the value of $\mathbf{h}_{ij}^H\mathbf{\Sigma}_{ij}^{-1}\mathbf{h}_{ij}$ to the central controller. iv) The controller sums up the values of $\mathbf{h}_{ij}^H\mathbf{\Sigma}_{ij}^{-1}\mathbf{h}_{ij}$ from all the BSs and updates q_i according to (17), (27) using fixed-point iterations. v) With the $\hat{\mathbf{w}}_{ij}$ computed locally at the BS-j using (26) corresponding to the final q_i, the scaling factor ρ_i can be obtained using the well-known distributed downlink power control algorithm [42] with the fixed effective channel for achieving the given SINR requirements γ_i. Therefore, the downlink transmit power p_{ij} in (21) can also be obtained at the BS-j. vi) The update of ν_j in (23) can be done on a per-BS basis. vii) The BSs send the calculated p_{ij} to the controller and the controller sorts them to decide the removal sequence.

In the semi-distributed implementation process, only scalars, e.g., the value of $\mathbf{h}_{ij}^H\mathbf{\Sigma}_{ij}^{-1}\mathbf{h}_{ij}$, need to be communicated from the BSs to the central controller. Full knowledge of the channel vectors are not required therein. Moreover, the central controller only performs simple calculations like summation, scalar division and sorting. Most of the computation loads, e.g., the matrix inversion, are distributed among the BSs. Therefore, the central controller can be "lightweight". The drawback of the semi-distributed implementation may be the delay induced in the iterations, which also depends on the convergence speed of the subgradient updates.

D. Complexity Analysis

Algorithm 1 needs to solve a series of SOC program. For each of the SOC program, the computational complexity using the interior-point algorithm is $O((KB)^{3.5}M^3 \log \epsilon_1^{-1})$, where ϵ_1 is the required accuracy of the duality gap termination [43]. In the worst case, $N_{\max}$ iterations must be performed. Therefore the worst case complexity for Algorithm 1 is $O(N_{\max}(KB)^{3.5}M^3 \log \epsilon_1^{-1})$. In Algorithm 2, the most computation-intensive step in the inner-minimization is the matrix inversion in (14) for all the K users. Each matrix inversion has the complexity of $O((BM)^3)$ [44]. Using the convergence results of the fixed-point algorithm in [6], [33], [45] and considering N_{m} subgradient updates required in the worst case, the complexity of of solving (9) is obtained as $O(N_{\mathrm{m}}K(BM)^3 \log c^{-1})$, where c^{-1} is a constant that determines the convergence rate of the fixed-point iterations. By applying the bisection method [28] to remove the links, the complexity of Algorithm 2 is therefore obtained as $O(N_{\mathrm{m}}K(BM)^3 \log(KB) \log \epsilon^{-1})$. Compared to the complexity of $O(2^{KB}(KBM)^3\sqrt{K+B} \log \epsilon_1^{-1})$ required by the exhaustive search method [46], both of the two algorithms save a lot of computational complexity.

V. Simulation Results

A. Convergence Behavior

We consider the case that the number of BSs and the number of MSs are $B = K = 5$. Each BS has $M = 3$ transmit antennas and each MS has a single receive antenna. The SINR requirement for each MS receiver is set to $\gamma_i = 6.02\text{dB}$, $\forall i \in \{1, \cdots, K\}$. The noise variances at the MSs are normalized to 1, i.e., $\sigma^2 = 1$. Furthermore, the per-BS transmit power constraint P_j is set according to $P_j/\sigma^2 = 10\text{dB}$, $\forall j \in \{1, \cdots, B\}$. A randomly generated channel is used to simulate the convergence results.

The convergence behavior for the reweighted ℓ_1-norm minimization based method, i.e., Algorithm 1, is shown in Fig. 2, where the parameter δ is set to be $10^{-2}, 10^{-3}$ and 10^{-4}, respectively. The number of active links corresponds to $\sum_{i=1}^{K} |\mathcal{B}_i|$ in (3). Fig. 2 shows that it decreases with the number of iterations. The first several iterations lead to the biggest improvement. As iterations go on, there is no further improvement after the tenth iteration. In the first iteration, i.e., in the unweighted case (7), twenty links are required to be active. That means $KB - 20 = 5$ entries in the data routing matrix $\mathbf{P}$ are zeros, which shows a sparse solution is obtained. In the final iteration, only six links are required to be active. Compared to full cooperation, which requires all the links to be active, we save $1-6/25 = 76\%$ of the user data transfer in the backhaul. In general, a smaller δ leads to better convergence behavior. This is because a small δ approximates the ℓ_0-norm better. However, the convergence for the different δ values does not differ much in Fig. 2. When $\delta = 10^{-2}$, only one more active link is required at the sixth iteration compared to the cases of $\delta = 10^{-3}$ and 10^{-4}.

When the iterative link removal based method, i.e., Algorithm 2, is applied to the considered system, the stopping term is chosen to be $\epsilon = 10^{-4}$ and the maximum number of subgradient updates is set to be $N_{\mathrm{m}} = 80$. The bisection method is not applied here to better show the convergence results. Only one active link is removed here in each iteration. We choose the step size t_n in (23) to be $t_n = 0.01, 0.1$ and 1, respectively. Starting from $KB = 25$ links, seventeen links are removed and only eight links remain active in the end for the considered choices of the step size t_n. This result is slightly worse than the final result obtained by Algorithm 1. Fig. 3 shows that the required number of subgradient updates is sensitive to the choice of the step size t_n. A smaller step size will lead to a larger number of updates. If the step size is chosen to be 0.01, more than twelve updates are required for each iteration. However, when we choose the step size to be 0.1 or 1, only two to three iterations are required per iteration. On the other hand, the step size t_n cannot be chosen to be too big because a large t_n will affect the accuracy of the results. We find it advisable to set $0.1 < t_n < 1$ in our simulations to achieve a good balance in the accuracy and program speed.

B. *Cellular Network Simulations*

We consider a three-cell network as shown in Fig. 4, where each BS is equipped with multiple antennas. The inter-cell distance between neighboring BSs is 0.5km. The transmission power constraint at each BS is 30dBm. The transmit antenna gain at each BS is 5dB. The pathloss model from the BS to the MS is

$$L(\text{dB}) = 128 + 37.6 \cdot \log_{10} D, \tag{28}$$

where D is in the unit of km. The log-normal shadowing parameter is 10dB. The available transmission spectrum is 10MHz and the noise figure at each MS is 10dB. The power of the noise plus out-of-cooperating cell interference is set to be -83.98dBm. The users are randomly and uniformly distributed within a disk of 100m radius at the center of the triangle formed by the three BSs. We perform 50 channel realizations for each user location and 20 different user locations are chosen in each simulation. Only those channel realizations that can support the required SINR requirement are admitted. Table I shows a summary of the above simulation parameters. Five methods are compared: the reweighted ℓ_1-norm minimization based method ("Algo1"), the iterative link removal based method ("Algo2"), the channel strength based clustering (CSBC) method [47], full cooperation, and exhaustive search ("ExSearch"). CSBC is similar to Algo2, but it removes the links according to the channel strength. Full cooperation distributes all the user data to all the BSs. ExSearch has been discussed in Section II-B, which produces the optimum solution in minimizing the backhaul data transfer but requires enormous computations. We set the maximum number of iterations N_{max} for Algorithm 1 to be 20 and the relative error threshold ε to be 10^{-4}. The step size t_n for the dual uplink noise variance update in Algorithm 2 is set to be 0.5 and the stopping term $\epsilon = 10^{-4}$. The maximum number of subgradient updates is set to be $N_{\text{m}} = 80$. The SINR requirements for different MSs are the same, i.e., $\gamma_i = \gamma$, $\forall i \in \{1, \cdots, K\}$.

Fig. 5–Fig. 6 show the average number of active links and their corresponding average sum transmission power with respect to the number of antennas M at each BS. The number of MSs that are simultaneously served by those BSs is $K = 6$. The SINR requirement for each MS is set to be $\gamma = 10.3$dB, which is the reference SINR for 16QAM transmission with spectral efficiency of 2.41bps/Hz [25]. The average number of active links is shown in Fig. 5(a). Full cooperation requires all the $KB = 18$ links to be active irrespective of M. This induces a large backhaul signaling overhead. For the other four methods, the required active links decreases dramatically with the increase of M. When $M = 2$, Algo1 and Algo2 achieve almost the same results for the average active links, which is about 16.5. CSBC and ExSearch require one more and one less link on average, respectively. When $M = 3$, the required links for the four methods decrease to only 11, 12, 13 and 10.5, respectively. When $M \geq 6$, the number of required links for the three methods all reduces to about 6. That is the minimum required number of active links, which equals K. Compared to full cooperation, the proposed methods save about $1 - 6/18 = 66.7\%$ of the backhaul user data transfer. This is due to the fact that more antennas at the BSs provides more spatial degrees of freedom and proper choice of beamformers reduces the inter-user interference. The same trend can be observed for the average sum transmit power in Fig. 6. The sum transmit power decreases

with the increase of M. Full cooperation serves as a lower bound for the sum transmit power because all the other three methods perform partial cooperation. To our surprise, Algo1 requires both fewer active links in Fig. 5(a) and lower sum transmit power in Fig. 6 compared to Algo2 and CSBC when $3 \leq M \leq 8$. This shows that the sparsity pattern in the data routing matrix also affects the required sum transmit power.

Fig. 7–Fig. 8 show the results when the number of antennas per BS is fixed to $M = 4$ and the SINR requirement is set to $\gamma = 10.3$dB. This system can serve roughly up to $BM = 12$ single-antenna MSs. Due to the huge complexity of ExSearch, the simulation for ExSearch is only performed up to $K = 6$ users. Full cooperation shows a linear curve according to BK in Fig. 7(a). When $K \leq 4$, it is far below the system's serving capacity and the average number of active links all approximately equals K for Algo1, Algo2, CSBC and ExSearch. The number of required active links increases with K. When $K = 8$, the first three methods require about 15, 16 and 17 active links, respectively. In this case, Algo1 saves about $1 - 15/24 = 37.5\%$ of the backhaul user data transfer compared to full cooperation. When $K = 10$, the system capacity is almost reached. Algo1 and Algo2 require about 26 active links and CSBC requires 28 active links. About 13.3% of the backhaul user data transfer can be saved by Algo1. In Fig. 8, the curves of average sum power for Algo2 and CSBC are quite similar. Algo1 shows lower sum power especially when $4 \leq K \leq 8$.

The simulation results for a system of $K = 4$ users and each BS with $M = 4$ antennas is shown in Fig. 9–Fig. 10. The SINR requirement is shown as the x-axis. When the SINR requirement $\gamma \leq 8.1$dB, the number of active links for Algo1, Algo2, CSBC and ExSearch in Fig. 9(a) are all about 4, which equals K. As γ increases, the performance gaps between different methods become obvious. The gap between Algo2 and CSBC remains about 0.6 when $14.1\text{dB} \leq \gamma \leq 18.7$dB. Algo1 remains very close to ExSearch in the whole SINR range. When $\gamma = 18.7$dB, Algo1 requires about 8 active links, which is about 1.4 below that of Algo2. In this case, Algo1 saves about 33.3% of the backhaul user data transfer compared to full cooperation. The required sum transmit power for Algo1 is slightly below those of Algo2 and CSBC when $\gamma \leq 11.7$dB in Fig. 10. However, when $\gamma \geq 14.1$dB, the sum transmit power for Algo1 surpass those of the other two methods. The difference is about 0.3W when $\gamma = 18.7$dB.

VI. Conclusion

We considered the problem of minimizing the user data transmission in the backhaul subject to QoS and per-BS power constraints in the CoMP downlink transmission scenario. We showed that it can be formulated into an ℓ_0-norm minimization problem. However, such a problem is combinatorial and NP-hard. Furthermore, the solution of such a problem needs to jointly determine the BS choice for user data and the BS beamformers. Inspired by recent results in compressive sensing, we first proposed an approach that solves the problem using the ℓ_1-relaxation and the reweighted ℓ_1-norm minimization, which requires to solve a series of convex ℓ_1-norm minimization problems. Then, we proposed an iterative link removal based approach employing the projected subgradient updates and iterative fixed-point algorithms. We also showed that this method can be implemented in a semi-distributed manner under certain assumptions and analyzed the computational complexity of the proposed algorithms. Simulations showed that both the proposed methods can significantly reduce the user data distribution to BSs in the backhaul. In terms of backhaul user data transfer, the reweighted ℓ_1-norm minimization based method achieved near-optimum results in our simulations, which saves the backhaul user data transfer by 13%–67%. The iterative link removal based method is also an attractive choice due to its good performance, which is only slightly worse than the reweighted ℓ_1-norm minimization based method, and its low implementation complexity.

Appendix A

Proof of Proposition 1

We introduce additional slack variables $\mathbf{z}_i \in \mathbb{C}^K, \forall i \in \{1, \cdots, K\}$. The problem (7) can then be expressed in the following form

$$
\begin{aligned}
\min_{\{\mathbf{w}_{ij}\},\{s_{ij}\}} \quad & \textstyle\sum_{i,j} s_{ij} \\
\text{subject to} \quad & \|\mathbf{z}_i\|_2 \leq \tfrac{1}{\sqrt{\gamma_i}} \textstyle\sum_{j=1}^{B} \mathbf{h}_{ij}^H \mathbf{w}_{ij}, \\
& \|\mathbf{w}_{ij}\|_2 \leq s_{ij}, \\
& \textstyle\sum_{i=1}^{K} s_{ij}^2 \leq P_j, \\
& \left[\textstyle\sum_{j=1}^{B} \mathbf{h}_{ij}^H \mathbf{w}_{1j}, \cdots, \sum_{j=1}^{B} \mathbf{h}_{ij}^H \mathbf{w}_{(i-1)j}, \sigma, \sum_{j=1}^{B} \mathbf{h}_{ij}^H \mathbf{w}_{(i+1)j}, \cdots, \sum_{j=1}^{B} \mathbf{h}_{ij}^H \mathbf{w}_{Kj}\right]^T = \mathbf{z}_i, \\
& \forall i \in \{1, \cdots, K\}, \quad j \in \{1, \ldots, B\}
\end{aligned}
$$

Note that $\mathbf{z}_i$ is complex, except that its ith element is real. By introducing $\lambda_i, \rho_{ij}, \mu_j \in \mathbb{R}_+$ for the first three constraints and $\boldsymbol{\nu}_i \in \mathbb{C}^K$ for the fourth constraint, respectively, the Lagrangian of the above problem can be expressed as

$$\begin{aligned}\mathcal{L}\left(\{\mathbf{w}_{ij}\}, \{s_{ij}\}\right) = & \sum_{i,j} s_{ij} + \sum_i \left(\lambda_i \|\mathbf{z}_i\|_2 - \frac{\lambda_i}{\sqrt{\gamma_i}} \sum_{j=1}^{B} \mathbf{h}_{ij}^H \mathbf{w}_{ij}\right) + \sum_{i,j} \left(\rho_{ij} \|\mathbf{w}_{ij}\|_2 - \rho_{ij} s_{ij}\right) \\ & + \sum_i \left(\sum_{k \neq i} \left(\nu_{ik}^* \sum_{j=1}^{B} \mathbf{h}_{ij}^H \mathbf{w}_{kj}\right) + \nu_{ii}\sigma - \boldsymbol{\nu}_i^H \mathbf{z}_i\right) + \sum_j \mu_j \left(\sum_i s_{ij}^2 - P_j\right) \quad (29) \\ = & \sum_{i,j} \left(\mu_j s_{ij}^2 + (1-\rho_{ij}) s_{ij}\right) + \sum_i \left(\lambda_i \|\mathbf{z}_i\|_2 - \boldsymbol{\nu}_i^H \cdot \mathbf{z}_i\right) + \sum_{i,j} \rho_{ij} \|\mathbf{w}_{ij}\|_2 \\ & + \sum_i \left(-\frac{\lambda_i}{\sqrt{\gamma_i}} \sum_{j=1}^{B} \mathbf{h}_{ij}^H \mathbf{w}_{ij} + \sum_{k \neq i} \nu_{ki}^* \sum_{j=1}^{B} \mathbf{h}_{kj}^H \mathbf{w}_{ij}\right) + \sum_i \nu_{ii}\sigma - \sum_j \mu_j P_j. \quad (30)\end{aligned}$$

Here ν_{ki} denotes the ith element of $\boldsymbol{\nu}_k$. Furthermore, s_{ij} is chosen from the range $s_{ij} \geq 0$.

According to the complementary slackness conditions [28], the dual variable $\mu_j = 0$ when the power constraint at the jth BS is not satisfied with equality. In this case, we must have $\rho_{ij} \leq 1$ otherwise the term $(1-\rho_{ij})s_{ij}$ is unbounded below because $s_{ij} \geq 0$. Furthermore, if $\rho_{ij} < 1$, the Lagrangian $\mathcal{L}\left(\{\mathbf{w}_{ij}\}, \{s_{ij}\}\right)$ is minimized when $s_{ij} = 0$. When the power constraint at the jth BS is satisfied with equality, i.e., $\mu_j > 0$, we also have that $s_{ij} = 0$ minimizes the Lagrangian $\mathcal{L}\left(\{\mathbf{w}_{ij}\}, \{s_{ij}\}\right)$ if $\rho_{ij} < 1$. Therefore, the optimal point of s_{ij} is obtained at $s_{ij} = 0$ when $\rho_{ij} < 1$.

APPENDIX B
PROOF OF PROPOSITION 2

For a given nonnegative real vector $\mathbf{x} \in \mathbb{R}_+^M$, its ℓ_0-norm can be approximated as [31]

$$\|\mathbf{x}\|_0 = \lim_{\delta \to 0} \sum_{i=1}^{M} \frac{\log(1 + |x_i|\,\delta^{-1})}{\log(1 + \delta^{-1})} \quad (31)$$

If we choose δ to be sufficiently small and replace the objective function of (6) with the RHS of (31), we obtain the following problem by neglecting some constant numbers

$$\begin{aligned} \min_{\{\mathbf{w}_{ij}\},\{s_{ij}\}} \quad & \textstyle\sum_{i,j} \log(s_{ij} + \delta) \\ \text{subject to} \quad & \text{Constraints of (6).} \end{aligned} \quad (32)$$

The cost function is a log-sum function. It is concave and below its tangent. Specifically, let $\left\{\mathbf{w}_{ij}^{(n)}, s_{ij}^{(n)}\right\}$ denote the solution in the nth iteration. Due to the concavity of the $\log$ function and

$s_{ij} \geq 0$ for feasible solutions, the first-order approximation of $\log(s_{ij} + \delta)$ yields

$$\log(s_{ij} + \delta) \leq \log(s_{ij}^{(n)} + \delta) + \frac{1}{s_{ij}^{(n)} + \delta} \left(s_{ij} - s_{ij}^{(n)} \right). \tag{33}$$

The RHS of (33) majorizes the $\log$ function, and can be used as a surrogate function for the original cost function in (32). Specifically, replacing the RHS of (33) into (32), the solution we get in the $(n+1)$-th iteration is

$$\left\{ \mathbf{w}_{ij}^{(n+1)}, s_{ij}^{(n+1)} \right\} = \arg\min \sum_{i,j} \frac{s_{ij}}{s_{ij}^{(n)} + \delta}, \quad \text{subject to: Constraints of (6).} \tag{34}$$

We can see that the weighting factor of s_{ij} in the objective function of (34) is same to the $\beta_{ij}^{(n)}$ in Algorithm 1. By iteratively minimizing the surrogate function, local optimal solutions can be obtained, which is along the idea of the majorization-minimization method [48].

For $s_{ij}^{(n)}$ and $s_{ij}^{(n+1)}$, we have

$$\sum_{i,j} \log(s_{ij}^{(n+1)} + \delta) \leq \log(s_{ij}^{(n)} + \delta) + \frac{s_{ij}^{(n+1)} - s_{ij}^{(n)}}{s_{ij}^{(n)} + \delta} \leq \log(s_{ij}^{(n)} + \delta) + \frac{s_{ij}^{(n)} - s_{ij}^{(n)}}{s_{ij}^{(n)} + \delta} = \sum_{i,j} \log(s_{ij}^{(n)} + \delta).$$

The two inequalities follow from (33) and (34), respectively. So $\sum_{i,j} \log(s_{ij}^{(n)} + \delta)$ is a monotonically decreasing sequence, lower bounded by $KB \log(\delta)$. Therefore, the optimal objective value of (32) must converge.

Appendix C
Proof of Theorem 1

Concatenate the channel vectors and beamformers to the ith MS into vectors $\mathbf{h}_i$ and $\mathbf{w}_i$ as in Section IV-A. Furthermore, we introduce dual variables $\{q_i\}$ and $\{\nu_j\}$ for the SINR constraints and per-BS power constraints of (8), respectively. The Lagrangian for (8) is obtained as follows

$$\begin{aligned}
&\mathcal{L}(\{\mathbf{w}_i\}, \{q_i\}, \{\nu_j\}) \\
&= \sum_i \mathbf{w}_i^H \mathbf{w}_i + \sum_i q_i \left(\sigma^2 + \sum_{k \neq i}^{K} \left| \mathbf{h}_i^H \mathbf{w}_k \right|^2 - \frac{\left| \mathbf{h}_i^H \mathbf{w}_i \right|^2}{\gamma_i} \right) + \sum_j \nu_j \left(\sum_i \mathbf{w}_i^H \mathbf{Q}_j \mathbf{w}_i - P_j \right) \\
&= \sum_i \mathbf{w}_i^H \left(\mathbf{I} + \sum_{j=1}^{B} \nu_j \mathbf{Q}_j \right) \mathbf{w}_i + \sum_i \left(\sum_{k \neq i}^{K} q_k \left| \mathbf{w}_i^H \mathbf{h}_k \right|^2 - \frac{q_i \left| \mathbf{w}_i^H \mathbf{h}_i \right|^2}{\gamma_i} \right) + \sum_i q_i \sigma^2 - \sum_j \nu_j P_j
\end{aligned}$$

where $\{\mathbf{Q}_j\}$ is introduced in Section IV-A. The Lagrange dual function is obtained as

$$g(\{q_i\},\{\nu_j\}) = \inf_{\{\mathbf{w}_i\}} \mathcal{L}(\{\mathbf{w}_i\},\{q_i\},\{\nu_j\})$$
$$= \begin{cases} \sum_{i=1}^{K} q_i\sigma^2 - \sum_{j=1}^{B}\nu_j P_j, & \text{if } \left(\mathbf{I}+\sum_{j=1}^{B}\nu_j\mathbf{Q}_j\right)+\sum_{k\neq i}^{K} q_k\mathbf{h}_k\mathbf{h}_k^H \geq \frac{q_i\mathbf{h}_i\mathbf{h}_i^H}{\gamma_i}, \forall i \\ -\infty, & \text{otherwise.} \end{cases}$$

Therefore, the Lagrange dual problem can then be expressed as

$$\begin{aligned} \max_{\{\nu_j\},\{q_i\}} \quad & \sum_{i=1}^{K} q_i\sigma^2 - \sum_{j=1}^{B}\nu_j P_j \\ \text{subject to} \quad & \left(\mathbf{I}+\sum_{j=1}^{B}\nu_j\mathbf{Q}_j\right)+\sum_{k\neq i}^{K} q_k\mathbf{h}_k\mathbf{h}_k^H \geq \frac{q_i\mathbf{h}_i\mathbf{h}_i^H}{\gamma_i}, \quad \forall i \\ & \nu_j \geq 0,\ q_i \geq 0, \quad \forall i\in\{1,\ldots,K\}, j\in\{1,\ldots,B\}. \end{aligned} \tag{35}$$

Introducing the dual uplink receive beamformers $\{\hat{\mathbf{w}}_i\}$, the first constraint can be rewritten as [23]

$$\mathsf{SINR}_i^{\mathrm{U}} = \frac{q_i\left|\hat{\mathbf{w}}_i\mathbf{h}_i\right|^2}{\sum_{k\neq i}^{K} q_k\left|\hat{\mathbf{w}}_i\mathbf{h}_k\right|^2+\hat{\mathbf{w}}_i^H\left(\mathbf{I}+\sum_{j=1}^{B}\nu_j\mathbf{Q}_j\right)\hat{\mathbf{w}}_i} \leq \gamma_i \tag{36}$$

Furthermore, $\{\hat{\mathbf{w}}_i\}$ is proportional to $\{\mathbf{w}_i\}$, $\forall i\in\{1,\cdots,K\}$, at the optimum solution following similar discussions as in [23]. Since the SINR constraint must be attained with equality for the optimal solution, the inequality in (36) can be reversed, and the maximization with respect to $\{q_i,\hat{\mathbf{w}}_i\}$ can be replaced by the minimization, which gives the Lagrange dual problem of (9).

REFERENCES

[1] 3GPP TR 36.819, "Coordinated multi-point operation for LTE physical layer aspects," Dec. 2011, v11.1.0.

[2] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[3] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.

[4] 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.

[5] W. Ho, T. Q. S. Quek, S. Sun, and R. Heath Jr, "Decentralized precoding for multicell MIMO downlink," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1798–1809, Jun. 2011.

[6] D. W. H. Cai, T. Q. S. Quek, C. W. Tan, and S. H. Low, "Max-min SINR coordinated multipoint downlink transmission – Duality and algorithms," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5384–5395, Oct. 2012.

[7] M. Bengtsson, "Jointly optimal downlink beamforming and base station assignment," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Process.*, Salt Lake City, UT, May, 2001, pp. 2961–2964.

[8] S. Shamai and B. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. 53th IEEE Veh. Tech. Conf.*, Rhodes, Greece, May, 2001, pp. 1745–1749.

[9] J. Zhang, R. Chen, J. Andrews, A. Ghosh, and R. Heath, “Networked MIMO with clustered linear precoding,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.

[10] R. Zhang, “Cooperative multi-cell block diagonalization with per-base-station power constraints,” *IEEE J. Select. Areas Commun.*, vol. 28, no. 9, pp. 1435–1445, Dec. 2010.

[11] C. T. K. Ng and H. Huang, “Linear precoding in cooperative MIMO cellular networks with limited coordination clusters,” *IEEE J. Select. Areas Commun.*, vol. 28, no. 9, pp. 1446–1454, Dec. 2010.

[12] H. Zhang, L. Venturino, N. Prasad, P. Li, S. Rangarajan, and X. Wang, “Weighted sum-rate maximization in multi-cell networks via coordinated scheduling and discrete power control,” *IEEE J. Select. Areas Commun.*, vol. 29, no. 6, pp. 1214–1224, Jun. 2011.

[13] L. Venturino, N. Prasad, and X. Wang, “Coordinated scheduling and power allocation in downlink multicell OFDMA networks,” *IEEE Trans. Veh. Technol.*, vol. 58, no. 6, pp. 2835–2848, Jul. 2009.

[14] T. Bogale and L. Vandendorpe, “Weighted sum rate optimization for downlink multiuser MIMO coordinated base station systems: Centralized and distributed algorithms,” *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1876–1889, Apr. 2012.

[15] O. Simeone, O. Somekh, H. Poor, and S. Shamai, “Local base station cooperation via finite-capacity links for the uplink of linear cellular networks,” *IEEE Trans. Inform. Theory*, vol. 55, no. 1, pp. 190–204, Jan. 2009.

[16] A. Sanderovich, O. Somekh, H. Poor, and S. Shamai, “Uplink macro diversity of limited backhaul cellular network,” *IEEE Trans. Inform. Theory*, vol. 55, no. 8, pp. 3457–3478, Aug. 2009.

[17] P. Marsch and G. Fettweis, “Uplink CoMP under a constrained backhaul and imperfect channel knowledge,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1730–1742, Jun. 2011.

[18] J. Hoydis, M. Kobayashi, and M. Debbah, “Optimal channel training in uplink network MIMO systems,” *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2824–2833, Jun. 2011.

[19] R. Zakhour and D. Gesbert, “Optimized data sharing in multicell MIMO with finite backhaul capacity,” *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6102–6111, Dec. 2011.

[20] M. Gastpar, “The Wyner-Ziv problem with multiple sources,” *IEEE Trans. Inform. Theory*, vol. 50, no. 11, pp. 2762–2768, Nov. 2004.

[21] D. Donoho, “Compressed sensing,” *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[22] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, Feb. 2006.

[23] W. Yu and T. Lan, “Transmitter optimization for the multi-antenna downlink with per-antenna power constraints,” *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, Jun. 2007.

[24] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G evolution: HSPA and LTE for mobile broadband*. Academic Press, 2008.

[25] A. Ghosh and R. Ratasuk, *Essentials of LTE and LTE-A*. Cambridge University Press, 2011.

[26] B. Natarajan, “Sparse approximate solutions to linear systems,” *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.

[27] E. Matskani, N. Sidiropoulos, Z. Luo, and L. Tassiulas, “Convex approximation techniques for joint multiuser downlink beamforming and admission control,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2682–2693, Jul. 2008.

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

[29] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[30] E. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[31] B. Sriperumbudur, D. Torres, and G. Lanckriet, "A majorization-minimization approach to the sparse generalized eigenvalue problem," *Machine learning*, vol. 85, no. 1, pp. 3–39, Oct. 2011.

[32] M. Fazel, H. Hindi, and S. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," in *Proc. American Control Conference*, Denver, CO, Jun. 2003, pp. 2156–2162.

[33] D. W. H. Cai, T. Q. S. Quek, and C. W. Tan, "A unified analysis of max-min weighted SINR for MIMO downlink system," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3850–3862, Aug. 2011.

[34] F. Rashid-Farrokhi, L. Tassiulas, and K. Liu, "Joint optimal power control and beamforming in wireless networks using antenna arrays," *IEEE Trans. Commun.*, vol. 46, no. 10, pp. 1313–1324, Oct. 1998.

[35] H. Van Trees, *Optimum array processing*. John Wiley & Sons, 2002.

[36] A. Wiesel, Y. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Jan. 2006.

[37] R. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Select. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, Sep. 1995.

[38] N. Shor, K. Kiwiel, and A. Ruszczynski, *Minimization methods for non-differentiable functions*. Springer-Verlag Berlin, 1985.

[39] J. Goffin, "On convergence rates of subgradient optimization methods," *Math. Program.*, vol. 13, no. 1, pp. 329–347, 1977.

[40] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex analysis and optimization*. Athena Scientific, 2003.

[41] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Athena Scientific, 1999.

[42] G. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 641–646, Nov. 1993.

[43] Y. Ye, *Interior point algorithms: theory and analysis*. John Wiley & Sons, 1997.

[44] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 1986.

[45] H. R. Feyzmahdavian, M. Johansson, and T. Charalambous, "Contractive interference functions and rates of convergence of distributed power control laws," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4494–4502, Dec. 2012.

[46] C. H. Papadimitriou and K. Steiglitz, *Combinatorial optimization: algorithms and complexity*. Prentice-Hall, Inc., 1982.

[47] J. Zhao and Z. Lei, "Clustering methods for base station cooperation," in *Proc. IEEE Wirel. Comm. and Netw. Conf.*, Paris, France, Apr., 2012, pp. 1–6.

[48] D. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

TABLE I
SUMMARY OF SIMULATION PARAMETERS

Number of cells	3
Inter-cell distance	0.5km
BS transmission power constraint	30dBm
BS transmit antenna gain	5dB
Log-normal shadowing	10dB
Transmission spectrum	10MHz
Noise variance plus out-of-cooperating cell interference	-83.98dBm
Disk radius for MS location	100m
Channel realizations for each MS location	50
MS location	20
MS location distribution	uniform
Number of antennas per MS	1

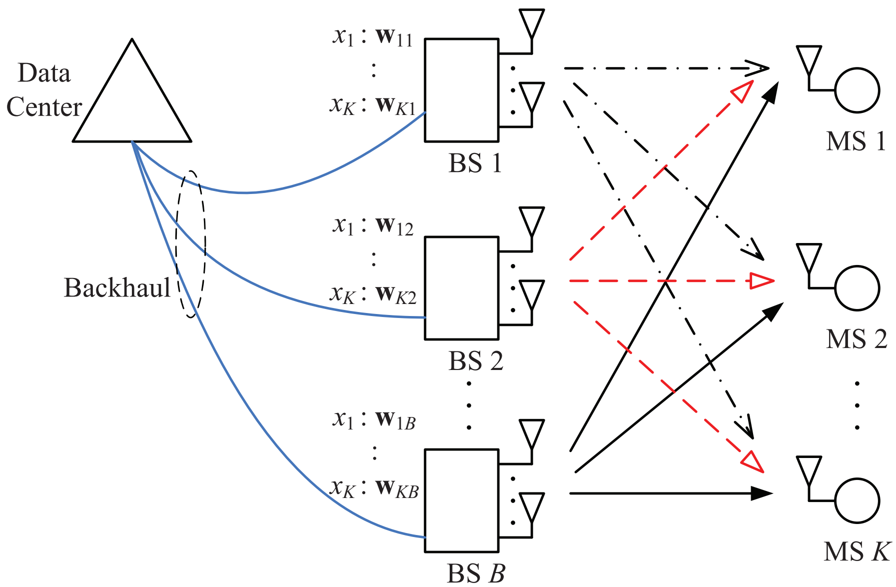


Fig. 1. CoMP downlink transmission scenario

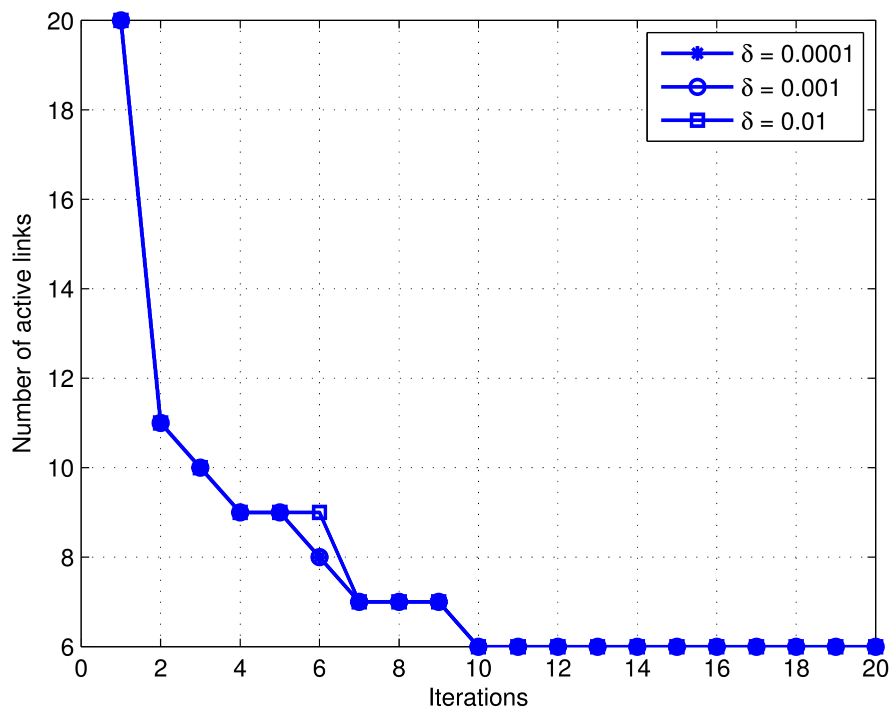


Fig. 2. Convergence behavior for the reweighted ℓ_1-norm minimization based method, $B = K = 5$. $M = 3$ antennas at each BS, the SINR requirement $\gamma_i = 6.02$dB, $\forall i \in \{1, \cdots, K\}$.

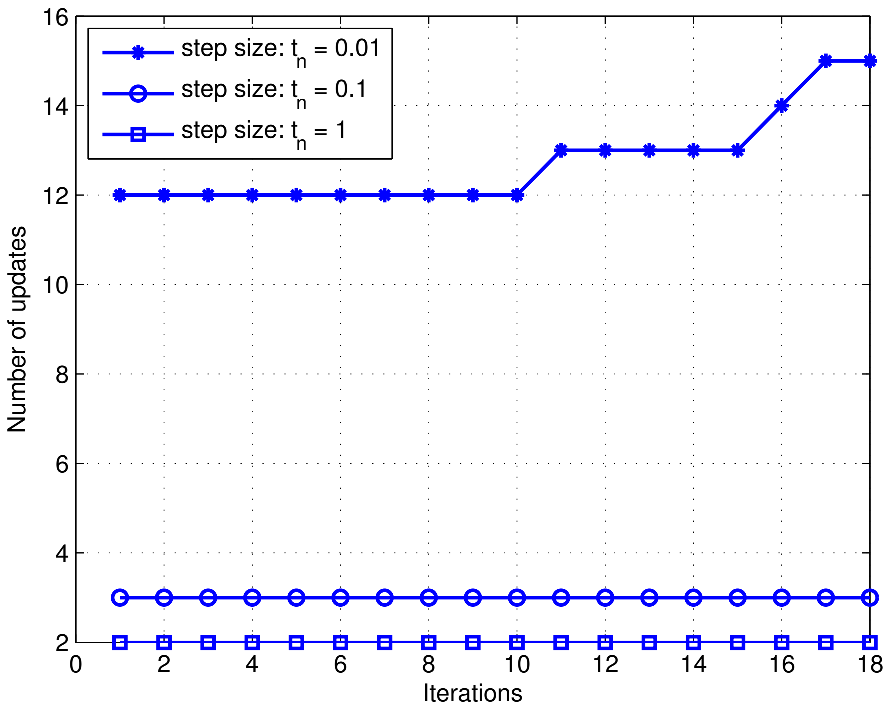


Fig. 3. Number of subgradient updates for dual noise $\{\nu_j\}$ in each iteration for the iterative link removal based method, $B = K = 5$. $M = 3$ antennas at each BS, the SINR requirement $\gamma_i = 6.02$dB, $\forall i \in \{1, \cdots, K\}$. The stopping term $\epsilon = 10^{-4}$.

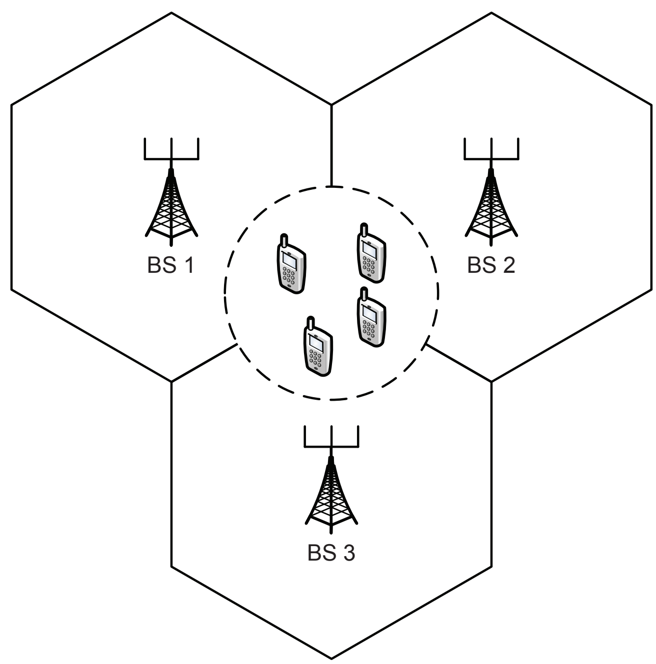
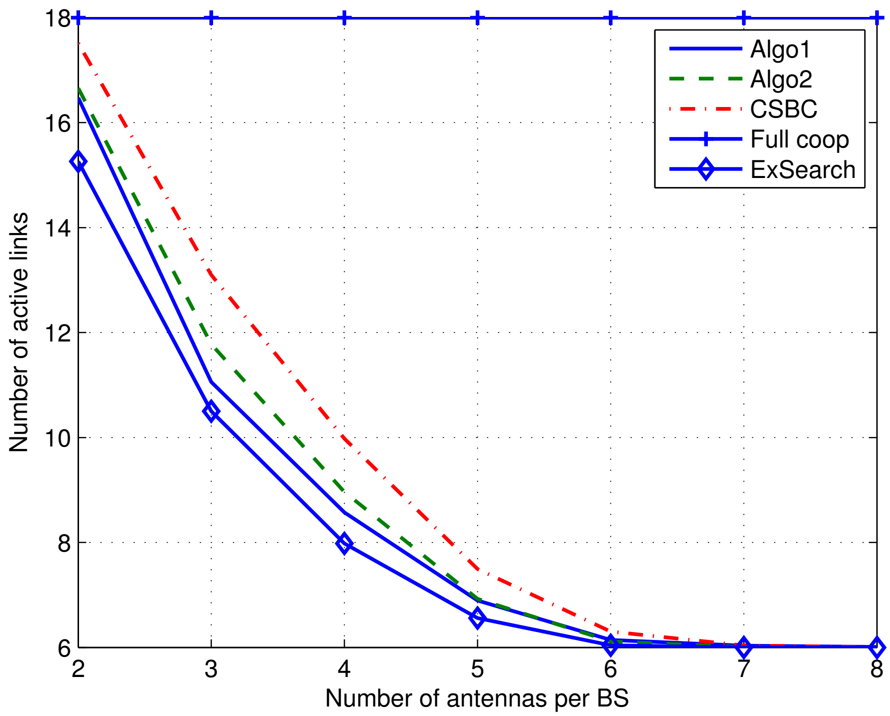
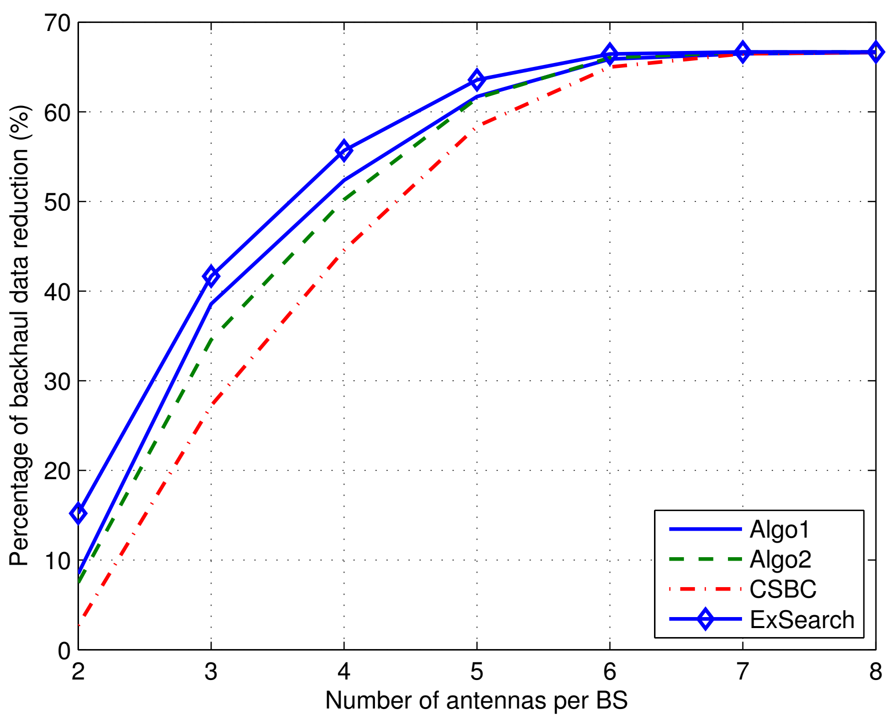


Fig. 4. Cellular network simulation scenario



(a) Number of transmission links



(b) Percentage of backhaul data transfer reduction compared to full cooperation

Fig. 5. Simulation results for $B = 3$ BSs, $K = 6$ MSs, SINR requirement $\gamma = 10.3$dB, where $\gamma_i = \gamma, \forall i \in \{1, \cdots, K\}$. Number of antennas M at each BS changes.

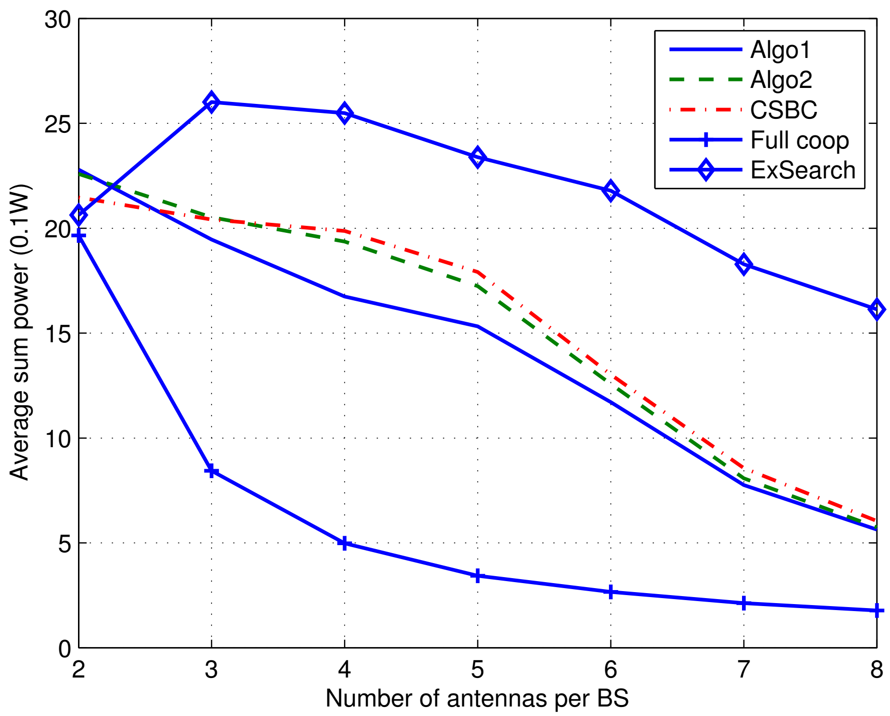
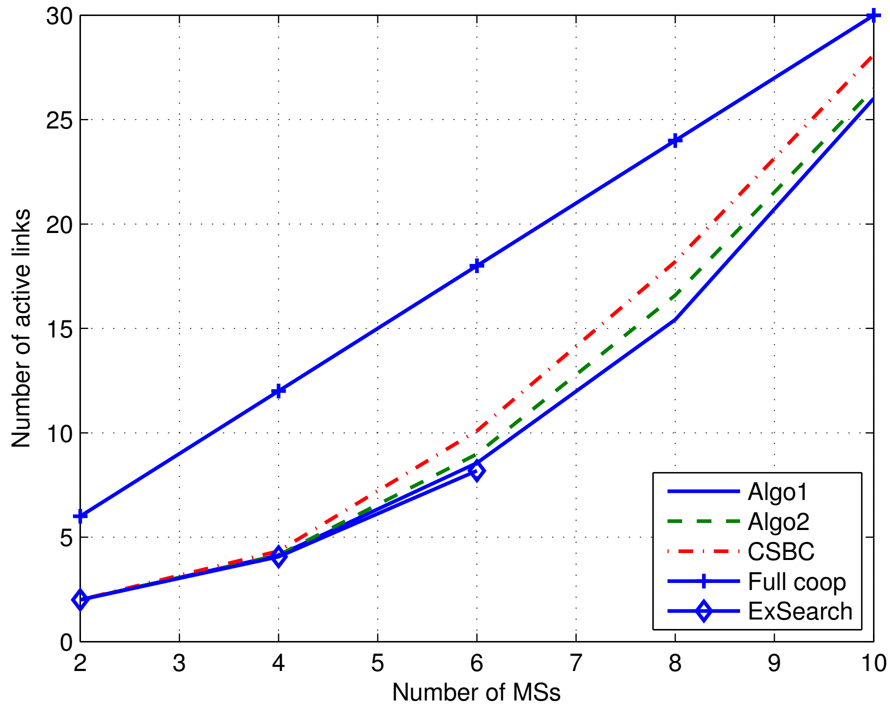
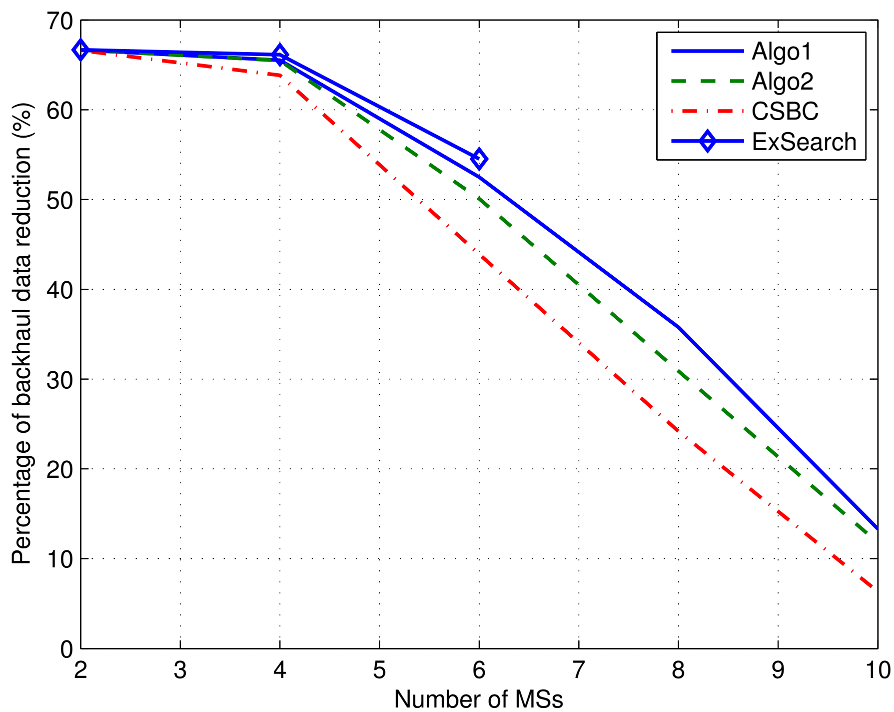


Fig. 6. Sum transmission power (in 0.1W unit) for $B = 3$ BSs, $K = 6$ MSs, SINR requirement $\gamma = 10.3$dB, where $\gamma_i = \gamma, \forall i \in \{1, \cdots, K\}$. Number of antennas M at each BS changes.



(a) Number of transmission links



(b) Percentage of backhaul data transfer reduction compared to full cooperation

Fig. 7. Simulation results for $B = 3$ BSs, $M = 4$ antennas per BS, SINR requirement $\gamma = 10.3$dB. Number of users K changes.

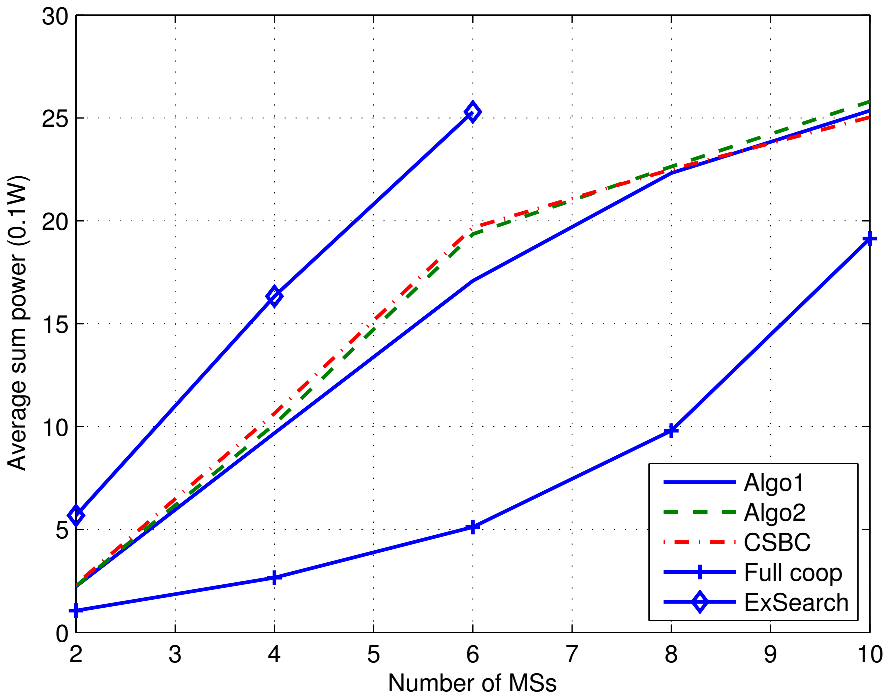
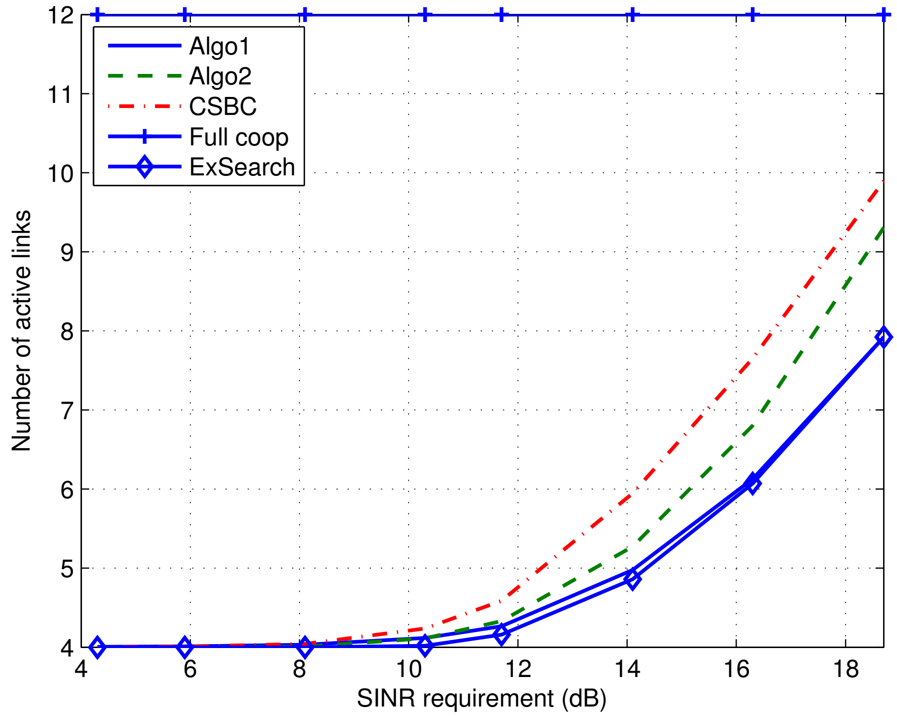
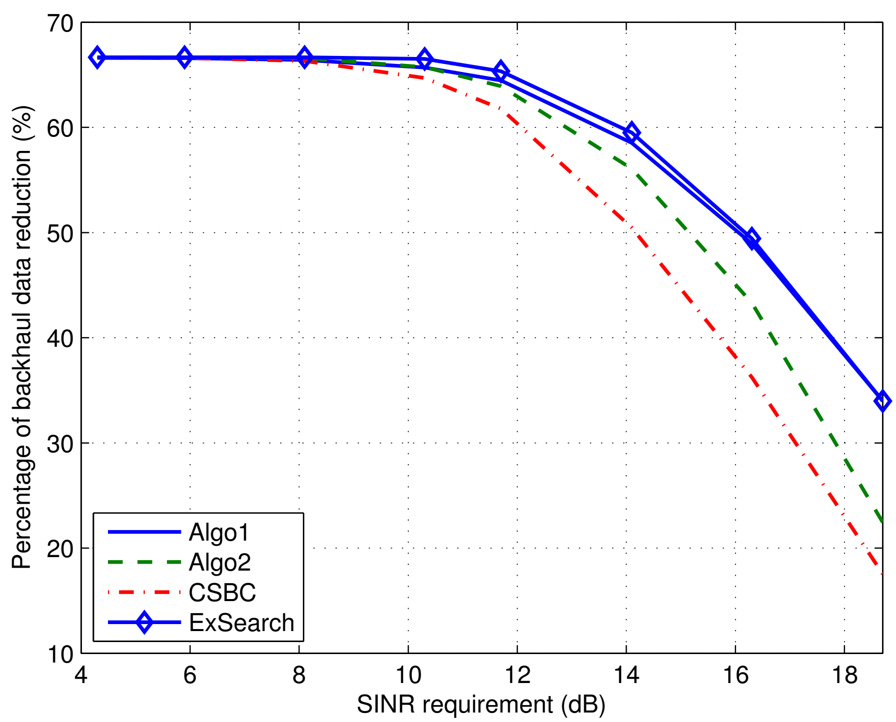


Fig. 8. Sum transmission power (in 0.1W unit) for $B = 3$ BSs, $M = 4$ antennas per BS, SINR requirement $\gamma = 10.3$dB. Number of users K changes.



(a) Number of transmission links



(b) Percentage of backhaul data transfer reduction compared to full cooperation

Fig. 9. Simulation results for $B = 3$ BSs, $K = 4$ MSs, $M = 4$ antennas per BS. SINR requirement γ at each MS changes.

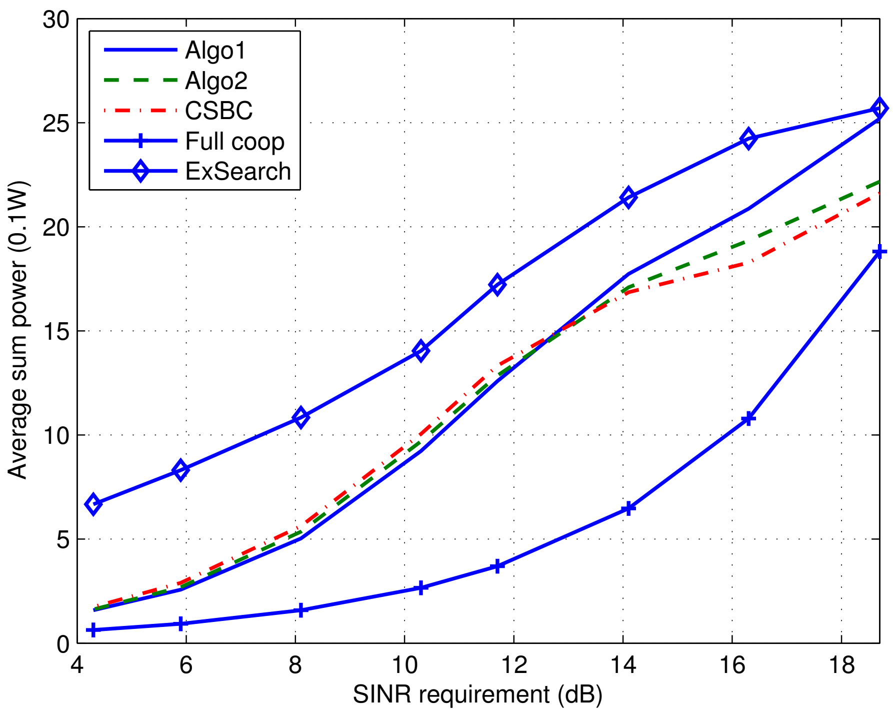


Fig. 10. Sum transmission power (in 0.1W unit) for $B = 3$ BSs, $K = 4$ MSs, $M = 4$ antennas per BS. SINR requirement γ at each MS changes.