BACKHAUL-CONSTRAINED MULTI-CELL COOPERATION USING COMPRESSIVE SENSING AND SPECTRAL CLUSTERING

Seung-Jun Kim, Swayambhoo Jain, and Georgios B. Giannakis

Dept. of ECE, University of Minnesota, Minneapolis, MN 55455, U.S.A.
E-mails: {seungjun,jainx174,georgios}@umn.edu

ABSTRACT

Multi-cell cooperative processing with limited backhaul traffic is considered for cellular uplinks. To parsimoniously select a set of cooperating base stations, a *sparse* multi-cell receive-filter is obtained through convex optimization using compressive sensing techniques. Clustered cooperation is also considered, where sparsity is promoted on inter-cluster feedback. A joint equalizer design and dynamic partitioning problem is formulated and solved using an iterative spectral clustering approach. Numerical tests verify the efficacy of proposed methods.

1. INTRODUCTION

Multi-cell cooperative processing (MCP) is a promising approach to mitigate inter-cell interference in cellular networks. Exploiting the increasingly high-speed backhaul networks that connect base stations (BSs), MCP allows cooperation of antennas distributed at different cell sites, thus creating a virtual networked MIMO system. As a result, inter-cell interference can be used constructively to better decode mobile stations (MSs) and enhance fairness, particularly for MSs residing at the cell edge [1, 2, 3]. Since the major processing burden is still at the BSs, complexity increase in MSs is minimal when MCP is in effect.

Information-theoretic studies in the uplink have revealed that MCP can effectively mitigate the capacity loss due to inter-cell interference, under a full cooperation model, where all BSs collaborate via an infinite-capacity backhaul link [4]. Downlink MCP with full cooperation has also been considered [5].

In practice, full cooperation is not feasible for large networks, due to excessive backhaul traffic [6]. Thus, the set of cooperating BSs must be chosen parsimoniously based on their contribution to system objectives. Moreover, in order to reduce the signaling overhead associated with backhauling, it is often prudent to form clusters of BSs. While clustering may be dictated by the network architecture, adaptive dynamic clustering based on channels estimates and MS locations is beneficial [3].

Direct sample exchanges between BSs were studied in the context of 3GPP LTE systems in [7], where distributed cooperation without central control was advocated. A greedy algorithm for dynamic clustering was proposed in [8] to maximize the uplink sum-rate. A clustered cooperation scheme with linear processing in the downlink was put forth in [9].

The contribution of this work is two-fold. First, an uplink distributed MCP scheme with limited backhaul traffic is proposed, where the BSs collect feedback from a small number of cooperating BSs. To judiciously select the relevant BSs to cooperate, recent developments in compressive sensing are leveraged to obtain a sparse linear receive-filter [10]. The approach is then extended to clustered scenarios, where clusters of BSs are formed dynamically based on channels and system objectives using a spectral clustering framework [11].

The rest of the paper is organized as follows. The system model is presented in Sec. 2. Sec. 3 develops backhaul-constrained distributed cooperation. Sec. 4 presents static and dynamic clustered cooperation. Numerical tests are performed in Sec. 5. Conclusions are provided in Sec. 6.

2. SYSTEM MODEL

2.1. Uplink Signal Model

Consider a multi-cell network uplink with N_B BSs and N_U MSs. Each BS $b \in \mathcal{B} := \{1, 2, \ldots, N_B\}$ is assumed to be equipped with A antennas. Thus, the total number of BS antennas in the network is $N_A := AN_B$. The set of antennas that belong to BS b is denoted as $\mathcal{A}_b$. Each MS user $u \in \mathcal{U} := \{1, 2, \ldots, N_U\}$ possesses a single antenna. The serving BS of the u-th MS is denoted as $b(u)$. The set of MSs served by BS b is denoted as $\mathcal{U}_b := \{u \in \mathcal{U} : b(u) = b\}$.

Let x_u denote the symbol transmitted by the u-th MS in a particular time slot and (sub-)band. Define $\mathbf{x} := [x_1, x_2, \ldots, x_{N_U}]^T$, where ^{T} stands for transposition. Similarly, the signal samples received by the N_A BS antennas are represented as $\mathbf{y} := [y_1, y_2, \ldots, y_{N_A}]^T$, where y_a for $a \in \mathcal{A} := \{1, 2, \ldots, N_A\}$ denotes the sample taken by the a-th antenna. The flat-fading channel matrix $\mathbf{H} \in \mathbb{C}^{N_A \times N_U}$ has its (a, u)-entry equal to the complex channel coefficient from the u-th MS to the a-th BS antenna. Thus, the input-output relation-

ship for the uplink can be expressed compactly as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{n}$ represents additive noise and possible interference from outside the network. Vectors $\mathbf{x}$ and $\mathbf{n}$ are assumed to be zero-mean complex with covariance $\mathbf{I}$, which can be justified through prewhitening and normalization, as well as by absorbing MS transmit-powers into $\mathbf{H}$. It is also assumed that $\mathbf{x}$ and $\mathbf{n}$ are uncorrelated.

With MCP, multiple BSs cooperate to decode $\mathbf{x}$. For simplicity of implementation, linear processing is considered. If all N_A BS antennas in the network could fully cooperate (which is unrealistic as it would require prohibitive amount of backhaul traffic), an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ could be obtained using a linear equalizer $\mathbf{W} \in \mathbb{C}^{N_U \times N_A}$ to obtain $\hat{\mathbf{x}} = \mathbf{W}\mathbf{y}$, where $\hat{\mathbf{x}}_u := [\hat{x}_1, \ldots, \hat{x}_{N_U}]^T$. Thus, in order to obtain the estimate $\hat{x}_u$ of x_u of MS $u \in \mathcal{U}$, BS $b(u)$ would need to collect samples received by other BS antennas, and compute $\mathbf{w}_u^T \mathbf{y}$, where $\mathbf{w}_u^T$ denotes the u-th row of $\mathbf{W}$.

A widely used linear equalizer, also adopted here, is the linear minimum mean-square error (LMMSE) one, given by

$$\mathbf{W}_{\mathrm{lmmse}} = (\mathbf{H}^{\mathcal{H}}\mathbf{H} + \mathbf{I})^{-1}\mathbf{H}^{\mathcal{H}} \tag{2}$$

where $^{\mathcal{H}}$ denotes Hermitian transposition.

2.2. Multi-Cell Cooperation Model

2.2.1. Distributed Cooperation

In the simplest cooperation model, henceforth termed distributed cooperation, each BS collects from cooperating BSs the received samples necessary for decoding its served MSs [7]. The BSs share the samples over the backhaul network. If unlimited backhaul traffic were allowed, every BS could collect samples from all other BSs to perform full cooperation. Since the backhaul capacity is limited in practice, each BS must judiciously choose the set of BSs whose samples are most helpful in decoding its intended MSs.

Given the equalizer matrix $\mathbf{W}$, the amount of backhaul traffic can be deduced. For example, in the simplest case of $A = 1$, assume that $N_B = N_U$ with $b(u) = u$ for all $u \in \mathcal{U}$, and that the b-th antenna belongs to BS b; that is $\mathcal{A}_b = \{b\}$ for $b \in \mathcal{B}$. Then, under the distributed cooperation model, the amount of backhaul traffic is proportional to the number of nonzero off-diagonal entries of $\mathbf{W}$. General cases will be delineated in Sec. 3.2.

2.2.2. Clustered Cooperation

In clustered cooperation, BSs are partitioned into a number of clusters, and the received samples are pooled by the cluster head in each cluster so the MSs served by the cluster are decoded jointly [8]. In some cases, natural clustering could emerge by the particular (e.g. hierarchical) deployment architecture of cellular networks. In other cases, dynamic formation and adaptation of clusters may be beneficial. In the latter case, clusters themselves can be optimized based on channel gains or traffic patterns [8]. In any case, inter-cluster backhaul traffic incurs much higher burden to the backhaul network than intra-cluster traffic. This is due to the network's inherent hierarchical architecture, or to significant signaling overhead required for inter-cluster backhauling.

3. DISTRIBUTED COOPERATION

3.1. Sparse LMMSE Equalizer

For simplicity of exposition, consider first the case with $A = 1$, $N_B = N_U$, $\mathcal{U}_b = \{b\}$ and $\mathcal{A}_b = \{b\}$ for all $b \in \mathcal{B}$, as in Sec. 2.2.1. Since the amount of backhaul traffic is proportional to the number of nonzero off-diagonal entries of $\mathbf{W}$, a natural approach to obtain a backhaul-constrained MCP solution is to promote sparsity in the off-diagonal entries of $\mathbf{W}$. Inspired by recent advances in compressive sensing techniques, the idea here is to optimize the cost functions related to the MMSE criterion, augmented with sparsity-promoting regularization terms.

Let $\tilde{\mathbf{W}} \in \mathbb{R}^{N_B \times N_B}$ denote the matrix with zero diagonal entries, and the off-diagonal entries equal to the absolute values of the corresponding entries in $\mathbf{W}$. The LMMSE equalizer seeks to minimize the mean-square error (MSE) $\mathbb{E}\{||\hat{\mathbf{x}} - \mathbf{x}||_2^2\} = ||\mathbf{I} - \mathbf{W}\mathbf{H}||_F^2 + ||\mathbf{W}||_F^2$. Thus, one can penalize the ℓ_0-"norm" of $\tilde{\mathbf{W}}$ as

$$\min_{\mathbf{W}} \|\mathbf{I} - \mathbf{W}\mathbf{H}\|_F^2 + \|\mathbf{W}\|_F^2 + \lambda\|\tilde{\mathbf{W}}\|_0 \tag{3}$$

where $\|\tilde{\mathbf{W}}\|_0$ denotes the total number of nonzero entries in $\tilde{\mathbf{W}}$, $\|\cdot\|_F$ the Frobenius norm, and λ controls the amount of backhaul traffic incurred by the solution.

Since the criterion in (3) is nonconvex, to obtain an efficient solution, the ℓ_0-"norm" can be replaced by the ℓ_1-norm, which offers the tightest convex relaxation. Thus, the relevant optimization problem is

$$\min_{\mathbf{W}} \left\|\bar{\mathbf{I}} - \mathbf{W}\bar{\mathbf{H}}\right\|_F^2 + \lambda\|\tilde{\mathbf{W}}\|_1 \tag{4}$$

where $\bar{\mathbf{I}} := \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix}$, $\bar{\mathbf{H}} := \begin{bmatrix} \mathbf{H} & \mathbf{I} \end{bmatrix}$, and $\|\tilde{\mathbf{W}}\|_1$ is the sum of the absolute values of all entries of $\tilde{\mathbf{W}}$.

Remark 1 It is expected that the receive-filter for multi-cell processing will have an inherent sparsity structure, since the filter coefficients applied to the samples from the BSs whose channel gains to the desired MSs are small should also be small. Due to the path loss effect, the uplink channel gains to faraway BSs will be typically small. Therefore, it is well justified to use the sparsity-promoting ℓ_0-norm penalty to design a sparse receive-filter with minimal performance degradation compared to the optimal LMMSE solution.

3.2. Multi-Antenna Case Using Group Sparsity

With $A > 1$, each BS may well serve more than one MS. When estimating the amount of backhaul traffic in the case of multi-antenna cell sites, one must note that once BS b has collected samples from a cooperating BS b' to decode a user $u \in \mathcal{U}_b$, no additional backhaul traffic is incurred even if the samples are used for decoding other MSs $u' \neq u$ that are in $\mathcal{U}_b$. In other words, once the (u, a')-entry of $\mathbf{W}$ is determined to be nonzero, where $u \in \mathcal{U}_b$ and $a' \notin \mathcal{A}_b$, one can improve the decoding performance without incurring additional backhaul traffic by allowing (u', a')-entries of $\mathbf{W}$ to be nonzero for $u' \in \mathcal{U}_b$. Such a structure can be effected using regularization terms promoting *group sparsity* [12].

There is another avenue where group sparsity is useful. When a signal sample received from antenna $a \in \mathcal{A}_b$ is collected, it may be desirable to collect together the samples from all the antennas in $\mathcal{A}_b$ associated with that BS in order to economize the overhead associated with identification of the cooperating BS.

Under these considerations, one can replace the terms in (4) with

$$\lambda \sum_{b,b' \in \mathcal{B}: b \neq b'} \sqrt{\sum_{u \in \mathcal{U}_b} \sum_{a \in \mathcal{A}_{b'}} |w_{ua}|^2} \qquad (5)$$

where w_{ua} is the (u, a)-entry of $\mathbf{W}$. In other words, one can define $\tilde{\mathbf{W}} \in \mathbb{R}^{N_B \times N_B}$ with (b, b')-entry $\tilde{w}_{bb'}$ for $b, b' \in \mathcal{B}$ given as

$$\tilde{w}_{bb'} = \begin{cases} \sqrt{\sum_{u \in \mathcal{U}_b} \sum_{a \in \mathcal{A}_{b'}} |w_{ua}|^2} & \text{if } b \neq b' \\ 0 & \text{otherwise.} \end{cases} \qquad (6)$$

Then, $\|\tilde{\mathbf{W}}\|_0$ represents the amount of backhaul traffic, and the relaxed problem (4) can be solved to obtain a sparse LMMSE equalizer efficiently. Note that (6) coincides with the definition of $\tilde{\mathbf{W}}$ in Sec. 3.1 when $A = 1$, $N_B = N_U$, $\mathcal{U}_b = \{b\}$ and $\mathcal{A}_b = \{b\} \, \forall b \in \mathcal{B}$.

4. CLUSTERED COOPERATION

4.1. Static Clustering

In the clustered cooperation scenario, the set of BSs is partitioned into N_C clusters $\{\mathcal{B}_c\}_{c=1}^{N_C}$, where

$$\bigcup_{c=1}^{N_C} \mathcal{B}_c = \mathcal{B} \quad \text{and} \quad \mathcal{B}_c \cap \mathcal{B}_{c'} = \varnothing \text{ for } c \neq c'. \qquad (7)$$

The inter-cluster backhaul traffic is proportional to the number of nonzero (b, b')-entries of $\tilde{\mathbf{W}}$, where b and b' belong to different clusters, namely, $b \in \mathcal{B}_c$ and $b' \in \mathcal{B}_{c'}$ with $c \neq c'$. By relaxing this cardinality number using the ℓ_1-norm, the corresponding regularizer is

$$\lambda \sum_{(c,c'):c \neq c'} s(\mathcal{B}_c, \mathcal{B}_{c'}) \qquad (8)$$

Table 1. Joint multi-cell processing and clustering algorithm.

Input: λ, number of clusters N_C, initial equalizer $\mathbf{W}_0$
Output: $\{\mathcal{B}_c\}_{c=1}^{N_C}$, $\mathbf{W}$
1: Initialize $\mathbf{W} \leftarrow \mathbf{W}_0$
2: Repeat:
3:     Compute $\tilde{\mathbf{W}}$ per (6).
4:     Set $\mathbf{D} \leftarrow \mathrm{diag}(\tilde{\mathbf{W}}\mathbf{1})$ ($\mathbf{1}$ is the all-one vector)
5:     Set $\mathbf{L} \leftarrow \mathbf{D} - \tilde{\mathbf{W}}$ and $\tilde{\mathbf{L}} \leftarrow (\mathbf{L} + \mathbf{L}^T)/2$
6:     Let ϕ_c be the eigenvector of $\tilde{\mathbf{L}}$ corresponding to the c-th smallest eigenvalue for $c = 1, \ldots, N_C$.
7:     Run the k-means algorithm on the rows of $\mathbf{\Phi} := [\phi_1, \ldots, \phi_{N_C}]$ to obtain clusters $\{\breve{\mathcal{B}}_c\}$.
8:     $\breve{\mathbf{W}} = \arg\min_{\mathbf{W}} \|\bar{\mathbf{I}} - \mathbf{W}\bar{\mathbf{H}}\|_F^2 + \lambda \cdot \mathrm{rcut}(\{\breve{\mathcal{B}}_c\})$
9:     If the objective in line 8 did not decrease, stop.
10:     Set $\mathbf{W} \leftarrow \breve{\mathbf{W}}$ and $\{\mathcal{B}_c\} \leftarrow \{\breve{\mathcal{B}}_c\}$.

where $s(\mathcal{B}_c, \mathcal{B}_{c'}) := \sum_{b \in \mathcal{B}_c} \sum_{b' \in \mathcal{B}_{c'}} \tilde{w}_{bb'}$. Note that in the special case of singleton clusters, i.e., if $N_C = N_B$ and $\mathcal{B}_b = \{b\}$ for all $b \in \mathcal{B}$, then (8) reduces to (5).

It is useful to view the backhaul penalty in (8) from the perspective of graph cuts. Consider a directed graph with the set of vertices comprising the BSs $\mathcal{B}$, and the set of directed edges $\mathcal{E} := \{(b, b') : b, b' \in \mathcal{B}\}$ connecting the vertices. The edge weight associated with edge (b, b') is given by $\tilde{w}_{bb'}$ for $b, b' \in \mathcal{B}$; i.e., $\tilde{\mathbf{W}}$ is the affinity matrix. Then, $s(\mathcal{B}_c, \mathcal{B}_{c'})$ is the sum of edge weights that connect the vertices in $\mathcal{B}_c$ to the vertices in $\mathcal{B}_{c'}$. Thus, (8) is proportional to the graph cut on $\mathcal{G} := (\mathcal{B}, \mathcal{E}, \tilde{\mathbf{W}})$ induced by the set of clusters $\{\mathcal{B}_c\}$, defined as

$$\mathrm{cut}(\{\mathcal{B}_c\}) := \sum_{c=1}^{N_C} s(\mathcal{B}_c, \overline{\mathcal{B}}_c) = \sum_{(c,c'):c \neq c'} s(\mathcal{B}_c, \mathcal{B}_{c'}) \qquad (9)$$

where $\overline{\mathcal{B}}_c := \cup_{c' \neq c} \mathcal{B}_{c'}$. Thus, the static clustering MCP solves

$$\min_{\mathbf{W}} \|\bar{\mathbf{I}} - \mathbf{W}\bar{\mathbf{H}}\|_F^2 + \lambda \cdot \mathrm{cut}(\{\mathcal{B}_c\}) \qquad (10)$$

which is a convex optimization problem.

4.2. Dynamic Partitioning Using Spectral Clustering

When the clusters are *not* given a priori, it is of interest to determine the best partitioning with limited inter-cluster backhaul traffic. However, finding the best $\{\mathcal{B}_c\}$ using the regularizer (8) leads to clusters of unbalanced sizes. For instance, taking $\mathcal{B}_1 = \mathcal{B}$ and $\mathcal{B}_c = \varnothing$ for $c = 2, \ldots, N_C$ incurs zero inter-cluster traffic. Obviously, this is undesirable, as the computational burden of large clusters may become excessive and the intra-cluster backhaul traffic also become significant.

To obtain clusters with balanced sizes, one may consider the *ratio cut* cost used in spectral graph partitioning [11]:

$$\mathrm{rcut}(\{\mathcal{B}_c\}) := \sum_{c=1}^{N_C} \frac{s(\mathcal{B}_c, \overline{\mathcal{B}}_c)}{|\mathcal{B}_c|} \qquad (11)$$

where $|\cdot|$ denotes set cardinality. Thus, a joint sparse LMMSE equalization and clustering problem is given by

$$\min_{\mathbf{W},\{\mathcal{B}_c\}_{c=1}^{N_C}} \|\bar{\mathbf{I}} - \mathbf{W}\bar{\mathbf{H}}\|_F^2 + \lambda \cdot \mathrm{rcut}(\{\mathcal{B}_c\}) \qquad (12)$$

where $\{\mathcal{B}_c\}$ must satisfy (7).

Even with fixed $\mathbf{W}$, problem (12) is known to be NP-complete. To obtain an approximate solution, it is first noted that with fixed $\mathbf{W}$, optimization w.r.t. $\{\mathcal{B}_c\}$ is precisely the minimum cut problem on graph $\mathcal{G}$, whose relaxation is what spectral clustering algorithms aim to solve [11, 13]. On the other hand, given $\{\mathcal{B}_c\}$, optimization w.r.t. $\mathbf{W}$ is a convex problem, as noted in Sec. 4.1. Thus, an iterative approach is pursued, where one alternately optimizes over $\mathbf{W}$ and $\{\mathcal{B}_c\}$ until no more reduction in the cost is obtained.

Spectral clustering has been mostly applied to undirected graphs [11]. To perform spectral clustering on directed graphs, some care must be taken [13]. Define matrix $\mathbf{D} \in \mathbb{R}^{N_B \times N_B}$ to be a diagonal matrix with b-th diagonal entry $\sum_{b' \in \mathcal{B}} \tilde{w}_{bb'}$. Define also $\mathbf{L} := \mathbf{D} - \tilde{\mathbf{W}}$. Consider an N_B-by-N_C cluster indicator matrix $\mathbf{\Phi}$ whose (b, c)-entry is defined as

$$\phi_{bc} = \begin{cases} \frac{1}{\sqrt{|\mathcal{B}_c|}} & \text{if } b \in \mathcal{B}_c \\ 0 & \text{otherwise} \end{cases} \qquad (13)$$

It can be seen that $\mathbf{\Phi}^T \mathbf{\Phi} = \mathbf{I}$. Then, it follows that

$$\mathrm{rcut}(\{\mathcal{B}_c\}) = \mathrm{tr}(\mathbf{\Phi}^T \mathbf{L} \mathbf{\Phi}) \qquad (14)$$

Thus, a minimum cut graph clustering problem based on the ratio cut objective can be stated as

$$\min_{\mathbf{\Phi}} \mathrm{tr}(\mathbf{\Phi}^T \mathbf{L} \mathbf{\Phi}) \text{ subject to } \mathbf{\Phi}^T \mathbf{\Phi} = \mathbf{I} \text{ and (13).} \qquad (15)$$

Spectral clustering drops constraint (13) such that $\mathbf{\Phi}$ takes real-valued entries. However, since $\mathbf{L}$ is not symmetric, one cannot apply the Rayleigh-Ritz theorem to the resulting relaxed problem directly. Noting that $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{A}^T)$, one can obtain an equivalent formulation as

$$\min_{\mathbf{\Phi}} \mathrm{tr}(\mathbf{\Phi}^T \tilde{\mathbf{L}} \mathbf{\Phi}) \text{ subject to } \mathbf{\Phi}^T \mathbf{\Phi} = \mathbf{I} \qquad (16)$$

where $\tilde{\mathbf{L}} := (\mathbf{L} + \mathbf{L}^T)/2$ is symmetric. The optimal solution to (16) is given by setting the columns of $\mathbf{\Phi}$ as the eigenvectors of $\tilde{\mathbf{L}}$ corresponding to the N_C smallest eigenvalues. Clusters $\{\mathcal{B}_c\}$ can then be found by running the k-means algorithm on the rows of $\mathbf{\Phi}$. The overall algorithm is listed in Table 1.

Remark 2 For distributed cooperation, computation of the sparse filter can be performed in a distributed fashion at each BS, provided that the channel state information $\mathbf{H}$ of the entire system is known to each BS. This is due to the fact that the objective in (3) can be decoupled to N_B subproblems involving only the individual rows of $\mathbf{W}$. The channel $\mathbf{H}$ may be

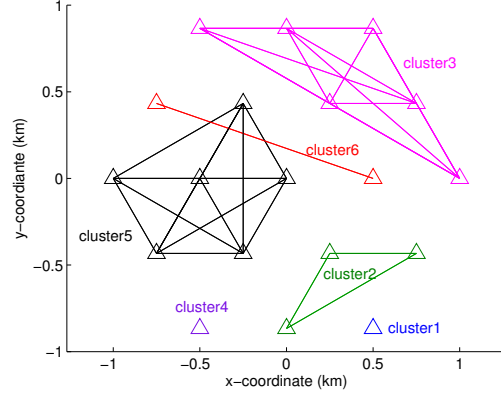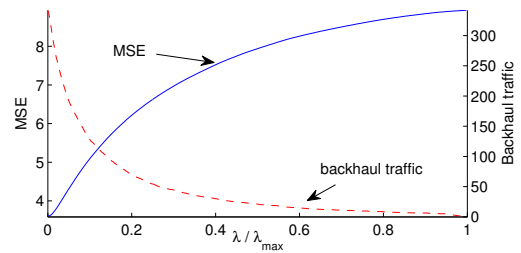**Fig. 1**. A 19-cell network.

Fig. 2. MSE and backhaul traffic.

broadcast infrequently (once per coherence time) for a static or slowly varying system. Similarly, the sparse filter for clustered cooperation with static clustering can be computed in a distributed way, provided the clusters are known to all BSs. Spectral clustering is a centralized operation, but even this task may be accomplished distributively, e.g., using the approach in [14], which is left for future investigation.

5. NUMERICAL TESTS

To test the proposed algorithms, a simple network comprising 19 single-antenna BSs with 500 m cell radius is considered. The locations of the BSs are shown in Fig. 1 as triangles. The MSs were dropped uniformly in the coverage area. The channels between the MSs and the BSs are composite of log-normal shadowing with standard deviation 8 dB, and Rayleigh small-scale fading. The path-loss followed the 3GPP urban model given as $\mathrm{PL(dB)} = 148.1 + 37.6 \log_{10}(d)$, where d is the distance between the BS and the MS in km. A path-loss exponent of $\alpha = 3.76$ was used. The MSs were assigned to the BSs with highest long-term channel gain.

Fig. 2 depicts the MSE and the amount of backhaul traffic as λ is varied under distributed cooperation. The system signal-to-noise-power-ratio (SNR), defined as the average SNR at which the BSs receive signals from cell-edge MSs without accounting for inter-cell interference, was 6.2 dB. $\lambda_{\max}$ denotes the value of λ that yields zero backhaul traffic. Thus, $\lambda = 0$ represents full cooperation, and $\lambda = \lambda_{\max}$,
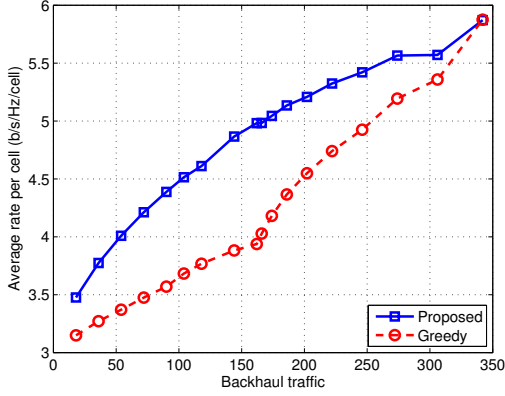
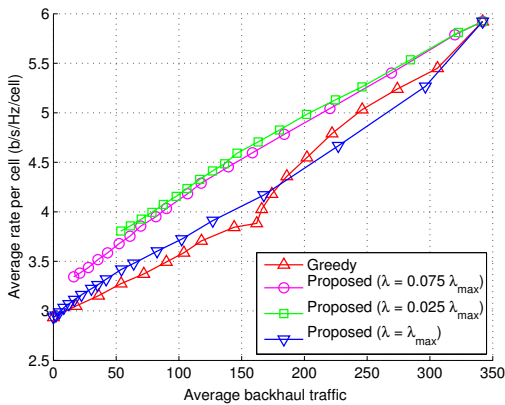**Fig. 3**. Average achievable rates for distributed cooperation.

Fig. 4. Average achievable rates with dynamic clustering.

no cooperation. It can be seen that at a small sacrifice of MSE, significant reduction in backhaul traffic can be achieved.

The performance of the proposed algorithm is compared to that of a greedy clustering algorithm in [8]. Given a cluster size, the greedy scheme picks the BS that yields the largest increase in the sum-rate as a member of the cluster sequentially. Once a cluster is formed, the algorithm moves on to the next cluster until all BSs belong to clusters. No inter-cluster backhauling is allowed in the greedy scheme. However, it is assumed that full cooperation is performed inside clusters. Fig. 3 depicts the achieved average per-cell rates of the proposed distributed cooperation and the greedy scheme as functions of backhaul traffic, at system SNR 11.2 dB. It is clearly seen that the proposed scheme achieves higher rates.

To test performance of the clustered cooperation, Fig. 4 shows the average per-cell rates under the proposed dynamic clustered cooperation for different values of λ. Each marker in the performance curves of the proposed scheme represents an average corresponding to a particular value of N_C. The backhaul traffic includes both inter- and intra-cluster backhauling. It can be seen that even without inter-cluster backhaul at $\lambda = \lambda_{\max}$, the spectral clustering-based algorithm is superior to the greedy scheme at low backhaul traffic. As inter-cluster backhauling is allowed, significant increase in rate can be achieved. Fig. 1 shows an example of the clus-

ter formed at $\lambda = 0.075\lambda_{\max}$.

6. CONCLUSIONS

Backhaul-constrained MCP was considered for cellular uplinks. Exploiting recent compressive sensing techniques, a reduced-backhaul linear equalizer was obtained. In the case of clustered cooperation, sparsity was promoted on inter-cluster feedback. For the dynamic clustering set-up, a joint cluster formation and equalization problem was formulated, and an iterative algorithm based on spectral clustering was developed. Numerical tests demonstrated that significant reduction in MSE is possible at small increase in backhaul traffic.

7. REFERENCES

[1] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun.*, vol. 13, no. 4, pp. 56–61, Aug. 2006.

[2] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: a new look at interference," *J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[3] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. 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.

[4] S. V. Hanly and P. Whiting, "Information-theoretic capacity of multi-receiver networks," *Telecommun. Syst.*, vol. 1, no. 1, pp. 1–42, Dec. 1993.

[5] S. Shamai and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. of the 53rd IEEE Vehicular Tech. Conf.*, May 2001, vol. 3, pp. 1745–1749.

[6] P. Marsch and G. Fettweis, "A framework for optimizing the uplink performance of a distributed antenna systems under a constrained backhaul," in *Proc. of the IEEE Intl. Conf. Commun.*, Glasgow, Scotland, Jun. 2007, pp. 975–979.

[7] C. Hoymann, L. Falconetti, and R. Gupta, "Distributed uplink signal processing of cooperating base stations based on IQ sample exchange," in *Proc. of the IEEE Intl. Conf. Commun.*, Dresden, Germany, Jun. 2009.

[8] A. Papadogiannis, D. Gesbert, and E. Hardouin, "A dynamic clustering approach in wireless networks with multi-cell cooperative processing," in *Proc. of the IEEE Intl. Conf. Commun.*, Beijing, China, May 2008, pp. 4033–4037.

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

[10] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. R. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 2006.

[11] U. von Luxburg, "A tutorial on spectral clustering," *Stat. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

[12] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Statist. Soc. B*, vol. 68, pp. 49–67, Feb. 2006.

[13] M. Meila and W. Pentney, "Clustering by weighted cuts in directed graphs," in *Proc. of the SIAM Intl. Conf. on Data Mining*, Minneapolis, MN, Apr. 2007, pp. 135–144.

[14] P. Forero, A. Cano, and G. B. Giannakis, "Distributed clustering using wireless sensor networks," *IEEE J. Sel. Topics Sig. Proc.*, vol. 5, no. 4, pp. 707–724, Aug. 2011.