Joint Base Station Selection and Distributed Compression for Cloud Radio Access Networks

[1]Seok-Hwan Park, [1]Osvaldo Simeone, [2]Onur Sahin and [3]Shlomo Shamai (Shitz)
[1]CWCSPR, New Jersey Institute of Technology, 07102 Newark, New Jersey, USA
[2]InterDigital Inc., Melville, New York, 11747, USA
[3]Department of Electrical Engineering, Technion, Haifa, 32000, Israel
Email: {seok-hwan.park, osvaldo simeone}@njit.edu, Onur.Sahin@interdigital.com, sshlomo@ee.technion.ac.il

***Abstract*—This work studies joint base station (BS) selection and distributed compression for the uplink of a cloud radio access network. Multiple multi-antenna BSs are connected to a central unit, also referred to as cloud decoder, via capacity-constrained backhaul links. Since the signals received at different BSs are correlated, distributed source coding strategies for communication to the cloud decoder are potentially beneficial. Moreover, reducing the number of active BSs can improve the network energy efficiency, since BS energy consumption provides a major contribution to the overall energy expenditure for the network. An optimization problem is formulated in which compression and BS selection are performed jointly by introducing a sparsity-inducing term into the objective function. An iterative algorithm is proposed. From numerical results, it is observed that the proposed joint BS selection and compression algorithm performs close to the more complex exhaustive search solution.**

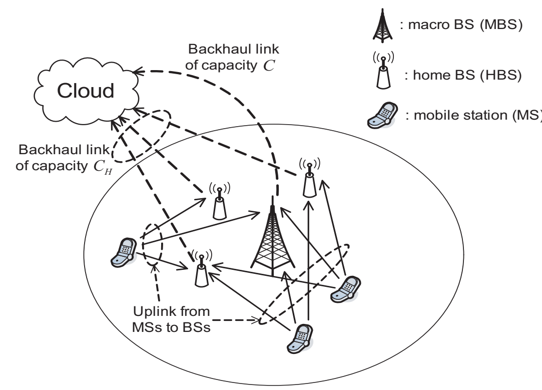

Figure 1. Illustration of a cell of a cloud radio access networks with BSs classified as Home BSs (HBSs) and Macro BSs (MBSs).

I. Introduction

In cloud radio access networks, the baseband processing of the base stations (BSs) is migrated to a central unit in the "cloud", to which the BSs are connected via backhaul links (see, e.g., [1]). On the uplink of a cloud radio access network, the BSs relay "soft" information to the cloud decoder regarding the received baseband signals. Since the signals received at different BSs are correlated, *distributed source coding* strategies are generally beneficial for the communication from the BSs to the cloud decoder. This was first demonstrated in [2] for single-antenna BSs and then extended to multi-antenna BSs in [3] (see also [4]-[6] for related work).

In this paper, we study the problem of compression with distributed source coding in the presence of multi-antenna BSs by focusing on the issue of *energy efficiency*. An efficient operation of cloud radio access networks is understood to require a parsimonious use of the BSs, whose energy consumption is among the most relevant contributions to the overall energy expenditure of the network. We address this problem by proposing a joint compression and BS selection approach.

The issue of energy efficient network operation via scheduling of the BSs was tackled for a multi-antenna downlink system in [7] and for a single-antenna uplink system in [8]. Here, we focus on a multi-antenna uplink system. An optimization problem is formulated in which *compression* and *BS selection* are performed jointly by introducing a sparsity-inducing term into the objective function (Sec. III). This is inspired by the strategy proposed in [7] for the design of beamforming vectors for the downlink. We emphasize that the downlink beamforming problem studied in [7] is significantly different from the uplink compression problem of interest here. An iterative block-coordinate ascent algorithm is then proposed (Sec. IV). Numerical results demonstrate that the proposed joint BS selection and compression algorithm performs close to the more complex exhaustive search solution (Sec. V).

Notation: The probability density function of a random variable X is denoted by $p(x)$ and similar notations are used for joint and conditional distributions. Given a vector $\mathbf{x} = [x_1, x_2.., x_n]^T$, we define $\mathbf{x}_{\mathcal{S}}$ for a subset $\mathcal{S} \subseteq \{1, 2, ..., n\}$ as the vector including, in arbitrary order, the entries x_i with $i \in \mathcal{S}$. Notation $\mathbf{\Sigma}_{\mathbf{x}}$ is used for the correlation matrix of random vector $\mathbf{x}$, i.e., $\mathbf{\Sigma}_{\mathbf{x}} = \mathrm{E}[\mathbf{x}\mathbf{x}^{\dagger}]$; and $\mathbf{\Sigma}_{\mathbf{x}|\mathbf{y}}$ represents the "conditional" correlation matrix of $\mathbf{x}$ given $\mathbf{y}$.

II. System Model

To fix the ideas, we consider a single cell with one Macro BS (MBS) with $n_{B,1}$ antennas and $N_B - 1$ Home BSs (HBSs), each with $n_{B,i}$ antennas ($i = 2, \ldots, M$), which serve N_M active MSs. We refer to Fig. 1 for an illustration. We denote the set of the indices of all HBSs as $\mathcal{S}_{\mathcal{H}} = \{2, ..., N_B\}$ and reserve index 1 for the MBS. The MBS is connected to the cloud decoder via a finite-capacity link of capacity C while the HBSs share a total backhaul capacity C_H. This assumption is reasonable since the MBS typically has a dedicated backhaul link, whereas smaller BSs may share a common wireless channel [9]. While the discussion focuses on a single cell, the discussion can be easily extended to multiple cells with minor changes to the model.

Defining $\mathbf{H}_{ij}$ as the $n_{B,i} \times n_{M,j}$ channel matrix between the jth MS and the ith BS, the overall channel matrix toward BS i is given as the $n_{B,i} \times n_M$ matrix

$$\mathbf{H}_i = [\mathbf{H}_{i1} \cdots \mathbf{H}_{iN_M}], \tag{1}$$

with $n_M = \sum_{i=1}^{N_M} n_{M,i}$. Assuming that all the N_M MSs in a cluster are synchronous, at any discrete-time channel use (c.u.), the signal received by the ith BS is given by

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{x} + \mathbf{z}_i. \tag{2}$$

In (2), vector $\mathbf{x} = [\mathbf{x}_1^\dagger \cdots \mathbf{x}_{N_M}^\dagger]^\dagger$ is the $n_M \times 1$ vector of symbols transmitted by all the MSs in the cluster at hand. The noise vectors $\mathbf{z}_i$ are independent over i and are distributed as $\mathbf{z}_i \sim \mathcal{CN}(0, \mathbf{I})$, for $i \in \{1, ..., N_B\}$. The channel matrix $\mathbf{H}_i$ is assumed to be constant in each time-slot and, unless stated otherwise, is considered to be known at the cloud decoder.

A. BS Selection and Achievable Sum-Rate

Using standard random coding arguments [10], the coding strategies employed by the MSs in each time-slot entail a distribution $p(\mathbf{x})$ on the transmitted signals that factorizes as $p(\mathbf{x}) = \prod_{i=1}^{N_M} p(\mathbf{x}_i)$, since the signals sent by different MSs are independent. We will assume throughout that the distribution $p(\mathbf{x}_i)$ of the signal transmitted by the ith MS is given as $\mathbf{x}_i \sim \mathcal{CN}(0, \mathbf{\Sigma}_{\mathbf{x}_i})$ for a given covariance matrix $\mathbf{\Sigma}_{\mathbf{x}_i}$.

The BSs communicate with the cloud by providing the latter with *soft information* derived from the received signal. Specifically, each active BS compresses the received signal and forwards it to the cloud. Note that the BSs do not need to be informed about the MSs' codebooks [2]. Using conventional rate-distortion theory arguments, a compression strategy for the ith BS is described by a *test channel* $p(\hat{\mathbf{y}}_i|\mathbf{y}_i)$ that characterizes the relationship between the signal to be compressed, namely $\mathbf{y}_i$, and its description $\hat{\mathbf{y}}_i$ to be communicated to the cloud (see, e.g., [11]). We will discuss in Sec. II-B the requirements on the backhaul capacity that are entailed by the choice of given test channels.

In order to operate the network efficiency, it is advantageous to let only a subset of BSs communicate to the cloud decoder. Assuming that the MBS is always active, we are then interested in selecting a subset $\mathcal{S}_{\mathcal{H}}^*$ of HBSs (i.e., $\mathcal{S}_{\mathcal{H}}^* \subseteq \mathcal{S}_{\mathcal{H}}$) under the given total backhaul constraint C_H for the HBSs.

The cloud decoder decodes jointly the signals $\mathbf{x}$ of all MSs based on the received descriptions $\hat{\mathbf{y}}_i$ for $i \in \mathcal{S}_{\text{active}}$ where $\mathcal{S}_{\text{active}} = \{1\} \cup \mathcal{S}_{\mathcal{H}}^*$ is the set of active BSs, including the MBS and the selected HBSs. From standard information-theoretic considerations (see, e.g., [11]), the achievable sum-rate is hence given by the mutual information

$$R_{sum} = I(\mathbf{x}; \hat{\mathbf{y}}_{\mathcal{S}_{\text{active}}}). \tag{3}$$

B. Distributed Compression

As mentioned, since the signals $\mathbf{y}_i$ measured by different BSs are correlated, distributed source coding techniques have the potential to improve the quality of the descriptions $\hat{\mathbf{y}}_i$ [2]. Specifically, given compression test channels $p(\hat{\mathbf{y}}_i|\mathbf{y}_i)$, thanks to distributed source coding strategies, the descriptions $\hat{\mathbf{y}}_i$ can be recovered at the cloud as long as the capacities C and C_H satisfy the following conditions (see, e.g., [11])

$$I(\mathbf{y}_1; \hat{\mathbf{y}}_1) \leq C, \tag{4}$$

$$\text{and } I(\mathbf{y}_{\mathcal{S}_{\mathcal{H}}^*}; \hat{\mathbf{y}}_{\mathcal{S}_{\mathcal{H}}^*} | \hat{\mathbf{y}}_1) \leq C_H. \tag{5}$$

This should be contrasted to the more restrictive conditions (4) and $\sum_{i \in \mathcal{S}_{\mathcal{H}}} I(\mathbf{y}_i; \hat{\mathbf{y}}_i) \leq C_H$ that would be necessary by using standard compression techniques that do not leverage correlation among the symbols $\{\mathbf{y}_i\}_{i \in \mathcal{S}_{\mathcal{H}}}$ (see, e.g., [11]).

Remark 1. In the approach discussed above, the central processor in the cloud decodes the descriptions $\hat{\mathbf{y}}_{\mathcal{S}_{\text{active}}}$ received from the BSs and then perform joint decoding of all of the MSs' signals. An alternative approach, discussed in [2], is instead that of performing joint decoding of both the descriptions $\hat{\mathbf{y}}_{\mathcal{S}_{\text{active}}}$ and of the signals $\mathbf{x}$ transmitted by the MSs. We will not further consider this approach here.

III. Problem Definition

In this section, we address the problem of selecting the subset $\mathcal{S}_{\mathcal{H}}^*$ of HBSs, along with the compression test channels $p(\hat{\mathbf{y}}_i|\mathbf{y}_i)$ for $i \in \mathcal{S}_{\mathcal{H}}^*$ in order to maximize the sum-rate (3) under the constraints (4) and (5). In general, the optimal BS selection requires an exhaustive search of exponential complexity in the number N_B of BSs. Here, inspired by [7], we propose an efficient approach based on the addition of a sparsity-inducing term to the objective sum-rate function (3).

To elaborate, we associate a cost q_i per spectral unit resource (i.e., per discrete-time channel use) to the ith BS. This measures the relative cost per spectral resource of activating the ith BS over the revenue per bit. Based on the results [3][12], we assume, without claim of optimality, that the Gaussian test channel $p(\hat{\mathbf{y}}_i|\mathbf{y}_i)$ is employed at each ith BS. This entails that the test channel $p(\hat{\mathbf{y}}_i|\mathbf{y}_i)$ is given as

$$\hat{\mathbf{y}}_i = \mathbf{A}_i \mathbf{y}_i + \mathbf{q}_i, \tag{6}$$

where $\mathbf{A}_i$ is a matrix to be calculated and $\mathbf{q}_i \sim \mathcal{CN}(0, \mathbf{I})$ is the compression noise, which is independent of $\mathbf{x}$ and $\mathbf{z}_i$.

The joint problem of HBS selection and compression via sparsity-inducing optimization is then formulated as

$$\begin{aligned} \underset{\{\mathbf{\Omega}_i \succeq \mathbf{0}\}_{i \in \mathcal{S}_{\mathcal{H}}}}{\text{maximize}} \; & I\left(\mathbf{x}; \hat{\mathbf{y}}_{\mathcal{S}_{\mathcal{H}}} | \hat{\mathbf{y}}_1\right) - q_H \sum_{i \in \mathcal{S}_{\mathcal{H}}} 1\left(\|\mathbf{\Omega}_i\|_F > 0\right) \\ \text{s.t. } & I\left(\mathbf{y}_{\mathcal{S}_{\mathcal{H}}}; \hat{\mathbf{y}}_{\mathcal{S}_{\mathcal{H}}} | \hat{\mathbf{y}}_1\right) \leq C_H, \end{aligned} \tag{7}$$

where matrix $\mathbf{\Omega}_i$ is defined as $\mathbf{\Omega}_i = \mathbf{A}_i^\dagger \mathbf{A}_i$, $1(\cdot)$ is the indicator function, which takes 1 if the argument state is true and 0 otherwise, and we have assumed that $q_2 = \ldots = q_{N_B} = q_H$ for simplicity. Note that the second term in the objective of problem (7) is the ℓ_0-norm of vector $[\,\mathrm{tr}(\mathbf{\Omega}_2) \cdots \mathrm{tr}(\mathbf{\Omega}_{N_B})\,]$. If the cost q_H is large enough, this term forces the solution

Algorithm 1 Two-Phase Joint HBS Selection and Compression Algorithm

Phase 1. Solve problem (8) via the block-coordinate ascent algorithm:

i) Initialize $n=0$ and $\mathbf{\Omega}_2^{(n)}=\ldots=\mathbf{\Omega}_{N_B}^{(n)}=\mathbf{0}$;

ii) For $i=2,\ldots,N_B$, update $\mathbf{\Omega}_i^{(n)}$ as a solution of the following problem.

$$\underset{\mathbf{\Omega}_i\succeq\mathbf{0}}{\text{maximize}}\ f\left(\mathbf{\Omega}_i,\mathbf{\Omega}^{(n-1)}_{\{2,\ldots,i-1\}},\mathbf{\Omega}^{(n)}_{\{i+1,\ldots,N_B\}}\right) \quad (13)$$
$$\text{s.t. } g\left(\mathbf{\Omega}_i,\mathbf{\Omega}^{(n-1)}_{\{2,\ldots,i-1\}},\mathbf{\Omega}^{(n)}_{\{i+1,\ldots,N_B\}}\right)\le C_H.$$

iii) Repeat step ii) if some convergence criterion is not satisfied and stop otherwise.

iv) Once the algorithm has terminated, denote the obtained $\mathbf{\Omega}_i$ by $\mathbf{\Omega}_i^*$ for $i=2,\ldots,N_B$, and set

$$\mathcal{S}_{\mathcal{H}}^*=\{i\in\mathcal{S}_{\mathcal{H}}:\mathbf{\Omega}_i^*\neq\mathbf{0}\}.$$

Phase 2. Apply the block-coordinate ascent algorithm to problem (8) with $q_H=0$ and considering only the HBSs in set $\mathcal{S}_{\mathcal{H}}^*$.

to set some of the matrices $\mathbf{\Omega}_i$ to zero, thus keeping the corresponding ith HBS inactive.

As is standard practice, in order to avoid the non-smoothness induced by the ℓ_0-norm, we modify problem (7) by replacing the ℓ_0-norm with the ℓ_1-norm of the same vector. We thus reformulate problem (7) as follows:

$$\underset{\{\mathbf{\Omega}_i\succeq\mathbf{0}\}_{i\in\mathcal{S}_{\mathcal{H}}}}{\text{maximize}}\ f(\mathbf{\Omega}_2,\ldots,\mathbf{\Omega}_{N_B}) \quad (8)$$
$$\text{s.t. } g(\mathbf{\Omega}_2,\ldots,\mathbf{\Omega}_{N_B})\le C_H,$$

where we have defined $f(\mathbf{\Omega}_2,\ldots,\mathbf{\Omega}_{N_B})=I(\mathbf{x};\hat{\mathbf{y}}_{\mathcal{S}_{\mathcal{H}}}|\hat{\mathbf{y}}_1)-q_H\sum_{i\in\mathcal{S}_{\mathcal{H}}}\text{tr}(\mathbf{\Omega}_i)$ and $g(\mathbf{\Omega}_2,\ldots,\mathbf{\Omega}_{N_B})=I(\mathbf{y}_{\mathcal{S}_{\mathcal{H}}};\hat{\mathbf{y}}_{\mathcal{S}_{\mathcal{H}}}|\hat{\mathbf{y}}_1)$. An explicit expression for $I(\mathbf{x};\hat{\mathbf{y}}_{\mathcal{S}_{\mathcal{H}}}|\hat{\mathbf{y}}_1)$ and $I(\mathbf{y}_{\mathcal{S}_{\mathcal{H}}};\hat{\mathbf{y}}_{\mathcal{S}_{\mathcal{H}}}|\hat{\mathbf{y}}_1)$ as a function of $\mathbf{\Omega}_2,\ldots,\mathbf{\Omega}_{N_B}$ can be easily obtained as

$$I(\mathbf{x};\hat{\mathbf{y}}_{\mathcal{S}_{\mathcal{H}}}|\hat{\mathbf{y}}_1)=\log\det(\mathbf{I}+\mathbf{C}), \quad (9)$$
$$\text{and } I(\mathbf{y}_{\mathcal{S}_{\mathcal{H}}};\hat{\mathbf{y}}_{\mathcal{S}_{\mathcal{H}}}|\hat{\mathbf{y}}_1)=\log\det(\mathbf{I}+\mathbf{C}) \quad (10)$$
$$+\sum_{i\in S_H}\log\det(\mathbf{I}+\mathbf{\Omega}_i),$$

where matrix $\mathbf{C}$ is defined as

$$\mathbf{C}=\mathbf{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_1}\sum_{j\in\mathcal{S}_{\mathcal{H}}}\mathbf{H}_j^\dagger(\mathbf{I}+\mathbf{\Omega}_j)^{-1}\mathbf{\Omega}_j\mathbf{H}_j. \quad (11)$$

The conditional covariance matrix $\mathbf{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_1}$ is given as

$$\mathbf{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_1}=\mathbf{\Sigma}_{\mathbf{x}}-\mathbf{\Sigma}_{\mathbf{x}}\bar{\mathbf{H}}_1^\dagger\left(\bar{\mathbf{H}}_1\mathbf{\Sigma}_{\mathbf{x}}\bar{\mathbf{H}}_1^\dagger+\mathbf{A}_1\mathbf{A}_1^\dagger\right)^{-1}\bar{\mathbf{H}}_1\mathbf{\Sigma}_{\mathbf{x}}, \quad (12)$$

with $\bar{\mathbf{H}}_1=\mathbf{A}_1\mathbf{H}_1$.

IV. Joint BS Selection and Compression via Sparsity-Inducing Optimization

In this section, we propose a joint BS selection and compression algorithm based on the sparsity-inducing optimization (8). Specifically, we propose a two-phase approach in Algorithm 1. As illustrated in the table, in the first phase we execute the block-coordinate ascent algorithm [13] to obtain a stationary point (i.e., local optimum solution) for problem (8). As a result, we obtain a subset $\mathcal{S}_{\mathcal{H}}^*\subseteq\mathcal{S}_{\mathcal{H}}$ of HBSs with nonzero $\mathbf{\Omega}_i$. In the second phase, the block-coordinate ascent algorithm is run on the subset $\mathcal{S}_{\mathcal{H}}^*$ obtained in the first phase by setting $\mathbf{\Omega}_i=\mathbf{0}$ for all $i\notin\mathcal{S}_{\mathcal{H}}^*$ and $q_H=0$. This second phase is needed to refine the test channels obtained in the first phase. It is noted that with $q_H=0$, the block-coordinate ascent method used here is the same as proposed in [3, Sec. IV].

It remains to discuss how to solve problem (13) at step ii) of the proposed algorithm. Note that this corresponds to a step of block coordinate ascent in which we update $\mathbf{\Omega}_i$ and all the other variables $\mathbf{\Omega}_{\mathcal{S}_{\mathcal{H}}\setminus\{i\}}$ are fixed to the values obtained from the previous iteration. The global maximum of problem (13) can be obtained as shown in Theorem 1.

Theorem 1. *A solution to problem (13) is given by (6), with matrix $\mathbf{A}_i$ such that $\mathbf{\Omega}_i=\mathbf{A}_i^\dagger\mathbf{A}_i$ is given as*

$$\mathbf{\Omega}_i^*=\mathbf{U}\text{diag}\left(\alpha_1,\ldots,\alpha_{n_{B,i}}\right)\mathbf{U}^\dagger, \quad (14)$$

where we have the eigenvalue decomposition $\mathbf{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_{\mathcal{B}}\setminus\{i\}}}=\mathbf{U}\text{diag}(\lambda_1,\ldots,\lambda_{n_{B,i}})\mathbf{U}^\dagger$*, and matrix* $\mathbf{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_{\mathcal{B}}\setminus\{i\}}}$ *is given as*

$$\mathbf{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_{\mathcal{B}}\setminus\{i\}}}=\mathbf{I}+\mathbf{H}_i\mathbf{R}_i^{-1}\mathbf{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_1}\mathbf{H}_i^\dagger, \quad (15)$$

with $\mathbf{R}_i=\mathbf{I}+\mathbf{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_1}\sum_{j\in\mathcal{S}_{\mathcal{H}}\setminus\{i\}}\mathbf{H}_j^\dagger(\mathbf{I}+\mathbf{\Omega}_j)^{-1}\mathbf{\Omega}_j\mathbf{H}_j$. *The diagonal elements $\alpha_1,\ldots,\alpha_{n_{B,i}}$ are obtained as $\alpha_l=\alpha_l(\mu^*)$, with*

$$\alpha_l(\mu)=\frac{\left[-a_l(\mu)+\sqrt{a_l^2(\mu)-2q'_H\lambda_l b_l(\mu)}\right]^+}{2q'_H\lambda_l}, \quad (16)$$

for $l=1,\ldots,n_{B,i}$ with $q'_H=\log_e 2\cdot q_H$, $a_l(\mu)=\lambda_l\mu+q'_H(\lambda_l+1)$ and $b_l(\mu)=2((\mu-1)\lambda_l+q'_H+1)$. The Lagrange multiplier μ^* is obtained as follows: if $h_i(0)\le\bar{C}_i$, where $\bar{C}_i$ is given by

$$\bar{C}_i=C_i-\log\det\mathbf{R}_i-\sum_{j\in\mathcal{S}_{\mathcal{H}\setminus\{i\}}}\log\det(\mathbf{I}+\mathbf{\Omega}_j), \quad (17)$$

then $\mu^*=0$; otherwise, μ^* is unique value $\mu\ge 0$ such that $h_i(\mu)=\bar{C}_i$ where $h_i(\mu)=\sum_{l=1}^{n_{B,i}}\log(1+\lambda_l\alpha_l(\mu))$.

Proof: The proof is given in Appendix A. ■

V. Numerical Results

In this section, we present numerical results in order to validate and complement the analysis. The sum-rate performance is evaluated for a single cell as described in Sec. II. We consider a practical scenario where the MBS is located at the cell center while the HBSs and MSs are divided into two groups: N_B^1 HBSs and N_M^1 MSs in group 1 and N_B^2 HBSs and N_M^2 MSs in group 2 such that $N_B^1+N_B^2=N_B$ and $N_M^1+N_M^2=N_M$. The HBSs and MSs in group 1 are uniformly distributed within the whole cell with cell-radius R_{cell}, while those in group 2 are distributed in smaller cell overlaid on the macrocell at hand and with radius $R_{\text{spot}}<$

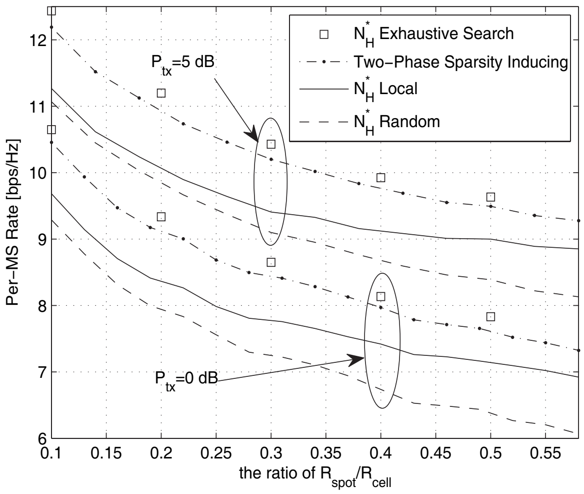

Figure 2. Average per-MS sum-rate versus the ratio $R_{\rm spot}/R_{\rm cell}$ in a single-cell heterogeneous network with $N_B = 13$ ($N_B^1 = 6$, $N_B^2 = 6$), $N_M = 14$ ($N_M^1 = 8$, $N_M^2 = 6$), $n_{B,i} = 8$, $C = 20$ bps/Hz, $C_H = 300$ bps/Hz and $q_H = 100$.

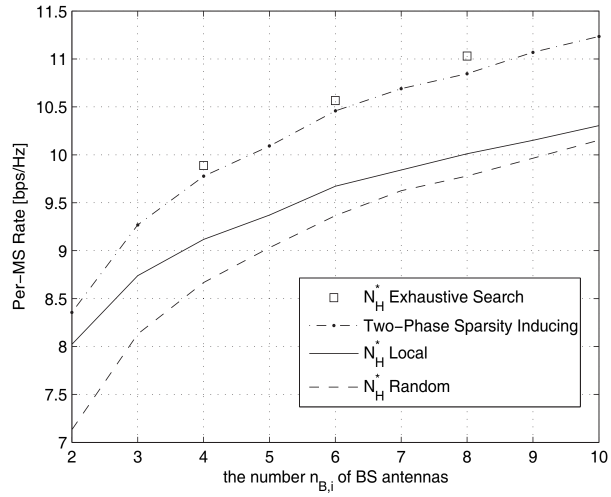

Figure 3. Average per-MS sum-rate versus the number $n_{B,i}$ of BS antennas in a single-cell heterogeneous network with $N_B = 13$ ($N_B^1 = 6$, $N_B^2 = 6$), $N_M = 14$ ($N_M^1 = 8$, $N_M^2 = 6$), $C = 20$ bps/Hz, $C_H = 300$ bps/Hz, $q_H = 100$, $P_{tx} = 5\,\mathrm{dB}$ and $R_{\rm spot}/R_{\rm cell} = 0.2$.

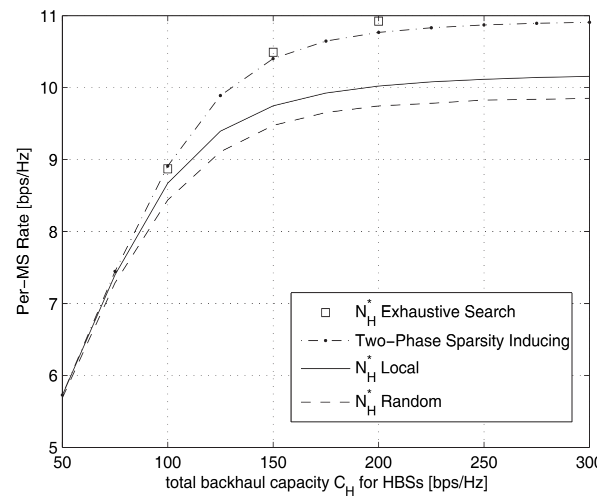

Figure 4. Average per-MS sum-rate versus the total backhaul capacity C_H of HBSs in a single-cell heterogeneous network with $N_B = 13$ ($N_B^1 = 6$, $N_B^2 = 6$), $N_M = 14$ ($N_M^1 = 8$, $N_M^2 = 6$), $n_{B,i} = 8$, $C = 20$ bps/Hz, $q_H = 100$, $P_{tx} = 5\,\mathrm{dB}$ and $R_{\rm spot}/R_{\rm cell} = 0.2$.

$R_{\rm cell}$, which models a *"hot spot*" such as a building or a public space. All channel elements of $\mathbf{H}_{i,j}$ are i.i.d. circularly symmetric complex Gaussian variables with zero mean and variance $(D_0/d_{i,j})^{\nu}$, where the path-loss exponent ν is chosen as 3.5 and $d_{i,j}$ is the distance from MS j to BS i. The reference distance D_0 is set to half of cell-radius, i.e., $D_0 = R_{\rm cell}/2$. For simplicity, each MS uses single antenna, i.e., $n_{M,i} = 1$ with transmit power $P_{\rm tx}$ such that the aggregated transmit vector $\mathbf{x}$ has a covariance of $\mathbf{\Sigma}_{\mathbf{x}} = P_{\rm tx}\mathbf{I}$.

We compare the proposed two-phase approach (Algorithm 1) with:

i) N_H^* *Exhaustive Search*: the scheme selects N_H^* HBSs that maximize the sum-rate via exhaustive search, where N_H^* is the cardinality of the set $\mathcal{S}_{\mathcal{H}}^*$ obtained after the first phase of the two-phase algorithm;

ii) N_H^* *Local*: the scheme selects the N_H^* HBSs with the largest value $C_i^{\rm local}$, where $C_i^{\rm local}$ is the capacity from the N_M MSs to BS i, i.e., $C_i^{\rm local} = \log\det\left(\mathbf{I} + \mathbf{H}_i\mathbf{\Sigma}_{\mathbf{x}}\mathbf{H}_i^{\dagger}\right)$. Note that this criterion is local in that it does not account for the correlation between the signals received by different HBSs.

iii) N_H^* *Random*: the scheme randomly selects N_H^* HBSs.

Fig. 2 presents the per-MS sum-rate versus the ratio of the ratio $R_{\rm spot}/R_{\rm cell}$ with $N_B = 13$ ($N_B^1 = 6$, $N_B^2 = 6$), $N_M = 14$ ($N_M^1 = 8$, $N_M^2 = 6$), $n_{B,i} = 8$, $C = 20$ bps/Hz, $C_H = 300$ bps/Hz and $q_H = 100$. From the figure, it is observed that the N_H^* Local approach provides a performance close to the N_H^* Exhaustive Search when the size of the *hot spot* is large. In fact, in this case, the correlation between pairs of signals received by different HBSs tends to be similar given the symmetry of the network topology. However, for sufficiently small *hot spot* size, the performance loss of the N_H^* Local approach becomes significant, while the proposed two-phase scheme still shows a performance almost identical to that of the N_H^* Exhaustive Search scheme which requires a search over $\binom{12}{6} = 924$ combinations of HBSs since we observed that the average number N_H^* of activated HBSs is about 6 for the simulated configurations. This is because, in this case, signals received by HBSs in the smaller *hot spot* tend to be more correlated, and thus it is more advantageous to select the HBSs judiciously in order to increase the sum-rate.

In Fig. 3 and Fig. 4, we plot the per-MS sum-rate versus the number $n_{B,i}$ of BS antennas and the total backahaul capacity C_H of HBSs, respectively, in the same configuration as in Fig. 2 except for the fixed ratio of $R_{\rm spot}/R_{\rm cell} = 0.2$. It is seen that for large number $n_{B,i}$ of antennas or large backhaul capacity

C_H, more gain can be harnessed by an appropriate scheduling of the HBSs. In particular, when the backhaul capacity C_H is very small, the compression noise dominates the performance and thus the effect of optimal scheduling is negligible.

VI. Conclusions

this work, we have studied distributed compression schemes for the uplink of cloud radio access networks. Focusing on the issue of energy efficiency, we have addressed the selection of a subset of BSs by proposing a joint BS selection and compression approach, in which a sparsity-inducing term is introduced into the objective function. We proposed an algorithm based on block coordinate ascent. It was verified that the proposed joint BS scheduling and compression method shows performance close to the more complex exhaustive search.

Appendix A
Proof of Theorem 1

The following lemma provides a problem formulation equivalent to (13).

Lemma 1. *Problem (13) is equivalent to*

$$\underset{\mathbf{\Omega}_i \succeq \mathbf{0}}{\text{maximize}} \left\{ \begin{array}{c} \log\det\left(\mathbf{I}+\mathbf{\Omega}_i\mathbf{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}}\right) \\ -\log\det\left(\mathbf{I}+\mathbf{\Omega}_i\right)-q_H\text{tr}\left(\mathbf{\Omega}_i\right) \end{array} \right\} \quad (18)$$
$$\text{s.t.}\ \log\det\left(\mathbf{I}+\mathbf{\Omega}_i\mathbf{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}}\right) \leq \bar{C}_i,$$

where $\mathbf{\Sigma}_{\tilde{\mathbf{y}}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}}$ *and* $\bar{C}_i$ *are defined in Theorem 1.*

Proof: We first observe that from the chain rule, the mutual information terms in problem (13) can be written as

$$I\left(\mathbf{x};\hat{\mathbf{y}}_{\mathcal{S}_\mathcal{H}}|\hat{\mathbf{y}}_1\right) = I\left(\mathbf{x};\hat{\mathbf{y}}_i|\hat{\mathbf{y}}_1,\hat{\mathbf{y}}_{\mathcal{S}_\mathcal{H}\setminus\{i\}}\right) + I\left(\mathbf{x};\hat{\mathbf{y}}_{\mathcal{S}_\mathcal{H}\setminus\{i\}}|\hat{\mathbf{y}}_1\right), \quad (19)$$

$$\text{and}\ I\left(\mathbf{y}_{\mathcal{S}_\mathcal{H}};\hat{\mathbf{y}}_{\mathcal{S}_\mathcal{H}}|\hat{\mathbf{y}}_1\right) = I\left(\mathbf{y}_i;\hat{\mathbf{y}}_i|\hat{\mathbf{y}}_1,\hat{\mathbf{y}}_{\mathcal{S}_\mathcal{H}\setminus\{i\}}\right) + I\left(\mathbf{y}_{\mathcal{S}_\mathcal{H}\setminus\{i\}};\hat{\mathbf{y}}_{\mathcal{S}_\mathcal{H}\setminus\{i\}}|\hat{\mathbf{y}}_1\right). \quad (20)$$

In (19), the matrix $\mathbf{\Omega}_i$ affects only the first term, which gives the objective function in (18). Similarly, calculating the first and second terms in (20) results in the constraint of problem (18). ■

Since problem (18) is non-convex, we first solve the KKT conditions, which can be proved to be necessary for optimality, and then show that the derived solution also satisfies the general sufficiency condition in [13, Proposition 3.3.4].

The Lagrangian of problem (18) is given as

$$\mathcal{L}\left(\mathbf{\Omega}_i,\mu,\mathbf{\Upsilon}\right) = (1-\mu)\log\det\left(\mathbf{I}+\mathbf{\Omega}_i\mathbf{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}}\right) - \log\det\left(\mathbf{\Omega}_i+\mathbf{I}\right) - q_H\text{tr}\left(\mathbf{\Omega}_i\right) + \text{tr}\left(\mathbf{\Upsilon}\mathbf{\Omega}_i\right), \quad (21)$$

with Lagrange multipliers $\mu \geq 0$ and $\mathbf{\Upsilon} \succeq \mathbf{0}$. We then write the KKT conditions as

$$\frac{\partial \mathcal{L}}{\partial \mathbf{\Omega}_i} = \mathbf{0}, \quad (22a)$$

$$\text{tr}\left(\mathbf{\Upsilon}\mathbf{\Omega}_i\right) = 0, \quad (22b)$$

$$\mu\left(\log\det\left(\mathbf{I}+\mathbf{\Omega}_i\mathbf{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}}\right)-\bar{C}_i\right) = 0, \quad (22c)$$

$$\text{and}\quad \log\det\left(\mathbf{I}+\mathbf{\Omega}_i\mathbf{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}}\right)-\bar{C}_i \leq 0. \quad (22d)$$

By direct calculation, we can see that setting $\mathbf{\Omega}_i$ as (14) with the eigenvalues $\alpha_1,\ldots,\alpha_{n_{B,i}}$ in (16) satisfy the conditions (22a)-(22d).

Now, we show that the solution (16) also satisfies the general sufficiency condition in [13, Proposition 3.3.4].

Lemma 2. *Let* $\mathbf{\Omega}_i^*, \mu^*$ *denote a pair obtained from Theorem 1. Then,* $\mathbf{\Omega}_i^*$ *is the global optimum of problem (18) since it satisfies the sufficiency condition in [13, Proposition 3.3.4], i.e., the equality*

$$\mathbf{\Omega}_i^* = \arg\max_{\mathbf{\Omega}_i \succeq \mathbf{0}} \mathcal{L}\left(\mathbf{\Omega}_i,\mu^*\right), \quad (23)$$

and the complementary slackness condition (22c), where function $\mathcal{L}\left(\mathbf{\Omega}_i,\mu\right)$ *is the Lagrangian function in (21) with* $\mathbf{\Upsilon}=\mathbf{0}$.

Proof: In problem (23), selecting the eigenvectors of $\mathbf{\Omega}_i$ as those of $\mathbf{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}}$ does not involve any loss of optimality due to the eigenvalue inequality $\log\det\left(\mathbf{I}+\mathbf{AB}\right) \leq \log\det\left(\mathbf{I}+\mathbf{\Gamma_A\Gamma_B}\right)$ where $\mathbf{\Gamma_A}$ and $\mathbf{\Gamma_B}$ are diagonal matrices with diagonal elements of the decreasing eigenvalues of $\mathbf{A}$ and $\mathbf{B}$, respectively, with $\mathbf{A},\mathbf{B} \succeq \mathbf{0}$ (see [3, Appendix B]). Then, (23) is equivalent to showing that the eigenvalues $\alpha_1^*,\ldots,\alpha_{n_{B,i}}^*$ in (16) satisfy

$$\alpha_l^* = \arg\max_{\alpha_l \geq 0}\left((1-\mu^*)\log\left(1+\alpha_l\lambda_l\right)-\log(1+\alpha_l)-q_H\alpha_l\right), \quad (24)$$

for $l=1,\ldots,n_{B,i}$. The above condition can be shown by following the same steps as in [3, Appendix B]. ■

References

[1] P. Marsch, B. Raaf, A. Szufarska, P. Mogensen, H. Guan, M. Farber, S. Redana, K. Pedersen and T. Kolding, "Future mobile communication networks: challenges in the design and operation," *IEEE Vehicular Technology Magazine*, vol.7, no.1, pp.16-23, March 2012.

[2] A. Sanderovich, O. Somekh, H. V. Poor and S. Shamai (Shitz), "Uplink macro diversity of limited backhaul cellular network," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3457-3478, Aug. 2009.

[3] A. del Coso and S. Simoens, "Distributed compression for MIMO coordinated networks with a backhaul constraint," *IEEE Trans. Wireless Comm.*, vol. 8, no. 9, pp. 4698-4709, Sep. 2009.

[4] B. Nazer, A. Sanderovich, M. Gastpar and S. Shamai (Shitz), "Structured superposition for backhaul constrained cellular uplink," in *Proc. IEEE ISIT '09*, Seoul, Korea, June 2009.

[5] S.-N. Hong and G. Caire, "Quantized compute and forward: a low-complexity architecture for distributed antenna systems," in *Proc. IEEE ITW '11*, Paraty, Brazil, Oct. 2011.

[6] S.-H. Park, O. Simeone, O. Sahin and S. Shamai (Shitz), "Robust Distributed Compression for Cloud Radio Access Networks," in *Proc. IEEE ITW '12*, Lausanne, Switzerland, Sep. 2012.

[7] M. Hong, R.-Y. Sun and Z.-Q. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogenous networks," arXiv:1203.6390.

[8] S. Ramanath, V. Kavitha and E. Altman, "Open loop optimal control of base station activation for green networks," in *Proc. WiOpt '11*, Princeton, NJ, May 2011.

[9] I. Maric, B. Bostjancic and A. Goldsmith, "Resource allocation for constrained backhaul in picocell networks," in *Proc. ITA '11*, UCSD, Feb. 2011.

[10] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: Wiley, 2006.

[11] A. E. Gamal and Y.-H. Kim, *Network information theory*, Cambridge University Press, 2011.

[12] G. Chechik, A. Globerson, N. Tishby and Y. Weiss, "Information bottleneck for Gaussian variables," *Jour. Machine Learn.*, Res. 6, pp. 165-188, 2005.

[13] D. Bertsekas, *Nonlinear programming*. New York: Athena Scientific, 1995.