Robust and Efficient Distributed Compression for Cloud Radio Access Networks

Seok-Hwan Park, Osvaldo Simeone, Onur Sahin, and Shlomo Shamai (Shitz), *Fellow, IEEE*

***Abstract*—This paper studies distributed compression for the uplink of a cloud radio access network where multiple multi-antenna base stations (BSs) are connected to a central unit, which is also referred to as a cloud decoder, via capacity-constrained backhaul links. Since the signals received at different BSs are correlated, distributed source coding strategies are potentially beneficial. However, they require each BS to have information about the joint statistics of the received signals across the BSs, and they are generally sensitive to uncertainties regarding such information. Motivated by this observation, a robust compression method is proposed to cope with uncertainties on the correlation of the received signals. The problem is formulated using a deterministic worst case approach, and an algorithm is proposed that achieves a stationary point for the problem. Then, BS selection is addressed with the aim of reducing the number of active BSs, thus enhancing the energy efficiency of 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 that is shown to converge to a locally optimal point. From numerical results, it is observed that the proposed robust compression scheme compensates for a large fraction of the performance loss induced by the imperfect statistical information. Moreover, the proposed BS selection algorithm is seen to perform close to the more complex exhaustive search solution.**

***Index Terms*—Cloud radio access networks, distributed source coding, multicell processing.**

I. Introduction

THE CURRENT deployments of cellular systems are facing the bandwidth crunch problem caused by the ever-increasing demand for high-data-rate applications. An integral part of many proposed solutions to this problem is the idea of cloud radio access networks, whereby the baseband processing of the base stations (BSs) is migrated to a central unit in the cloud to which the BSs are connected via finite-capacity backhaul links [1]–[7]. Cloud radio access networks can be seen as an effective implementation of the idea of multicell processing, which is also referred to as network multiple-input multiple-output (MIMO), which has been widely investigated as summarized in [8]. The promises of network MIMO and, thus of cloud radio access networks, include energy efficiency, load balancing, and capacity improvement due to the cooperative processing of the signals received/transmitted by distributed BSs, while requiring an efficient data sharing strategy between the BSs and the cloud on the capacity-constrained backhaul links [4].

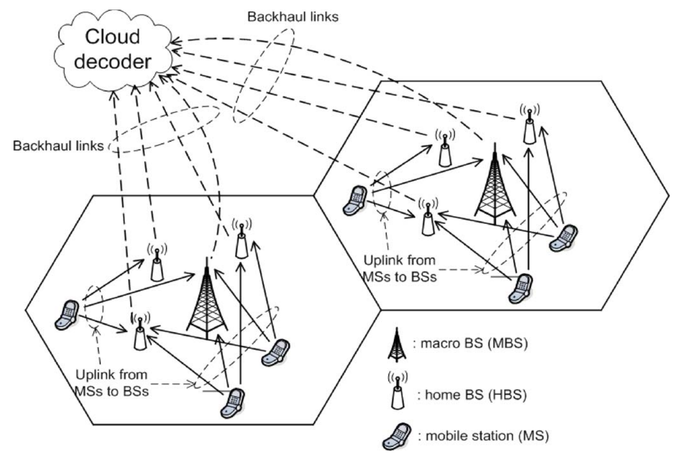


Fig. 1. Uplink of a cloud radio access networks with BSs classified as HBSs and MBSs.

On the uplink of a cloud radio access network, the BSs operate as terminals that relay "soft" information to the cloud decoder regarding the received baseband signals (see Fig. 1). Since the signals received at different BSs are correlated, distributed source coding strategies are generally beneficial, as first demonstrated in [9]. In this paper, we study the problem of compression with distributed source coding in the presence of multiantenna BSs by focusing on the issues of *robustness* and *efficiency*, as discussed in the following.

A. Contributions and Related Work

Related works on uplink multicell processing with constrained backhaul can be found in [9]–[14] and references therein. In [9], compress-and-forward strategies with joint decompression and decoding were proposed for the uplink

Manuscript received June 15, 2012; revised August 25, 2012; accepted October 13, 2012. Date of publication November 12, 2012; date of current version February 12, 2013. This work was supported by InterDigital Inc. This paper was presented in part at the IEEE Information Theory Workshop 2012, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, September 3–7, and in part at the 2012 IEEE Global Communications Conference, Disneyland Hotel, Anaheim, CA, December 3–7. The review of this paper was coordinated by Dr. X. Wang.

S.-H. Park and O. Simeone are with the Center for Wireless Communications and Signal Processing Research, Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: seok-hwan.park@njit.edu; osvaldo.simeone@njit.edu).

O. Sahin is with InterDigital Inc., Melville, NY 11747 USA (e-mail: Onur.Sahin@interdigital.com).

S. Shamai is with the Department of Electrical Engineering, Technion Israel Institute of Technology, Haifa 32000, Israel (e-mail: sshlomo@ee.technion.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2012.2226945

of a Wyner model with single-antenna terminals. The setup with multiple antennas but with a single transmitter is studied in [10]. Various relaying schemes, including decode-and-forward and compress-and-forward techniques, were compared in [11]. In [12], a different relaying scheme was proposed, which is referred to as compute-and-forward, in which structured codes are used by mobile stations (MSs) and the BSs decode a function of the transmitted messages. An efficient implementation of the compute-and-forward scheme, which is referred to as quantized compute-and-forward, was proposed in [13].

In this paper, we focus on the distributed compression of the received signals at multiple multiantenna BSs. Distributed compression can be implemented via sequential source coding with side information (SI) [15]: At each step of this process, a BS compresses its received signal by leveraging the available statistical information about the signals compressed by the BSs active at the previous steps.[1] The problem of compressing Gaussian random vector signals (as with a multiantenna receiver in the presence of Gaussian codebooks and Gaussian noise) with receiver SI has been investigated in [16]–[20]. Specifically, it was shown in [16] and [17] that independent coding of the signals obtained at the output of the so-called conditional Karhunen–Loeve transform (KLT) achieves the optimal performance in terms of minimizing the mean square error (MMSE) and maximizing achievable rate (Max-Rate), respectively (a review of these techniques can be found in [21]).

In this paper, we focus on the design of distributed compression strategies for the uplink of a cloud radio access cellular network that employs sequential source coding with SI. First, we point out the connection with the so-called *information bottleneck problem* [19], [22] and some consequence of this observation (see Section III). Then, we observe that the performance of distributed source coding is very sensitive to errors in the knowledge of the joint statistics of the received signals at the BSs due to the potential inability of the cloud decoder to decompress the signal received by a BS. This is because distributed source coding is based on the idea of reducing the rate of the compressed stream by introducing some uncertainty on the compressed signal that is resolved with the aid of the SI [23]. The amount of rate reduction that is allowed without incurring decompression errors thus depends critically on the quality of the SI, which should be known to the encoder.

Motivated by this observation, in Section IV, we propose a *robust compression scheme* by assuming the knowledge of the joint statistics, which amount here to a covariance matrix, available at each BS is imperfect. To model the uncertainty, we adopt a deterministic additive error model with bounds on eigenvalues of the error matrix similar to [24]–[26] (see also [27]). We remark that bounding the eigenvalues is equivalent to bounding any norm of the error [28, App. A]. The problem is formulated following a deterministic worst case approach, and a solution that achieves a stationary point of this problem is provided by solving Karush–Kuhn–Tucker (KKT) conditions [28], [29], which are also shown to be necessary for optimality.

We then tackle the issue of *energy-efficient network operation* via scheduling of the BSs. This problem was studied for a single-antenna uplink system in [30] without accounting for the effect of intercell interference and for joint decoding at the central unit. In Section V, instead, an optimization problem is formulated in which compression and BS scheduling are performed jointly for the given cloud radio access setup, by introducing a sparsity-inducing term into the objective function. This approach is inspired by the strategy proposed in [31] for the design of beamforming vectors for a multiantenna downlink system. An iterative block-coordinate ascent algorithm is proposed that is shown to converge to a locally optimal point. We conclude this paper with numerical results in Section VI.

Notation: We use the same notation for probability mass functions (pmfs) and probability density functions (pdfs), namely $p(x)$ represents the distribution, either pmf or pdf, of random variable X. Similar notations are used for joint and conditional distributions. All logarithms are in base two unless specified. Given vector $\mathbf{x} = [x_1, x_2 \ldots, x_n]^T$, we define $\mathbf{x}_{\mathcal{S}}$ for subset $\mathcal{S} \subseteq \{1, 2, \ldots, n\}$ 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., $\mathbf{\Sigma}_{\mathbf{x}} = E[\mathbf{x}\mathbf{x}^\dagger]$; $\mathbf{\Sigma}_{\mathbf{xy}}$ represents the cross-correlation matrix $\mathbf{\Sigma}_{\mathbf{xy}} = \mathrm{E}[\mathbf{x}\mathbf{y}^\dagger]$; and $\mathbf{\Sigma}_{\mathbf{x}|\mathbf{y}}$ represents the "conditional" correlation matrix of $\mathbf{x}$ given $\mathbf{y}$, namely $\mathbf{\Sigma}_{\mathbf{x}|\mathbf{y}} = \mathbf{\Sigma}_{\mathbf{x}} - \mathbf{\Sigma}_{\mathbf{xy}}\mathbf{\Sigma}_{\mathbf{y}}^{-1}\mathbf{\Sigma}_{\mathbf{xy}}^\dagger$. Notation $\mathcal{H}^n$ represents the set of all $n \times n$ Hermitian matrices. The circularly symmetric complex Gaussian distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{R}$ is denoted by $\mathcal{CN}(\mathbf{m}, \mathbf{R})$. Given vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$, we define $\mathbf{x}_{\mathcal{S}}$ for subset $\mathcal{S} \subseteq \{1, \ldots, m\}$ as the vector including, in ascending order, the vectors $\mathbf{x}_i$ with $i \in \mathcal{S}$. Similarly, given matrices $\mathbf{X}_1, \ldots, \mathbf{X}_m$, we denote by $\mathbf{X}_{\mathcal{S}}$ the matrix obtained by stacking the matrices $\mathbf{X}_i$ with $i \in \mathcal{S}$ vertically in ascending order.

II. System Model

We consider a *cluster* of cells,[2] which includes a total number N_B of BSs, each being either a macro BS (MBS) or a Home BS (HBS), and there are N_M active MSs (see Fig. 1). We denote the set of all BSs as $\mathcal{N}_B = \{1, \ldots, N_B\}$ and define π as the permutation of the set $\mathcal{N}_B$ with $\pi(i)$ standing for the ith element of π. Each ith BS is connected to the cloud decoder via a finite-capacity link of capacity C_i and has $n_{B,i}$ antennas, whereas each MS has $n_{M,i}$ antennas. Throughout this paper, we focus on the uplink.

Defining $\mathbf{H}_{ij}$ as the $n_{B,i} \times n_{M,j}$ complex 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, i.e.,

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

[1]This argument assumes, as done throughout this paper, that the cloud decoder performs decompression of the received signals and decoding separately. For analysis of joint decompression and decoding for single-antenna BSs, see [9] (see also Section VII).

[2]The model applies also to a cluster of sectors.

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 (CU) of a given time slot, 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 with $\mathbf{x}_i$ being the $n_{M,i} \times 1$ vector of symbols transmitted by MS i. The noise vectors $\mathbf{z}_i$ are independent over i and are distributed as $\mathbf{z}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, for $i \in \{1, \ldots, N_B\}$. Note that the noise covariance matrix is selected as the identity without loss of generality since the received signal can always be whitened by the BSs. 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.

Using standard random coding arguments [32], the coding strategies employed by the MSs in each time slot entail distribution $p(\mathbf{x})$ on the transmitted signals that factorizes as

$$p(\mathbf{x}) = \prod_{i=1}^{N_M} p(\mathbf{x}_i) \tag{3}$$

since the signals sent by different MSs are independent. Note that signals $\mathbf{x}$ are typically discrete, e.g., taken from discrete constellation, but can be well approximated by continuous (e.g., Gaussian) distributions for capacity-achieving codes over Gaussian channels [33]. If not stated otherwise, we will thus assume throughout that the distribution $p(\mathbf{x}_i)$ of the signal transmitted by the ith MS is given as $\mathbf{x}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{x}_i})$ 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. We consider compression strategies that do not require the BSs to know the codebooks employed by the MSs [9]. Using conventional rate–distortion theory arguments, a compression strategy for the ith BS is described by test channel $p(\hat{\mathbf{y}}_i|\mathbf{y}_i)$ that describes the relationship between the signal to be compressed, namely $\mathbf{y}_i$, and its description $\hat{\mathbf{y}}_i$ of size $n_{B,i} \times 1$ to be communicated to the cloud (see, e.g., [23]). It is recalled that such a description is limited to C_i bits per received symbol and that decoding at the cloud is based on the received descriptions $\hat{\mathbf{y}}_i$ for $i \in \mathcal{N}_\mathcal{B}$. Specifically, the cloud decoder performs joint decoding of signals $\mathbf{x}$ transmitted by all MSs so that, from standard information-theoretic considerations, the achievable sum rate is given by

$$R_{sum} = I(\mathbf{x}; \hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}}). \tag{4}$$

Since signals $\mathbf{y}_i$ measured by different BSs are correlated, distributed source coding techniques have the potential to improve the quality of the descriptions $\hat{\mathbf{y}}_i$ [9]. An efficient way to implement distributed source coding is via successive quantization [34].[3] Accordingly, one fixes a permutation π of the indices of the BSs. Then, the description $\hat{\mathbf{y}}_i$ for $i \in \mathcal{N}_\mathcal{B}$ can be successively recovered at the cloud as long as the backhaul capacities C_i for $i \in \mathcal{N}_\mathcal{B}$ satisfy the following conditions:

$$I\left(\mathbf{y}_{\pi(i)}; \hat{\mathbf{y}}_{\pi(i)} | \hat{\mathbf{y}}_{\{\pi(1),\ldots,\pi(i-1)\}}\right) \leq C_{\pi(i)} \tag{5}$$

for all $i = 1, \ldots, N_B$ [34].

Remark 1: The Gaussian distribution assumed here for the transmitted signals $\mathbf{x}$ maximizes the capacity of a Gaussian channel but is not, in general, optimal for the setting at hand, in which the receiver observes a compressed version of the received signal, even in a system with only one BS $(N_B = 1)$ [9].

III. Maximizing the Sum Rate

Here, we first discuss a greedy approach to find a suboptimal solution to the problem of maximizing the sum rate (4). Then, we will focus on the main step of this greedy procedure by reviewing known results and pointing out some new observation along the way.

A. Problem Definition and Greedy Solution

Here, we aim at maximizing the sum rate (4) under the constraints (5), i.e.,

$$\begin{aligned} \underset{\pi, \{p(\hat{\mathbf{y}}_i|\mathbf{y}_i)\}_{i\in\mathcal{N}_B}}{\text{maximize}} \quad & I(\mathbf{x}; \hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}}) \\ \text{s.t.} \quad & I\left(\mathbf{y}_{\pi(i)}; \hat{\mathbf{y}}_{\pi(i)} | \hat{\mathbf{y}}_{\{\pi(1),\ldots,\pi(i-1)\}}\right) \leq C_{\pi(i)} \\ & \text{for all } i = 1, \ldots, N_B \end{aligned} \tag{6}$$

where the optimization space includes also the BS permutation π.

The optimization (6) is generally still complex since it requires an exhaustive search over all possible permutations of the order of the BSs, which requires a search of size $N_B!$. We thus propose a greedy approach in Algorithm 1 to the selection of the permutation π while optimizing the test channels $p(\hat{\mathbf{y}}_i|\mathbf{y}_i)$ at each step of the greedy algorithm. The greedy algorithm is based on the chain rule for the mutual information that allows the sum rate (4) to be written as

$$I(\mathbf{x}; \hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}}) = \sum_{i=1}^{N_B} I\left(\mathbf{x}; \hat{\mathbf{y}}_{\pi(i)} | \hat{\mathbf{y}}_{\{\pi(1),\ldots,\pi(i-1)\}}\right) \tag{7}$$

for any permutation π of set $\{1, \ldots, N_B\}$. As a result of the algorithm, we obtain permutation π^* and feasible [in the sense of satisfying constraint (5)] test channels $p^*(\hat{\mathbf{y}}_i|\mathbf{y}_i)$. From now on, we refer to the compression based on (8) in Algorithm 1 as Max-Rate compression.

Algorithm 1 Greedy approach to the selection of the ordering π and the test channels $p(\hat{\mathbf{y}}_i|\mathbf{y}_i)$

1. Initialize set $\mathcal{S}$ to be an empty set, i.e., $\mathcal{S}^{(0)} = \varnothing$.
2. For $j = 1, \ldots, N_B$, perform the following steps.

[3] A more general successive quantization approach was proposed in [15] based on the idea of source splitting. While this approach may lead to performance gains, it is not investigated further here.

i) Each ith BS with $i \in \mathcal{N}_B - \mathcal{S}$ evaluates the test channel $p(\hat{\mathbf{y}}_i|\mathbf{y}_i)$ by solving the problem, i.e.,

$$\begin{aligned}&\underset{p(\hat{\mathbf{y}}_i|\mathbf{y}_i)}{\text{maximize}}\ I(\mathbf{x};\hat{\mathbf{y}}_i|\hat{\mathbf{y}}_{\mathcal{S}})\\ &\text{s.t. } I(\mathbf{y}_i;\hat{\mathbf{y}}_i|\hat{\mathbf{y}}_{\mathcal{S}}) \leq C_i.\end{aligned} \tag{8}$$

Denote the optimal value of this problem as $\phi_i^\star$ and an optimal test channel as $p^*(\hat{\mathbf{y}}_i|\mathbf{y}_i)$.

ii) Choose the BS $i \in \mathcal{N}_B - \mathcal{S}$ with the largest optimal value $\phi_i^\star$ and add it to the set $\mathcal{S}$ i.e., $\mathcal{S}^{(j)} = \mathcal{S}^{(j-1)} \cup \{i\}$ and set $\pi^*(j) = i$.

Remark 2: The implementation of the greedy algorithm in Algorithm 1 requires solving the problem (8) for each ith BS for a given order π. In practice, problem (8) can be solved at the cloud decoder, which then communicates the result to the ith BS. As it will be further clarified in the following, this approach requires the cloud center to know the channel matrices $\mathbf{H}_i$, $i = 1, \ldots, N_B$. Alternatively, once the order π is fixed by the cloud, problem (8) can be solved at each ith BS. This second approach requires the cloud to communicate some information to each BS, as shown in the following.

B. Max-Rate Compression

We now discuss the solution to the problem (8) of optimizing the compression test channel $p(\hat{\mathbf{y}}_i|\mathbf{y}_i)$ at the ith BS under the assumption that the cloud decoder has SI $\hat{\mathbf{y}}_{\mathcal{S}}$ with $\mathcal{S} = \{\pi(1), \ldots, \pi(i-1)\}$. Note that the random vectors involved in problem (8) satisfy the Markov chain $\hat{\mathbf{y}}_{\mathcal{S}} \leftrightarrow \mathbf{x} \leftrightarrow \mathbf{y}_i \leftrightarrow \hat{\mathbf{y}}_i$. We first review the solution of problem (8) given in [17] and [18]. We also point out the relationship of the solution found in [17] and [18] with the information bottleneck method for Gaussian variables of [19]. This connection does not seem to have been observed before and allows for a solution of problem (8) in the presence of a generic discrete distribution (3) of the transmitted signals, as briefly discussed in Remark 4. A more thorough discussion of this background material can be found in [21]. To describe the optimal solution to problem (8), we first define the covariance matrix as follows:

$$\mathbf{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_{\mathcal{S}}} = \mathbf{\Sigma}_{\mathbf{x}} - \mathbf{\Sigma}_{\mathbf{x}}\bar{\mathbf{H}}_{\mathcal{S}}^\dagger\left(\bar{\mathbf{H}}_{\mathcal{S}}\mathbf{\Sigma}_{\mathbf{x}}\bar{\mathbf{H}}_{\mathcal{S}}^\dagger + \mathbf{\Sigma}_{\mathbf{t}_{\mathcal{S}}}\right)^{-1}\bar{\mathbf{H}}_{\mathcal{S}}\mathbf{\Sigma}_{\mathbf{x}} \tag{9}$$

where $\bar{\mathbf{H}}_j = \mathbf{A}_j\mathbf{H}_j$ and $\mathbf{\Sigma}_{\mathbf{t}_j} = \mathbf{A}_j\mathbf{A}_j^\dagger + \mathbf{I}$. Matrix $\mathbf{A}_j$ will be defined in Proposition 1. We then have the following result.

Proposition 1: An optimal solution $p(\hat{\mathbf{y}}_i|\mathbf{y}_i)$ to problem (8) is given by [17], [18]

$$\hat{\mathbf{y}}_i = \mathbf{A}_i\mathbf{y}_i + \mathbf{q}_i \tag{10}$$

where $\mathbf{q}_i \sim \mathcal{CN}(0, \mathbf{I})$ is the compression noise, which is independent of $\mathbf{x}$ and $\mathbf{z}_i$, and matrix $\mathbf{A}_i$ is such that $\mathbf{\Omega}_i = \mathbf{A}_i^\dagger\mathbf{A}_i$, with

$$\mathbf{\Omega}_i = \mathbf{U}\text{diag}(\alpha_1, \ldots, \alpha_{n_{B,i}})\mathbf{U}^\dagger. \tag{11}$$

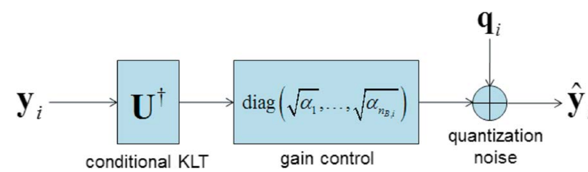


Fig. 2. Max-Rate compression solution. $\mathbf{U}$ is the conditional KLT [16], and $\mathbf{q}_i \sim \mathcal{CN}(0, \mathbf{I})$ represents the compression noise.

In (11), we have used the eigenvalue decomposition $\mathbf{H}_i\mathbf{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_{\mathcal{S}}}\mathbf{H}_i^\dagger + \mathbf{I} = \mathbf{U}\text{diag}(\lambda_1, \ldots, \lambda_{n_{B,i}})\mathbf{U}^\dagger$ with unitary $\mathbf{U}$ and ordered eigenvalues $\lambda_1 \geq \cdots \geq \lambda_{n_{B,i}}$. The diagonal elements $\alpha_1, \ldots, \alpha_{n_{B,i}}$ are computed as

$$\alpha_l = \left[\frac{1}{\mu}\left(1 - \frac{1}{\lambda_l}\right) - 1\right]^+, \quad l = 1, \ldots, n_{B,i} \tag{12}$$

where μ is such that condition $\sum_{l=1}^{n_{B,i}} \log(1 + \alpha_l\lambda_l) = C_i$ is satisfied.

Remark 3: The linear transform $\mathbf{U}$ coincides with the conditional KLT derived in [16]. We observe that, in [16], the goal is that of minimizing the MSE and not maximizing the rate. The same transformation thus turns out to be optimal under both criteria. The conditional KLT $\mathbf{U}$ has the effect of making the components of the output conditionally uncorrelated given the SI $\hat{\mathbf{y}}_{\mathcal{S}}$. Therefore, intuitively, one can then compress each element of the output independently without loss of optimality. This is done by selecting the compression gains $\alpha_1, \ldots, \alpha_{n_{B,i}}$. It is observed that the optimal choice of these gains is different, depending on whether one adopts the MMSE criterion as in [16] or the Max-Rate criterion as done here (see further discussion in [17]). An illustration of the optimal Max-Rate compression is shown in Fig. 2.

Remark 4: An alternative formulation of the optimal solution of Proposition 1 can be obtained using the results in [19, Th. 3.1]. In fact, problem (8) can be interpreted as an instance of the information bottleneck problem, which was introduced in [22]. We recall that the information bottleneck method consists in the maximization of $I(\mathbf{x};\hat{\mathbf{y}}) - 1/\beta I(\mathbf{y};\hat{\mathbf{y}})$ for a given $\beta > 0$ over the conditional distribution $p(\hat{\mathbf{y}}|\mathbf{y})$ for random variables $\mathbf{x}$, $\mathbf{y}$, and $\hat{\mathbf{y}}$ satisfying the Markov chain $\mathbf{x} \leftrightarrow \mathbf{y} \leftrightarrow \hat{\mathbf{y}}$. This connection between the information bottleneck problem and that of compression for the cloud radio access network allows us to import tools developed for the information bottleneck problem to the setup at hand [22]. For instance, to solve problem (8) in the case where the distribution of the transmitted symbols (3) is not Gaussian but discrete, it may be advantageous to use a discrete alphabet for $\hat{\mathbf{y}}_i$. Assuming this alphabet to be given, the information bottleneck approach enables us to maximize $I(\mathbf{x};\hat{\mathbf{y}}_i|\hat{\mathbf{y}}_{\mathcal{S}}) - 1/\beta I(\mathbf{y}_i;\hat{\mathbf{y}}_i|\hat{\mathbf{y}}_{\mathcal{S}})$ or equivalently to minimize $I(\mathbf{y}_i;\hat{\mathbf{y}}_i|\hat{\mathbf{y}}_{\mathcal{S}}) - \beta I(\mathbf{x};\hat{\mathbf{y}}_i|\hat{\mathbf{y}}_{\mathcal{S}})$, over the pmf $p(\hat{\mathbf{y}}_i|\mathbf{y}_i)$ for some Lagrange multiplier $\beta > 0$. To this end, an iterative algorithm from [22, Sec. 3.3] can be adopted with minor modifications, which are due to the conditioning on $\hat{\mathbf{y}}_{\mathcal{S}}$ that does not appear in [22]. We recall that the key idea of the algorithm is to maximize the objective function over the three distributions $p(\hat{\mathbf{y}}_i|\hat{\mathbf{y}}_{\mathcal{S}})$, $p(\hat{\mathbf{y}}_i|\hat{\mathbf{y}}_i)$, and $p(\mathbf{x}|\hat{\mathbf{y}}_i, \hat{\mathbf{y}}_{\mathcal{S}})$ in turn. Since the objective function can be seen to be a concave function in the domain of the three distributions, an iterative block-coordinate ascent algorithm is known to converge to a locally optimal solution [29]. We do not detail further this approach here.

IV. Robust Optimal Compression

As shown earlier, the optimal compression resulting from the solution of problem (8) at the ith BS depends on covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$ in (9) of the vector of transmitted signals conditioned on the compressed version $\hat{\mathbf{y}}_\mathcal{S}$ of the signals received by the BSs in set $\mathcal{S}$. In general, when solving (8), particularly in the case in which the optimization is done at the ith BS (see Remark 2), it might not be realistic to assume that matrix $\boldsymbol{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$ is perfectly known since it depends on the channel matrices of all the BSs in the set $\mathcal{S}$ [see (9)]. Motivated by this observation, we propose a robust version of the optimization problem (8) by assuming that, when solving problem (8), only an estimate of $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$ is available that is related to the actual matrix $\boldsymbol{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$ as

$$\boldsymbol{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} + \boldsymbol{\Delta}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} \tag{13}$$

where $\boldsymbol{\Delta}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} \in \mathcal{H}^{n_M}$ is a deterministic Hermitian matrix that models the estimation error. We assume that error matrix $\boldsymbol{\Delta}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$ is only known to belong to set $\mathcal{U}_{\boldsymbol{\Delta}} \subseteq \mathcal{H}^{n_M}$, which models the uncertainty at the ith BS regarding matrix $\boldsymbol{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$. Using (9), it can be seen that the additive uncertainty model (13) arises, for instance, if only an estimate $\hat{\bar{\mathbf{H}}}_\mathcal{S} = \bar{\mathbf{H}}_\mathcal{S} + \boldsymbol{\Delta}$ of the channel matrix $\bar{\mathbf{H}}_\mathcal{S}$ relative to the other BSs is available, where $\boldsymbol{\Delta}$ accounts for the estimation error, as long as the eigenvalues of $\boldsymbol{\Delta}$ are sufficiently small.

In general, to define the uncertainty set $\mathcal{U}_{\boldsymbol{\Delta}}$, one can impose some bounds on given measures of the eigenvalues and/or eigenvectors of matrix $\boldsymbol{\Delta}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$. Based on the observation that the mutual information $I(\mathbf{x};\hat{\mathbf{y}}_i|\hat{\mathbf{y}}_\mathcal{S})$ can be written as

$$I(\mathbf{x};\hat{\mathbf{y}}_i|\hat{\mathbf{y}}_\mathcal{S}) = f_i(\boldsymbol{\Omega}_i, \tilde{\boldsymbol{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}) - \log\det(\mathbf{I}+\boldsymbol{\Omega}_i) \tag{14}$$

where we have defined here as $f_i(\boldsymbol{\Omega}_i, \tilde{\boldsymbol{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}) = \log\det(\mathbf{I}+\boldsymbol{\Omega}_i(\mathbf{H}_i\hat{\boldsymbol{\Sigma}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}\mathbf{H}_i^\dagger + \tilde{\boldsymbol{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} + \mathbf{I}))$ and $\tilde{\boldsymbol{\Delta}}_{\mathbf{x}|\mathbf{y}_\mathcal{S}} = \mathbf{H}_i\boldsymbol{\Delta}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}\mathbf{H}_i^\dagger$, we take the approach of bounding the uncertainty on the eigenvalues of $\tilde{\boldsymbol{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$. This is equivalent to bounding, within some constant, any norm of matrix $\tilde{\boldsymbol{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$ (see, e.g., [28, App. A]). Specifically, we define the uncertainty set $\mathcal{U}_{\boldsymbol{\Delta}}$ as the set of Hermitian matrices $\tilde{\boldsymbol{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$ such that the following conditions:

$$\lambda_{\min}(\tilde{\boldsymbol{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}) \geq \lambda_{\mathrm{LB}} \quad \text{and} \quad \lambda_{\max}(\tilde{\boldsymbol{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}) \leq \lambda_{\mathrm{UB}} \tag{15}$$

hold for given lower and upper bounds[4] $(\lambda_{\mathrm{LB}}, \lambda_{\mathrm{UB}})$ on the eigenvalues of matrix $\tilde{\boldsymbol{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$.

[4] Note that, when using the additive model (13), a nontrivial lower bound λ_{LB} must satisfy $\lambda_{\mathrm{LB}} \geq \lambda_{\min}(\mathbf{H}_i\hat{\boldsymbol{\Sigma}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}\mathbf{H}_i^\dagger)$ to guarantee that matrix $\mathbf{H}_i\boldsymbol{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}\mathbf{H}_i^\dagger$ is positive semidefinite.

Under this model, the problem of deriving the optimal robust compression strategy can be formulated as

$$\begin{aligned}&\underset{\boldsymbol{\Omega}_i\succeq\mathbf{0}}{\text{maximize}} \min_{\tilde{\boldsymbol{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}\in\mathcal{H}^{n_M}} f_i(\boldsymbol{\Omega}_i, \tilde{\boldsymbol{\Delta}}_{x|\hat{y}_\mathcal{S}}) - \log\det(\mathbf{I}+\boldsymbol{\Omega}_i)\\ &\text{s.t.} \begin{cases} f_i(\boldsymbol{\Omega}_i, \tilde{\boldsymbol{\Delta}}_{x|\hat{y}_\mathcal{S}}) \leq C_i \\ \lambda_{\min}(\tilde{\boldsymbol{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}) \geq \lambda_{\mathrm{LB}} \\ \lambda_{\max}(\tilde{\boldsymbol{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}) \leq \lambda_{\mathrm{UB}}. \end{cases}\end{aligned} \tag{16}$$

Problem (16) is not convex, and a closed-form solution appears prohibitive. Theorem 1 derives a solution to the KKT conditions for problem (16), which is also referred to as a stationary point. It is shown in Appendix A that the KKT conditions are necessary for the optimality of problem (16). To state the main result compactly, let us define the following scalar functions:

$$g_i^L(\alpha_1,\ldots,\alpha_{n_{B,i}}) = \sum_{l=1}^{n_{B,i}} \log\left(1+\alpha_l c_l^L\right)$$

$$\text{and} \quad g_i^U(\alpha_1,\ldots,\alpha_{n_{B,i}}) = \sum_{l=1}^{n_{B,i}} \log\left(1+\alpha_l c_l^U\right)$$

with $c_l^L = \lambda_l + \lambda_{\mathrm{LB}}$ and $c_l^U = \lambda_l + \lambda_{\mathrm{UB}}$, and the discrete set

$$\mathcal{P}_l(\mu) = \begin{cases} \{0\}, & \text{if } Q_l \geq 0,\ S_l \geq 0 \\ \left\{\frac{-Q_l+\sqrt{Q_l^2-4S_l}}{2}\right\}, & \text{if } Q_l \geq 0,\ S_l < 0 \\ \left\{\frac{-Q_l\pm\sqrt{Q_l^2-4S_l}}{2}, 0\right\}, & \text{if } Q_l < 0,\ S_l \geq 0 \\ & \quad -\frac{Q_l^2}{4}+S_l \leq 0 \\ \{0\}, & \text{if } Q_l < 0,\ S_l \geq 0 \\ & \quad -\frac{Q_l^2}{4}+S_l > 0 \\ \left\{\frac{-Q_l+\sqrt{Q_l^2-4S_l}}{2}\right\}, & \text{if } Q_l < 0,\ S_l < 0 \end{cases} \tag{17}$$

with Q_l and S_l given as

$$Q_l = \frac{c_l^U\left(1+\mu+(\mu-1)c_l^L\right)}{\mu c_l^U c_l^L} \quad \text{and} \quad S_l = \frac{\mu c_l^U + 1 - c_l^L}{\mu c_l^U c_l^L}. \tag{18}$$

Theorem 1: A stationary point for problem (16) can be found as (10) with matrix $\mathbf{A}_i$ such that $\boldsymbol{\Omega}_i = \mathbf{A}_i^\dagger\mathbf{A}_i$ is given as (11), where matrix $\mathbf{U}$ is obtained from the eigenvalue decomposition $\mathbf{H}_i\hat{\boldsymbol{\Sigma}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}\mathbf{H}_i^\dagger + \mathbf{I} = \mathbf{U}\mathrm{diag}(\lambda_1,\ldots,\lambda_{n_{B,i}})\mathbf{U}^\dagger$, and the diagonal elements $\alpha_1,\ldots,\alpha_{n_{B,i}}$ are calculated by solving the following mixed integer problem:

$$\max_{\mu,\alpha_1\ldots,\alpha_{n_{B,i}}} g_i^L(\alpha_1,\ldots,\alpha_{n_{B,i}}) - \sum_{l=1}^{n_{B,i}}\log(1+\alpha_l) \tag{19}$$

$$\text{s.t. } 0 < \mu < 1 \tag{20a}$$

$$\alpha_l \in \mathcal{P}_l(\mu), \quad l = 1,\ldots,n_{B,i} \tag{20b}$$

$$g_i^U(\alpha_1,\ldots,\alpha_{n_{B,i}}) = C_i. \tag{20c}$$

Proof: See Appendix A. ■

Remark 5: The solution identified in Theorem 1 is based on the conditional KLT as in the case of perfect knowledge of matrix $\mathbf{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_S}$ (see Remark 3). However, here the conditional KLT is obtained using the nominal covariance matrix $\hat{\mathbf{\Sigma}}_{\mathbf{x}|\hat{\mathbf{y}}_S}$ to construct the matrix $\mathbf{H}_i\hat{\mathbf{\Sigma}}_{\mathbf{x}|\hat{\mathbf{y}}_S}\mathbf{H}_i^\dagger + \mathbf{I}$. Moreover, the selection of the compression gains $\alpha_1, \ldots, \alpha_{n_{B,i}}$ is more complex than in the ideal case of Proposition 1 and comes from the solution of the KKT conditions of problem (16) (see Appendix A for more details).

Remark 6: The optimal $\alpha_1, \ldots, \alpha_{n_{B,i}}$ for problem (19) can be found from an exhaustive scalar search over μ between 0 and 1. For each μ, the search for values α_l in problem (19) is restricted to the set $\mathcal{P}_l(\mu)$ that contains at most three elements for $l = 1, \ldots, n_{B,i}$.

In fact, the following corollary shows that, in some special case, the search over parameters α_l is not necessary since sets $\mathcal{P}_l(\mu)$ only contain one element.

Corollary 1: If $\lambda_{\mathrm{UB}} - \lambda_{\mathrm{LB}} < 1$, a stationary point for problem (16) is given by $\alpha_l = ([-Q_l + \sqrt{Q_l^2 - 4S_l}]^+)/2$ for $l = 1, \ldots, n_{B,i}$ with μ such that the constraint (20c) is satisfied.

Proof: If $\lambda_{\mathrm{UB}} - \lambda_{\mathrm{LB}} < 1$, α_l is computed from (17) as

$$\alpha_l = \begin{cases} \frac{-Q_l + \sqrt{Q_l^2 - 4S_l}}{2}, & \text{if } \mu < \frac{c_l^L - 1}{c_l^U} \\ 0, & \text{if } \mu \geq \frac{c_l^L - 1}{c_l^U} \end{cases} \tag{21}$$

which entails the claimed result by direct calculation. ■

Remark 7: If $\lambda_{\mathrm{UB}} = \lambda_{\mathrm{LB}} = 0$, the solution to problem (19) is unique and reduces to the solution (12) obtained if matrix $\mathbf{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_S}$ is perfectly known at BS i. This shows that the proposed robust solution reduces to the optimal design for the case in which $\mathbf{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_S}$ is perfectly known to BS i.

Proof: It follows by substituting $c_l^U = c_l^L = \lambda_l$ into (21). ■

Remark 8: As shown in Appendix A, any values of $\mu, \alpha_1, \ldots, \alpha_{n_{B,i}}$ that satisfy the constraints (20a)–(20c) are a solution to the KKT conditions for the robust problem (16).

V. Joint Base Station Selection and Compression via Sparsity-Inducing Optimization

To operate the network efficiently, it is generally advantageous to let only a subset $\mathcal{S} \subseteq \mathcal{N}_B$ of the N_B available BSs communicate to the cloud decoder in a given time slot. This is the case in scenarios in which different BSs share the same backhaul resources [35] or energy consumption and green networking are critical issues [30]. Therefore, under this assumption, the system design entails the choice of the subset $\mathcal{S}$, along with that of the compression test channels $p(\hat{\mathbf{y}}_i|\mathbf{y}_i)$ for $i \in \mathcal{S}$. In general, this BS selection requires an exhaustive search of exponential complexity in the number N_B of BSs. Here, inspired by Hong *et al.* [31], we propose an efficient approach based on the addition of a sparsity-inducing term to the objective function.

To elaborate, we associate cost q_i per spectral unit resource (i.e., per discrete-time CU) to the ith BS. This measures the relative cost per spectral resource of activating the ith BS over the revenue per bit. To highlight the ideas, consider a single cell with one MBS and $N_B - 1$ HBSs. We assume that the HBSs share the same total backhaul capacity C_H to the cloud decoder [35]. Assuming that the MBS is active, we are interested in selecting a subset of the HBSs to provide additional information to the cloud decoder under the given total backhaul constraint. Note that the solution proposed here can also be generalized to more complex systems with multiple cells.

Let $\mathcal{S}_\mathcal{M} = \{1\}$ and $\mathcal{S}_\mathcal{H} = \{2, \ldots, N_B\}$ denote the set that includes the MBS and the HBSs, respectively. Assuming that the Gaussian test channel (10) is employed at each ith BS with a given covariance matrix $\mathbf{\Omega}_i \succeq \mathbf{0}$, the joint problem of HBS selection and compression via sparsity-inducing optimization is formulated as

$$\begin{aligned} &\underset{\{\mathbf{\Omega}_i \succeq \mathbf{0}\}_{i \in \mathcal{S}_\mathcal{H}}}{\text{maximize}} \; I(\mathbf{x}; \hat{\mathbf{y}}_{\mathcal{S}_\mathcal{H}} | \hat{\mathbf{y}}_1) - q_H \sum_{i \in \mathcal{S}_\mathcal{H}} 1\left(\|\mathbf{\Omega}_i\|_F > 0\right) \\ &\text{s.t. } I(\mathbf{y}_{\mathcal{S}_\mathcal{H}}; \hat{\mathbf{y}}_{\mathcal{S}_\mathcal{H}} | \hat{\mathbf{y}}_1) \leq C_H \end{aligned} \tag{22}$$

where $1(\cdot)$ is the indicator function, which takes 1 if the argument state is true and 0 if otherwise, and we have assumed that $q_2 = \cdots = q_{N_B} = q_H$ for simplicity. In (22), we have conditioned on $\hat{\mathbf{y}}_1$ to account for the fact that the MBS is assumed to be active. Note that the second term in the objective of problem (22) is the ℓ_0-norm of vector $[\mathrm{tr}(\mathbf{\Omega}_2) \cdots \mathrm{tr}(\mathbf{\Omega}_{N_B})]$. If cost q_H is large enough, this term forces the solution to set some of matrices $\mathbf{\Omega}_i$ to zero, thus keeping the corresponding ith HBS inactive. To avoid the nonsmoothness induced by the ℓ_0-norm, we modify problem (22) by replacing the ℓ_0-norm with the ℓ_1-norm of the same vector. The rationale behind this approximation is that the ℓ_0-norm is reasonably well approximated by the ℓ_1-norm when it comes to identifying sparse patterns, as widely reported (see, e.g., [36]). We thus reformulate problem (22) as follows:

$$\begin{aligned} &\underset{\{\mathbf{\Omega}_i \succeq \mathbf{0}\}_{i \in \mathcal{S}_\mathcal{H}}}{\text{maximize}} \; f(\mathbf{\Omega}_2, \ldots, \mathbf{\Omega}_{N_B}) \\ &\text{s.t. } g(\mathbf{\Omega}_2, \ldots, \mathbf{\Omega}_{N_B}) \leq C_H \end{aligned} \tag{23}$$

where we have defined here as $f(\mathbf{\Omega}_2, \ldots, \mathbf{\Omega}_{N_B}) = I(\mathbf{x}; \hat{\mathbf{y}}_{\mathcal{S}_\mathcal{H}} | \hat{\mathbf{y}}_1) - q_H \sum_{i \in \mathcal{S}_\mathcal{H}} \mathrm{tr}(\mathbf{\Omega}_i)$ 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 expansion for $f(\mathbf{\Omega}_2, \ldots, \mathbf{\Omega}_{N_B})$ and $g(\mathbf{\Omega}_2, \ldots, \mathbf{\Omega}_{N_B})$ as a function of $\mathbf{\Omega}_2, \ldots, \mathbf{\Omega}_{N_B}$ can be easily obtained and is not regarded now.

Algorithm 2 Two-Phase Joint HBS Selection and Compression Algorithm

---

Phase 1. Solve problem (23) via the block-coordinate ascent algorithm:

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

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

$$\begin{aligned} &\underset{\mathbf{\Omega}_i \succeq \mathbf{0}}{\text{maximize}} \; f\left(\mathbf{\Omega}_i, \mathbf{\Omega}_{\{2,\ldots,i-1\}}^{(n)}, \mathbf{\Omega}_{\{i+1,\ldots,N_B\}}^{(n-1)}\right) \\ &\text{s.t. } g\left(\mathbf{\Omega}_i, \mathbf{\Omega}_{\{2,\ldots,i-1\}}^{(n)}, \mathbf{\Omega}_{\{i+1,\ldots,N_B\}}^{(n-1)}\right) \leq C_H. \end{aligned} \tag{24}$$

iii) Repeat step (ii) if some convergence criterion is not satisfied and stop if otherwise. 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 (23) with $q_H = 0$ and consider only the HBSs in set $\mathcal{S}_{\mathcal{H}}^*$.

Based on this formulation, we propose a two-phase approach to the problem of joint HBS selection and compression in Algorithm 2. As shown in the table, in the first phase, we execute the block-coordinate ascent algorithm [29] to tackle problem (23). In the block-coordinate ascent algorithm, we iteratively optimize $\mathbf{\Omega}_i$ for fixed other variables. 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 only on the subset $\mathcal{S}_{\mathcal{H}}^*$ 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 [17, Sec. IV].

It remains to discuss how to solve problem (24) at step (ii) of the proposed algorithm. Note that this corresponds to the update of $\mathbf{\Omega}_i$ when all the other variables $\mathbf{\Omega}_{\mathcal{S}_{\mathcal{H}}\setminus\{i\}}$ are fixed to the values obtained from the earlier iterations. The global maximum of problem (24) can be obtained, as shown in Theorem 2.

Theorem 2: A solution to problem (24) is given by (10), with matrix $\mathbf{A}_i$ such that $\mathbf{\Omega}_i = \mathbf{A}_i^\dagger \mathbf{A}_i$ is given as (11), where we have the eigenvalue decomposition $\mathbf{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}} = \mathbf{U}\mathrm{diag}(\lambda_1, \ldots, \lambda_{n_{B,i}})\mathbf{U}^\dagger$, and matrix $\mathbf{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}}$ is given as

$$\mathbf{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}} = \mathbf{I} + \mathbf{H}_i \mathbf{R}_i^{-1} \mathbf{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_1} \mathbf{H}_i^\dagger \tag{25}$$

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

$$\alpha_l(\mu) = \frac{\left[-a_l(\mu) + \sqrt{a_l(\mu)^2 - b_l(\mu)}\right]^+}{2q'_H \lambda_l} \tag{26}$$

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 $b_l(\mu) = 4q'_H \lambda_l((\mu - 1)\lambda_l + q'_H + 1)$. The Lagrange multiplier μ^* is obtained as follows: If $h_i(0) \leq \bar{C}_i$, where $\bar{C}_i$ is given by

$$\bar{C}_i = C_i - \log\det \mathbf{R}_i - \sum_{j\in\mathcal{S}_{\mathcal{H}\setminus\{i\}}} \log\det(\mathbf{I} + \mathbf{\Omega}_j) \tag{27}$$

then $\mu^* = 0$; otherwise, μ^* is unique value $\mu \geq 0$ 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 B. ■

Remark 9: The key difference between problems (8) and (24) is the sparsity-inducing term $q_H \sum_{i\in\mathcal{S}_\mathcal{H}} \mathrm{tr}(\mathbf{\Omega}_i)$ in the objective function in (24). The presence of this term explains the difference between (12) and (26). Moreover, if $q_H \to 0$, the solution (26) reduces to (12) derived in [17] by L'Hopital's rule.

Remark 10: We note that $\mathbf{\Omega}_i$ tends to zero; thus, the ith BS becomes inactive if the cost q'_H is large enough since, in (26), we have $\alpha_l = 0$ if $(\mu - 1)\lambda_l + q'_H + 1 \geq 0$.

VI. Numerical Results

Here, we present numerical results to validate and complement the analysis. It is assumed that the considered cell has radius R_{cell}, a single MBS, and multiple HBSs. The MBS is located at the cell center, whereas the HBS and MS are randomly dropped within a circular cell according to uniform distribution. All channel elements of $\mathbf{H}_{i,j}$ are independent identically distributed 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 the cell radius, i.e., $D_0 = R_{\mathrm{cell}}/2$. For simplicity, each MS uses a single antenna, i.e., $n_{M,i} = 1$ with transmit power P_{tx}, such that the aggregated transmit vector $\mathbf{x}$ has a covariance of $\mathbf{\Sigma}_\mathbf{x} = P_{\mathrm{tx}}\mathbf{I}$. Then, the signal-to-noise ratio is defined as P_{tx} since we have assumed unit variance noise in Section II. We remark that more extensive numerical results can be found in [21].

Robust Compression: We first present numerical results for the case in which there is uncertainty on the conditional covariance matrix $\mathbf{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$. According to the considered uncertainty model, matrix $\hat{\mathbf{\Sigma}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$ is assumed to be known at the ith BS with uncertainty matrix $\tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} = \mathbf{H}_i \mathbf{\Delta}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} \mathbf{H}_i^\dagger$ (see Section IV), whose eigenvalues are limited in the range (15) for given bounds $(\lambda_{\mathrm{LB}}, \lambda_{\mathrm{UB}})$. We have generated the eigenvectors of $\tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$ randomly according to isotropic distribution on the column space of $\mathbf{H}_i$ and the eigenvalues uniformly in the set (15), where $\lambda_{\mathrm{UB}} = \lambda_{\min}(\mathbf{H}_i \mathbf{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} \mathbf{H}_i^\dagger)$, and $\lambda_{\mathrm{LB}} = -\lambda_{\min}(\mathbf{H}_i \mathbf{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} \mathbf{H}_i^\dagger)$, respectively. This implies that the uncertainty on the eigenvalues is the maximum (symmetric) uncertainty consistent with the positive semidefiniteness of matrix $\mathbf{H}_i \hat{\mathbf{\Sigma}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} \mathbf{H}_i^\dagger$ (see Section IV). Moreover, it is assumed that the MBS is connected to the cloud via a backhaul link of capacity C, whereas the HBSs' backhaul is of capacity that is equal to a fraction of C, namely, ωC with $0 < \omega \leq 1$.

In Fig. 3, per-MS sum-rate performance is evaluated for a single cell with $N_B = 4$ (one MBS and three randomly placed HBSs), $N_M = 8$, $n_{B,i} = 2$, $\omega = 0.5$, and $P_{\mathrm{tx}} = 10$ dB versus the MBS backhaul capacity C. For reference, we plot the performance attained by a variation of the Max-Rate scheme that ignores SI[5] and that of a scheme that operates by assuming that $\hat{\mathbf{\Sigma}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$ is the true covariance matrix. Note that, in this case, the backhaul constraint (8) can be violated, which implies that the cloud decoder cannot recover the corresponding compressed signal $\hat{\mathbf{y}}_i$ (labeled as "imperfect SI" in the figure). The figure shows that assuming the incorrect matrix $\hat{\mathbf{\Sigma}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$ as being true can result in a severe performance degradation. However, this

[5]This is easily obtained from (11) and (12) by substituting the covariance matrix $\mathbf{\Sigma}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$ with $\mathbf{\Sigma}_\mathbf{x}$.

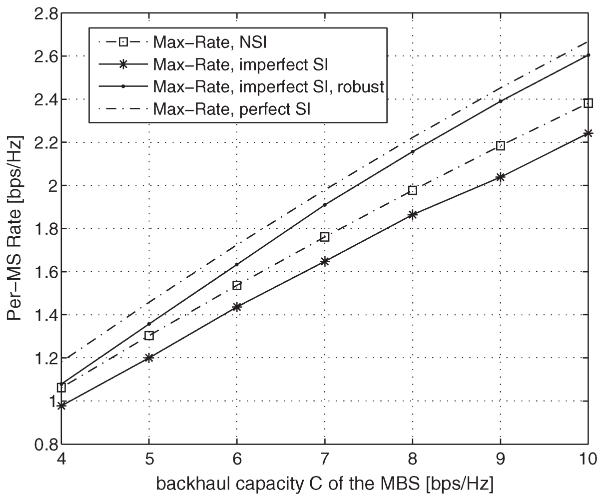


Fig. 3. Average per-MS sum rate versus the backhaul capacity C. The MBS obtained with the Max-Rate compression scheme in the presence of uncertainty for a single-cell heterogeneous network with $N_B = 4$, $N_M = 8$, $n_{B,i} = 2$, and $\omega = 0.5$ at $P_{\rm tx} = 10$ dB.

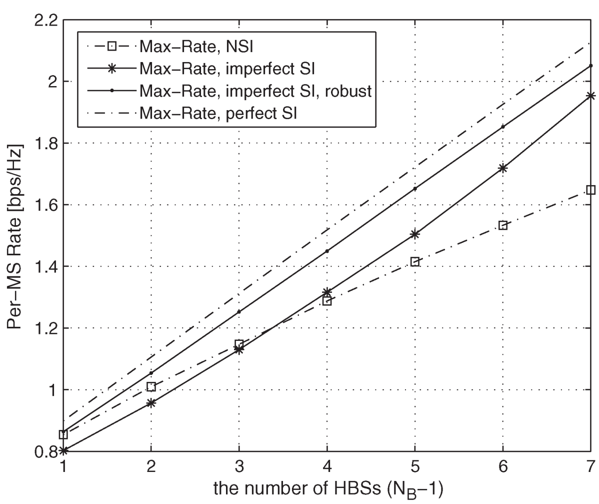


Fig. 4. Average per-MS sum rate versus the number $N_B - 1$ of HBSs obtained with the Max-Rate compression scheme in the presence of uncertainty for a single-cell heterogeneous network with $N_M = 10$, $n_{B,i} = 8$, and $C = 7$ bps/Hz at $P_{\rm tx} = 0$ dB.

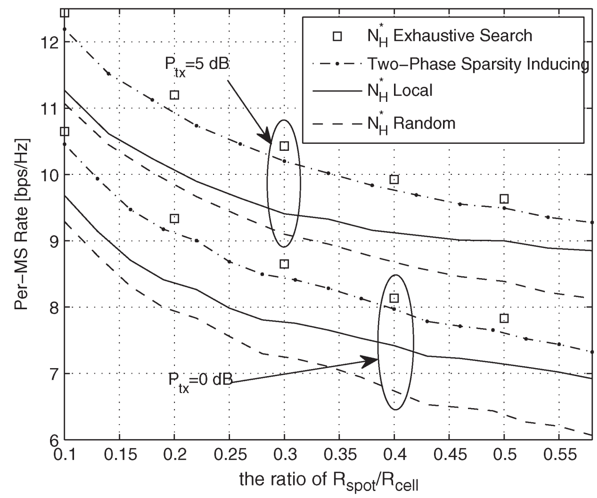


Fig. 5. Average per-MS sum rate versus the ratio $R_{\rm spot}/R_{\rm cell}$ in a single-cell heterogeneous network with $N_B = 13$ $(N_B^1 = 6, N_B^2 = 6)$, $N_M = 14$ $(N_M^1 = 8, N_M^2 = 6)$, $n_{B,i} = 8$, $C = 20$ bps/Hz, $C_H = 300$ bps/Hz, and $q_H = 100$.

performance loss can be overcome by adopting the proposed robust algorithm, which shows intermediate performance between the ideal setting with perfect SI and that with no SI. We also observe the more pronounced performance gain of the proposed robust solution for a larger backhaul link capacity. In a similar vein, in Fig. 4, we investigate the effect of the number $N_B - 1$ of HBSs. It is seen that, as the number of BSs grows, leveraging SI provides more relevant gains so that even assuming imperfect SI can be useful. As in Fig. 4, the proposed algorithm shows a rate performance close to the ideal case of perfect SI.

BS Selection: We now consider the single-cell setup of Section V, where we aim at scheduling a subset of the $N_B - 1$ HBSs under total backhaul constraint C_H. We compare the proposed two-phase approach (see Algorithm 2) with the following. First, the N_H^* *exhaustive search* is the scheme that selects N_H^* HBSs that maximize the sum rate via an 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. Second, the N_H^* *local* is the scheme that 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., $C_i^{\rm local} = \log\det(\mathbf{I} + \mathbf{H}_i\mathbf{\Sigma}_{\mathbf{x}}\mathbf{H}_i^\dagger)$ [23]. Note that this criterion is local in that it does not account for the correlation between the signals received by different HBSs. Finally, the N_H^* *Random* is the scheme that randomly selects N_H^* HBSs.

We consider a practical scenario in which 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, whereas those in group 2 are distributed in a smaller cell overlaid on the macrocell at hand and with radius $R_{\rm spot} < R_{\rm cell}$, which models a "*hot spot*" such as a building or a public space. Fig. 5 presents the per-MS sum rate versus 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 approach 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, whereas 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[6] of HBSs). This is because, in this case, signals received by HBSs in the smaller *hot spot* tend to be more correlated; thus, it is more advantageous to select the HBSs judiciously to increase the sum rate.

[6]In the simulation, it was observed that the average number N_H^* of activated HBSs is about 6 for the simulated configurations.

VII. Conclusion

In this paper, we have studied distributed compression schemes for the uplink of cloud radio access networks. We proposed a robust compression scheme for a practical scenario with inaccurate statistical information about the correlation among the BSs' signals. The scheme is based on a deterministic worst case problem formulation and the solution of the corresponding KKT conditions. Via numerical results, we have demonstrated that, while errors in the statistical model of the SI make a distributed source coding strategy virtually useless, the proposed robust compression scheme allows to tolerate sizable errors without drastic performance degradation and while still reaping the benefits of distributed source coding. In this regard, we remark that the robust strategy could be further improved in at least two ways, which are subject of current work. First, one could deploy a layered compression strategy that attempts to opportunistically leverage a more advantageous SI (see, e.g., [37]). Second, one could enhance the decoding operation by performing joint decompression and decoding, as discussed in [9] and [38], for related models. Moreover, we have addressed the issue of selecting a subset of BSs with the aim of improving the energy efficiency of the network. This was done by proposing a joint BS selection and compression approach, in which a sparsity-inducing term is introduced into the objective function. It was verified that the proposed joint BS selection and compression method shows performance close to exhaustive search.

Appendix A
Proof of Theorem 1

Since the problem (16) involves infinitely many inequality constraints, we first convert it into a problem with a finite number of inequalities, following the standard robust optimization approach reviewed in [39]. This step will also allow us to eliminate variable $\tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$ from the problem formulation. We then show that the KKT conditions are necessary for optimality. Finally, we formulate the KKT conditions and verify that a solution to problem (19) also satisfies the KKT conditions.

Lemma 1: Problem (16) is equivalent to the following problem:

$$\begin{aligned} \underset{\mathbf{\Omega}_i \succeq \mathbf{0}}{\text{maximize}}\ & f_i(\mathbf{\Omega}_i, \lambda_{\text{LB}}\mathbf{I}) - \log\det(\mathbf{\Omega}_i + \mathbf{I}) \\ \text{s.t.}\ & f_i(\mathbf{\Omega}_i, \lambda_{\text{UB}}\mathbf{I}) - C_i \le 0. \end{aligned} \tag{28}$$

Proof: First, we consider the epigraph form of problem (16) (see, e.g., [28, Sec. 4.1.3]) as follows:

$$\begin{aligned} \underset{\mathbf{\Omega_i} \succeq \mathbf{0},\, t,\, \tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} \in \mathcal{H}^{n_M}}{\text{maximize}}\quad & t - \log\det(\mathbf{I} + \mathbf{\Omega}_i) \\ \text{s.t.}\ \max & \left\{ \begin{array}{l} t - f_i(\mathbf{\Omega}_i, \tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}) \\ f_i(\mathbf{\Omega}_i, \tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}) - C_i \end{array} \right\} \le 0 \\ & \lambda_{\min}(\tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}) \ge \lambda_{\text{LB}}, \lambda_{\max}(\tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}) \le \lambda_{\text{UB}}. \end{aligned} \tag{29}$$

Then, we observe that problem (29) is equivalent to the following problem with one inequality constraint:

$$\begin{aligned} \underset{\mathbf{\Omega}_i \succeq \mathbf{0},\, t}{\text{maximize}}\ & t - \log\det(\mathbf{I} + \mathbf{\Omega}_i) \\ \text{s.t.}\ & \max_{\tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} \text{ s.t. } (15)} \max \left\{ \begin{array}{l} t - f_i(\mathbf{\Omega}_i, \tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}), \\ f_i(\mathbf{\Omega}_i, \tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}) - C_i \end{array} \right\} \le 0. \end{aligned} \tag{30}$$

Since the ordering of the maximum operators can be interchanged [28, Sec. 4.1.3], the inequality constraint in (30) can be written as

$$\max \left\{ \begin{array}{l} t - \min\limits_{\tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} \text{ s.t. } (15)} f_i(\mathbf{\Omega}_i, \tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}) \\ \max\limits_{\tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} \text{ s.t. } (15)} f_i(\mathbf{\Omega}_i, \tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}) - C_i \end{array} \right\} \le 0. \tag{31}$$

To proceed, we need to maximize and minimize function f_i with respect to $\tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$ for given $\mathbf{\Omega}_i$ under constraint (15). To this end, note that function f_i can be written as the sum of $\log\det(\mathbf{I} + \mathbf{K}_i\tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}) = \sum_{l=1}^{n_{B,i}} \log(1 + \lambda_l(\mathbf{K}_i\tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}))$ and a term that is independent of $\tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}$, where $\mathbf{K}_i = (\mathbf{I} + \mathbf{\Omega}_i(\mathbf{H}_i\hat{\mathbf{\Sigma}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}}\mathbf{H}_i^\dagger + \mathbf{I}))^{-1}\mathbf{\Omega}_i$, and $\lambda_l(\mathbf{X})$ represents the lth largest eigenvalue of $\mathbf{X}$. Finally, using the following eigenvalue inequalities [40]:

$$\begin{aligned} \lambda_l(\mathbf{K}_i)\lambda_{\min}(\tilde{\mathbf{\Delta}}_{\mathbf{x}|\mathbf{y}_\mathcal{S}}) &\le \lambda_l(\mathbf{K}_i\tilde{\mathbf{\Delta}}_{\mathbf{x}|\mathbf{y}_\mathcal{S}}) \\ &\le \lambda_l(\mathbf{K}_i)\lambda_{\max}(\tilde{\mathbf{\Delta}}_{\mathbf{x}|\mathbf{y}_\mathcal{S}}) \end{aligned} \tag{32}$$

for $l = 1, \ldots, n_{B,i}$, the optimal values for the maximization and minimization of f_i are obtained by setting $\tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} = \lambda_{\text{UB}}\mathbf{I}$ and $\tilde{\mathbf{\Delta}}_{\mathbf{x}|\hat{\mathbf{y}}_\mathcal{S}} = \lambda_{\text{LB}}\mathbf{I}$, respectively. This leads to problem (28). ■

Problem (28) is not convex due to the nonconvexity of constraint set. In the next two lemmas, we list some necessary conditions for the optimality of problem (28).

Lemma 2: The KKT conditions are necessary conditions for optimality in problem (28).

Proof: It follows from [29] and direction calculation. ■

Lemma 3: At any optimal point $\mathbf{\Omega}_i^*$ for problem (28), the backhaul capacity should be fully utilized, i.e., $f_i(\mathbf{\Omega}_i^*, \lambda_{\text{UB}}\mathbf{I}) = C_i$.

Proof: Suppose that $\mathbf{\Omega}_i^0$ is optimal but does not fully utilize the backhaul capacity. Then, we can set $\mathbf{\Omega}_i = \eta\mathbf{\Omega}_i^0$ with some $\eta > 1$ to increase the objective function in (28) without violating the backhaul capacity constraint of problem (28). Thus, $\mathbf{\Omega}_i^0$ cannot be the optimal solution. ■

Now, without loss of optimality, we can consider only the points satisfying the necessary conditions described in Lemmas 2 and 3. To this end, with the choice (11), using standard steps [28, Sec. 5.5.3] and [29], the KKT conditions can be written as

$$\frac{c_l^L}{1+\alpha_l c_l^L} - \frac{1}{1+\alpha_l} - \frac{\mu c_l^U}{1+\alpha_l c_l^U} + \theta_l = 0,\ l = 1, \ldots, n_{B,i} \tag{33a}$$

$$\theta_l\alpha_l = 0,\ \theta_l \ge 0, \quad l = 1, \ldots, n_{B,i} \tag{33b}$$

$$g_i^U(\alpha_1, \ldots, \alpha_{n_{B,i}}) - C_i = 0 \tag{33c}$$

with Lagrange multipliers $\theta_l \geq 0$ for $l = 1, \ldots, n_{B,i}$ and $\mu \geq 0$. We note that, similar to Lemmas 2 and 3, the conditions (33a)–(33c) can be shown to be necessary for the optimality of the following problem:

$$\underset{\alpha_1 \geq 0, \ldots, \alpha_{n_{B,i}} \geq 0}{\text{maximize}} \; g_i^L(\alpha_1, \ldots, \alpha_{n_{B,i}}) - \sum_{l=1}^{n_{B,i}} \log\det(1 + \alpha_l)$$
$$\text{s.t. } g_i^U(\alpha_1, \ldots, \alpha_{n_{B,i}}) - C_i = 0. \tag{34}$$

However, according to the Weierstrass theorem [29], problem (34) has a solution due to the compact constraint set. Thus, we can find parameters $\alpha_i, \ldots, \alpha_{n_{B,i}}$ satisfying the KKT conditions (33a)–(33c) with a proper choice of μ.

The earlier discussion shows that any solution of problem (34) provides a solution to the KKT conditions (33a)–(33c). Moreover, we show that α_l must lie in the set $\mathcal{P}_l(\mu)$ with $\mu \in (0, 1)$ to satisfy the conditions (33a)–(33c); thus, we can limit the domain of the optimization (34), as done in (19)–(20c). This is shown in the following. First, from the following lemma, the search region for μ can be restricted to the interval $\mu \in (0, 1)$.

Lemma 4: For $\mu = 0$ and $\mu \geq 1$, the conditions (33a)–(33c) cannot be satisfied simultaneously.

Proof: With $\mu = 0$, it is impossible to satisfy (33a) and (33b) simultaneously. For $\mu \geq 1$, (33a) does not hold with nonnegative α_l. ■

Lemma 5: A value of α_l with $\alpha_l \notin \mathcal{P}_l$ cannot satisfy the conditions (33a)–(33c).

Proof: In order for (33a) and (33b) to hold together, parameter α_l should be such that

$$\alpha_l^2 + Q_l\alpha_l + S_l = 0, \text{ if } \alpha_l > 0 \tag{35}$$
$$\alpha_l^2 + Q_l\alpha_l + S_l \geq 0, \text{ if } \alpha_l = 0. \tag{36}$$

By direct calculation, it follows that candidate $\alpha_l \in \mathcal{P}_l(\mu)$ must hold to satisfy both (35) and (36). ■

Appendix B
Proof of Theorem 2

The following lemma provides a problem formulation equivalent to (24).

Lemma 6: Problem (24) is equivalent to

$$\underset{\boldsymbol{\Omega}_i \succeq \mathbf{0}}{\text{maximize}} \; \log\det\left(\mathbf{I} + \boldsymbol{\Omega}_i \boldsymbol{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}}\right)$$
$$- \log\det(\mathbf{I} + \boldsymbol{\Omega}_i) - q_H \operatorname{tr}(\boldsymbol{\Omega}_i)$$
$$\text{s.t. } \log\det\left(\mathbf{I} + \boldsymbol{\Omega}_i \boldsymbol{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}}\right) \leq \bar{C}_i \tag{37}$$

where $\boldsymbol{\Sigma}_{\tilde{\mathbf{y}}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}}$ and $\bar{C}_i$ are defined in Theorem 2.

Proof: Follows by using the chain rule for mutual information, see [21] for details. ■

Since problem (37) is nonconvex, we first solve the KKT conditions, which can be proved to be necessary for optimality as in Lemma 2, and then show that the derived solution also satisfies the general sufficiency condition in [29, Prop. 3.3.4]. If we set $\boldsymbol{\Omega}_i$ as (11) with the eigenvalue decomposition $\boldsymbol{\Sigma}_{\tilde{\mathbf{y}}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}} = \mathbf{U}\text{diag}(\lambda_1, \ldots, \lambda_{n_{B,i}})\mathbf{U}^\dagger$, the KKT conditions can be written as

$$\frac{(1-\mu)\lambda_l}{1+\alpha_l\lambda_l}\frac{1}{1+\alpha_l} - q'_H + \theta_l = 0, \; l = 1, \ldots, n_{B,i} \tag{38a}$$
$$\theta_l\alpha_l = 0, \quad l = 1, \ldots, n_{B,i} \tag{38b}$$
$$\mu\left(\sum_{l=1}^{n_{B,i}} \log(1 + \alpha_l\lambda_l) - \bar{C}_i\right) = 0 \tag{38c}$$
$$\sum_{l=1}^{n_{B,i}} \log(1 + \alpha_l\lambda_l) - \bar{C}_i \leq 0 \tag{38d}$$

with Lagrange multipliers $\theta_l \geq 0$ for $l = 1, \ldots, n_{B,i}$ and $\mu \geq 0$. By direct calculation, we can see that the eigenvalues $\alpha_1, \ldots, \alpha_{n_{B,i}}$ in (26) satisfy the conditions (38a)–(38d).

Now, we show that the solution (26) also satisfies the general sufficiency condition in [29, Prop. 3.3.4].

Lemma 7: Let $\boldsymbol{\Omega}_i^*, \mu^*$ denote a pair obtained from Theorem 2. Then, $\boldsymbol{\Omega}_i^*$ is the global optimum of problem (37) since it satisfies the sufficiency condition in [29, Prop. 3.3.4], i.e., the following equality:

$$\boldsymbol{\Omega}_i^* = \arg\max_{\boldsymbol{\Omega}_i \succeq \mathbf{0}} \mathcal{L}(\boldsymbol{\Omega}_i, \mu^*) \tag{39}$$

where we have the Lagrangian as follows:

$$\mathcal{L}(\boldsymbol{\Omega}_i, \mu) = (1-\mu)\log\det\left(\mathbf{I} + \boldsymbol{\Omega}_i \boldsymbol{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}}\right)$$
$$- \log\det(\boldsymbol{\Omega}_i + \mathbf{I}) - q_H \operatorname{tr}(\boldsymbol{\Omega}_i) \tag{40}$$

and the complementary slackness condition $\mu(\log\det(\mathbf{I} + \boldsymbol{\Omega}_i \boldsymbol{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}}) - \bar{C}_i) = 0$.

Proof: In problem (39), selecting the eigenvectors of $\boldsymbol{\Omega}_i$ as those of $\boldsymbol{\Sigma}_{\mathbf{y}_i|\hat{\mathbf{y}}_{\mathcal{N}_\mathcal{B}\setminus\{i\}}}$ does not involve any loss of optimality due to the eigenvalue inequality $\log\det(\mathbf{I} + \mathbf{AB}) \leq \log\det(\mathbf{I} + \boldsymbol{\Gamma}_\mathbf{A}\boldsymbol{\Gamma}_\mathbf{B})$, where $\boldsymbol{\Gamma}_\mathbf{A}$ and $\boldsymbol{\Gamma}_\mathbf{B}$ 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 [17, App. B]). Then, (39) is equivalent to showing that the eigenvalues $\alpha_1^*, \ldots, \alpha_{n_{B,i}}^*$ in (26) satisfy

$$\alpha_l^* = \arg\max_{\alpha_l \geq 0} \left((1-\mu^*)\log(1 + \alpha_l\lambda_l) - \log(1 + \alpha_l) - q_H\alpha_l\right) \tag{41}$$

for $l = 1, \ldots, n_{B,i}$. The given condition can be shown by following the same steps in [17, App. B]. ■

References

[1] S. Liu, J. Wu, C. H. Koh, and V. K. N. Lau, "A 25 Gb/s($/\text{km}^2$) urban wireless network beyond IMT-advanced," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 122–129, Feb. 2011.

[2] Intel Cor., Solution brief Intel heterogeneous network solution brief, Solution brief.

[3] J. Segel and M. Weldon, Technology White Paper 1, Alcatel-Lucent, Lightradio portfolio-technical overview.

[4] "C-RAN: The road towards green RAN," China Mobile Res. Inst., Beijing, China, Oct. 2011, White Paper, ver. 2.5.

[5] T. Flanagan, “Creating cloud base stations with TI’s keystone multicore architecture,” Texas Inst., Dallas, TX, Oct. 2011, White Paper.

[6] “Heterogeneous networks,” Ericsson, Stockholm, Sweden, Feb. 2012, White Paper.

[7] 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 Veh. Technol. Mag.*, vol. 7, no. 1, pp. 16–23, Mar. 2012.

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

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

[10] A. Sanderovich, S. Shamai, and Y. Steinberg, “Distributed MIMO receiver—Achievable rates and upper bounds,” *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4419–4438, Oct. 2009.

[11] 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.

[12] B. Nazer, A. Sanderovich, M. Gastpar, and S. Shamai, “Structured superposition for backhaul constrained cellular uplink,” in *Proc. IEEE ISIT*, Seoul, Korea, Jun. 2009, pp. 1530–1534.

[13] S.-N. Hong and G. Caire, “Quantized compute and forward: A low-complexity architecture for distributed antenna systems,” in *Proc. IEEE ITW*, Paraty, Brazil, Oct. 2011, pp. 420–424.

[14] L. Zhou and W. Yu, “Uplink multicell processing with limited backhaul via successive interference cancellation,” presented at the IEEE Globe Communications Conf. (Globecom 2012), Anaheim, CA, Dec. 2012.

[15] J. Chen and T. Berger, “Successive Wyner-Ziv coding scheme and its application to the quadratic Gaussian CEO problem,” *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1586–1603, Apr. 2008.

[16] M. Gastpar, P. L. Dragotti, and M. Vetterli, “The distributed Karhunen-Loeve transform,” *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5177–5196, Dec. 2006.

[17] A. del Coso and S. Simoens, “Distributed compression for MIMO coordinated networks with a backhaul constraint,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4698–4709, Sep. 2009.

[18] C. Tian and J. Chen, “Remote vector Gaussian source coding with decoder side information,” *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4676–4680, Oct. 2009.

[19] G. Chechik, A. Globerson, N. Tishby, and Y. Weiss, “Information bottleneck for Gaussian variables,” *J. Mach. Learn. Res.*, vol. 6, pp. 165–188, Dec. 2005.

[20] C. Yu and G. Sharma, “Distributed estimation and coding: A sequential framework based on a side-informed decomposition,” *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 759–773, Feb. 2011.

[21] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, “Robust and efficient distributed compression for cloud radio access networks,” arXiv:1206.3602.

[22] N. Tishby, F. C. Pereira, and W. Bialek, “The information bottleneck method,” in *Proc. 37th Allerton Conf.*, Sep. 1999, pp. 368–377.

[23] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[24] Y. C. Eldar, “Robust competitive estimation with signal and noise covariance uncertainties,” *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4532–4547, Oct. 2006.

[25] R. Mittelman and E. L. Miller, “Robust estimation of a random parameter in a Gaussian linear model with joint eigenvalue and elementwise covariance uncertainties,” *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1001–1011, Mar. 2010.

[26] S. Loyka and C. D. Charalambous, “On the compound capacity of a class of MIMO channels subject to normed uncertainty,” *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2048–2063, Apr. 2012.

[27] A. Wiesel, Y. C. Eldar, and S. Shamai, “Optimization of the MIMO compound capacity,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 1094–1101, Mar. 2007.

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[29] D. Bertsekas, *Nonlinear Programming*. New York: Athena Scientific, 1995.

[30] S. Ramanath, V. Kavitha, and E. Altman, “Open loop optimal control of base station activation for green networks,” in *Proc. IEEE WiOpt*, Princeton, NJ, May 2011, pp. 161–166.

[31] 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.

[32] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 2006.

[33] U. Erez and R. Zamir, “Achieving $(1/2)\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding,” *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[34] X. Zhang, J. Chen, S. B. Wicker, and T. Berger, “Successive coding in multiuser information theory,” *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2246–2254, Jun. 2007.

[35] I. Maric, B. Bostjancic, and A. Goldsmith, “Resource allocation for constrained backhaul in Picocell networks,” in *Proc. IEEE ITA*, Feb. 2011, pp. 1–6.

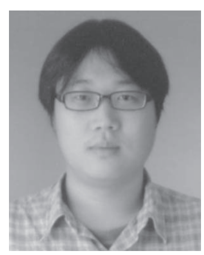
[36] R. G. Baraniuk, “Compressive sensing,” *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[37] C. T. K. Ng, C. Tian, A. J. Goldsmith, and S. Shamai, “Minimum expected distortion in Gaussian source coding with fading side information,” *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5725–5739, Sep. 2012.

[38] O. Simeone, O. Somekh, E. Erkip, H. V. Poor, and S. Shamai, “Robust communication via decentralized processing with unreliable backhaul links,” *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4187–4201, Jul. 2011.

[39] D. Bertsimas, D. B. Brown, and C. Caramanis, “Robust communication via decentralized processing with unreliable backhaul links,” *SIAM Rev.*, vol. 53, no. 3, pp. 464–501, Aug. 2011.

[40] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.



Seok-Hwan Park received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Korea University, Seoul, Korea, in 2005, 2007, and 2011, respectively.

From February 2011 to January 2012, he was a Research Engineer with the Agency for Defense Development, Daejeon, Korea. Since January 2012, he has been a Postdoctoral Research Associate with the Center for Wireless Communication and Signal Processing Research, New Jersey Institute of Technology, Newark. His research interests include signal processing techniques for multicell multiple-input–multiple-output (MIMO) systems, including MIMO beamforming, distributed compression, broadcast coding, delay-tolerant design, hybrid automatic repeat request, and noisy network coding.

Dr. Park received the Best Paper Award at the 2006 Asia-Pacific Conference on Communications and an Excellent Paper Award at IEEE Student Paper Contest in 2006.

Osvaldo Simeone received the M.Sc. degree (with honors) and the Ph.D. degree in information engineering from Polytechnic University of Milan, Milan, Italy, in 2001 and 2005, respectively.

He is currently with the Center for Wireless Communications and Signal Processing Research, New Jersey Institute of Technology, Newark, where he is an Associate Professor. His current research interests include the cross-layer analysis and design of wireless networks with emphasis on information-theoretic, signal processing, and queuing aspects; cognitive radio; cooperative communications; rate–distortion theory; ad hoc, sensor, mesh, and hybrid networks; distributed estimation; and synchronization.

Dr. Simeone coreceived the Best Paper Awards of the 2007 IEEE Signal Processing Advances in Wireless communications Workshop and the 2007 IEEE Wireless Rural and Emergency Communications Conference. He currently serves as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS.

Onur Sahin received B.S. degree in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, in 2003 and the M.Sc. and Ph.D. degrees in electrical engineering from the Polytechnic Institute of New York University, Brooklyn, in 2005 and 2009, respectively.

Since November 2009, he has been with the Advance Air Interface Group, InterDigital Inc., Melville, NY, as a Staff Engineer. He conducts research on the cross-layer design of the next-generation cellular and Wi-Fi systems such as ultradense networks, Long-Term Evolution Advanced, and beyond. He is the author of more than 20 publications and eight international patent applications on the next-generation wireless communication techniques and system design. His recent research interest include small cell networks with the utilization of millimeter-wave (60 GHz) link technology to achieve multi-Gbps throughput experience.

Dr. Sahin received the 2011 InterDigital Innovation Award.

Shlomo Shamai (Shitz) (F'94) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, Haifa, Israel, in 1975, 1981, and 1986 respectively.

From 1975 to 1985, he was with the Communications Research Laboratory as a Senior Research Engineer. Since 1986, he has been with the Department of Electrical Engineering, Technion, where he is currently a Technion Distinguished Professor and holds the William Fondiller Chair of Telecommunications. His research interests include information theory and statistical communications.

Dr. Shamai is a member of the Israeli Academy of Sciences and Humanities. He has also served twice on the Board of Governors of the IEEE Information Theory Society. He has served as Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY and as a member of the Executive Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY. He received the 1985 Alon Grant for Distinguished Young Scientists, the 1999 van der Pol Gold Medal of the Union Radio Scientifique Internationale, the 2000 Technion Henry Taub Prize for Excellence in Research, and the 2011 Claude E. Shannon Award. He coreceived the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003 and 2004 joint IT/COM Societies Paper Award, the 2007 IEEE Information Theory Society Paper Award, the 2009 European Commission Seventh Framework Programme, the Network of Excellence in Wireless COMmunications (NEWCOM++) Best Paper Award, and the 2010 Thomson Reuters Award for International Excellence in Scientific Research.