Multihop Backhaul Compression for the Uplink of 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 investigates efficient backhaul compression strategies for the uplink of cloud radio access networks with a general multihop backhaul topology. In these systems, each radio unit (RU) communicates with the managing control unit (CU) through a set of intermediate RUs. A baseline multiplex-and-forward (MF) scheme is first studied in which each RU forwards the bit streams received from the connected RUs without any processing. It is observed that this strategy may cause significant performance degradation in the presence of a dense deployment of RUs. To obviate this problem, a scheme is proposed in which each RU decompresses the received bit streams and performs linear in-network processing of the decompressed signals. For both the MF and the decompress-process-and-recompress (DPR) backhaul schemes, the optimal design is addressed with the aim of maximizing the sum-rate under the backhaul capacity constraints. Based on the analysis, numerical results are provided to compare the performance of the MF and DPR schemes, highlighting the potential advantage of in-network processing.**

***Index Terms*—Cloud radio access network, multihop backhaul, mesh backhaul, compression, in-network processing.**

I. Introduction

Cloud radio access networks (C-RANs) are composed of a wireless access channel, which connects mobile stations (MSs) to radio units (RUs), and a backhaul network, which links the RUs to a central unit (CU). Focusing on the uplink, the backhaul network carries digitized baseband signals from the RUs to the CU, which performs baseband processing [1]-[3].

Most of the research activity on the subject assumes a single-hop, or *star,* backhaul topology in which each RU is directly connected to its managing CUs via an orthogonal backhaul link (see [4] for review). In this work, instead, we study a more general *multihop* backhaul topology in which each RU may communicate with the managing CU through a set of intermediate RUs as shown in Fig. 1. This backhaul topology is especially relevant for heterogeneous small-cell networks in which RUs of various sizes such as pico/femto or macro base stations are connected by a mesh backhaul network [5] (see also the standard [1]).

Reference [6] provides a simulation-based study of the performance of uplink C-RANs over multihop networks under the assumption that each RU is able to evaluate the log-likelihood ratios of the transmitted bits of the connected MSs. In-network processing of the log-likelihood ratios is proposed to enhance the effectiveness of the use of the backhaul network. In this paper, we instead focus on RUs that directly compress the received baseband signal without performing any demodulation, following the standard set-up for C-RAN [1][2]. Specifically, we investigate the *Multiplex-and-Forward* (MF) scheme, whereby each RU forwards the bit streams received from the connected RUs without any processing (Sec. III). We then propose and investigate the *Decompress-Process-and-Recompress* (DPR) scheme that performs linear in-network processing of the compressed baseband signals (Sec. IV). The DPR scheme is related to the linear in-network processing strategy studied in [7] in the context of *estimation* (and not reliable digital communication).

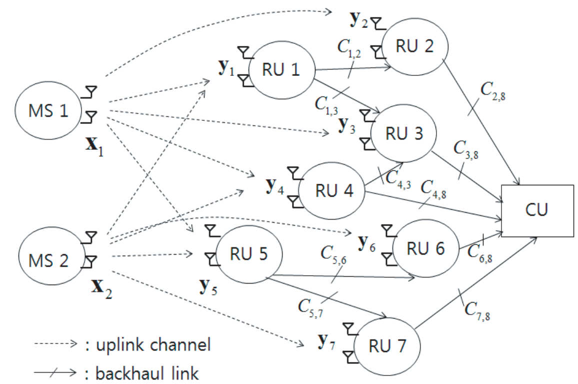

Figure 1. Illustration of the uplink of a C-RAN with a multihop backhaul network (RU: Radio Unit, CU: Control Unit).

II. System Model

We consider the uplink of a C-RAN in which N_M MSs transmit information over a shared wireless medium to N_R RUs as depicted in Fig. 1. The RUs are connected among themselves and to the CU that performs decoding of the MSs' information via a multihop network of backhaul links. We define as $\mathcal{N}_M = \{1, \ldots, N_M\}$ and $\mathcal{N}_R = \{1, \ldots, N_R\}$ the sets of MSs and RUs, respectively. MS k and RU i are equipped with $n_{M,k}$ and $n_{R,i}$ antennas, respectively, for $k \in \mathcal{N}_M$ and $i \in \mathcal{N}_R$. The total number of MSs' antennas is denoted as $n_M = \sum_{k\in\mathcal{N}_M} n_{M,k}$.

1. Channel Model: Here we discuss the wireless uplink channel between MSs and RUs and the multihop backhaul network connecting RUs and the CU.

Uplink: On the uplink channel, the signal $\mathbf{y}_i \in \mathbb{C}^{n_{R,i}\times 1}$ received by RU i at a given time is given by

$$\mathbf{y}_i = \mathbf{H}_i\mathbf{x} + \mathbf{z}_i, \quad (1)$$

where $\mathbf{x} = [\mathbf{x}_1; \ldots; \mathbf{x}_{N_M}]$ is the signal transmitted by all MSs with $\mathbf{x}_k \in \mathbb{C}^{n_{M,k}\times 1}$ denoting the signal transmitted by MS k; $\mathbf{H}_i \in \mathbb{C}^{n_{R,i}\times n_M}$ is the flat-fading channel response matrix from all MSs toward RU i; and $\mathbf{z}_i \in \mathbb{C}^{n_{R,i}\times 1}$ is the additive noise at RU i, which is distributed as $\mathbf{z}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. The signal $\mathbf{x}$ is distributed as $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{x}})$ with covariance matrix $\mathbf{\Sigma}_{\mathbf{x}} = \mathrm{diag}(\mathbf{\Sigma}_{\mathbf{x}_1}, \ldots, \mathbf{\Sigma}_{\mathbf{x}_{N_M}})$. Note that the signals $\mathbf{x}_k$ are independent for $k \in \mathcal{N}_M$, since the MSs are not able to cooperate. As a result, the signal $\mathbf{y} = [\mathbf{y}_1; \ldots; \mathbf{y}_{N_R}]$ received by all RUs is distributed as $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{y}})$ with $\mathbf{\Sigma}_{\mathbf{y}} = \mathbf{H}\mathbf{\Sigma}_{\mathbf{x}}\mathbf{H}^\dagger + \mathbf{I}$ and $\mathbf{H} = [\mathbf{H}_1; \ldots; \mathbf{H}_{N_R}]$.

Backhaul network: In order to model the backhaul multi-hop network connecting the RUs and the CU, we define a capacitated directed acyclic graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ (see, e.g., [8]). Accordingly, the set of vertex nodes of the directed acyclic graph is $\mathcal{V} = \mathcal{N}_R \cup \{N_R+1\}$, where the node i represents the ith RU for $i \in \mathcal{N}_R$ and the last node N_R+1 stands for the CU. Also, the set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ contains the edges, where an edge $e = (i, j)$ represents the backhaul link of capacity C_e bits/s/Hz connecting node i to node j. The capacity C_e is normalized by the bandwidth used on the uplink wireless channel. Note that this enables the capacity C_e to be equivalently measured in bits per channel use of the uplink. The head and tail of edge $e = (i, j)$ with respect to the direction $i \to j$ are denoted by $\mathrm{head}(e) = j$ and $\mathrm{tail}(e) = i$, respectively.

2. Backhaul Routing: As discussed in Sec. I, we will consider different strategies for the transmission of the RUs' baseband received signals to the CU on the backhaul network. For all schemes, routing from the RUs to the CU can be described as detailed in this subsection following similar treatments in [7]. To this end, we fix an ordered partition of the set $\mathcal{V}$, which includes the RUs and the CU, into layers $\mathcal{V}_1, \ldots, \mathcal{V}_L$, so that $\mathcal{V} = \bigcup_{l=1}^{L} \mathcal{V}_l$ and $\mathcal{V}_m \bigcap \mathcal{V}_l = \varnothing$ for $m \neq l$ with $N_R + 1 \in \mathcal{V}_L$. Each partition gives rise to a specific routing schedule, as discussed next.

Given a partition $\mathcal{V}_1, \ldots, \mathcal{V}_L$, we consider as active, and hence available for routing, only the edges, i.e., the backhaul links, that connect nodes belonging to successive layers. More precisely, we define the set $\mathcal{E}_{\mathrm{act}}$ of the active edges as

$$\mathcal{E}_{\mathrm{act}} = \{e \in \mathcal{E} | \mathrm{tail}(e) \in \mathcal{V}_l \,\mathrm{and}\, \mathrm{head}(e) \in \mathcal{V}_k \text{ with } l < k\}.$$

Moreover, we define as $\Gamma_I(i) = \{e_1^i, \ldots, e_{|\Gamma_I(i)|}^i\}$ and $\Gamma_O(i)$ the sets of active edges that end or originate at node i, respectively. In other words, we have $\Gamma_I(i) = \{e \in \mathcal{E}_{\mathrm{act}} | \mathrm{head}(e) = i\}$ and $\Gamma_O(i) = \{e \in \mathcal{E}_{\mathrm{act}} | \mathrm{tail}(e) = i\}$.

A given ordered partition $\mathcal{V}_1, \ldots, \mathcal{V}_L$ defines a routing strategy as follows. Each node i in the first layer, i.e., with $i \in \mathcal{V}_1$, transmits on the active backhaul links $e \in \Gamma_O(i)$ to the nodes in the next layers $\mathcal{V}_l$, $l > 1$. The nodes in the second layer $\mathcal{V}_2$ wait until all the nodes in the same layer receive from the connected nodes in $\mathcal{V}_1$ and then transmit on the active backhaul links to the nodes in the next layers $\mathcal{V}_l$ with $l > 2$. In general, the nodes in each layer $\mathcal{V}_l$ wait for all the nodes in the same layer to receive from the previous layers $\mathcal{V}_1, \ldots, \mathcal{V}_{l-1}$ and then transmit on the active backhaul links to the nodes in the next layers $\mathcal{V}_{l+1}, \ldots, \mathcal{V}_L$.

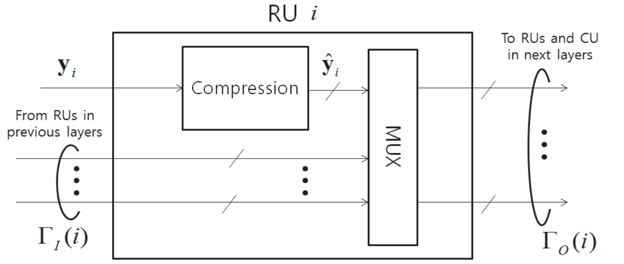

Figure 2. Illustration of the operation at a RU i in the "Multiplex-and-Forward" (MF) scheme studied in Sec. III (Plain arrows "$\rightarrow$" indicate baseband signals, while broken arrows "$\nrightarrow$" denote bit streams).

We now discuss how the choice of the routing strategy and the the tolerated delay for communication from the RUs to the CU affect the use of the capacity of each backhaul link. To start, for a given set $\mathcal{E}_{\mathrm{act}}$ of active edges, we define as D_i the number of edges in the longest path connecting the node i to the CU $N_R + 1$. Then, we define the *depth* D of a routing strategy defined by the partition $\mathcal{V}_1, \ldots, \mathcal{V}_L$ as $D = \max_{i \in \mathcal{N}_R} D_i$. Finally, we define as T the maximum delay allowed for transmission of the received baseband signals from the RUs to the CU. We normalize T by the duration of the transmission on the uplink, so that $T = 1$ means that the delay allowed for transmission on the backhaul network equals the duration of the uplink transmission. Assuming for simplicity that each active backhaul link is used for the same amount of time, we then obtain that each active edge is used only for a period equal to T/D uplink slots. Therefore, the *effective backhaul capacity* $\tilde{C}_e$ used on an edge $e \in \mathcal{E}_{\mathrm{act}}$, i.e., the number of bits per channel use of the uplink that are transmitted on a given active edge e, equals $\tilde{C}_e = C_e \cdot T/D$.

III. Multiplex-and-Forward

In this section, we present a reference scheme, which we refer to as Multiplex-and-Forward (MF). In this scheme, as illustrated in Fig. 2, each RU i performs compression of its received baseband signal $\mathbf{y}_i$ using a given quantization codebook and then simply multiplexes the bit streams received from the previous layers and its compressed signal without any further processing.

We assume a Gaussian test channel (without claim of optimality), so that the compressed signal $\hat{\mathbf{y}}_i$ is given by

$$\hat{\mathbf{y}}_i = \mathbf{y}_i + \mathbf{q}_i, \quad (2)$$

where $\mathbf{q}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_i)$ is the quantization noise. The compressed signal $\hat{\mathbf{y}}_i$ in (2) can be obtained at the output of the compressor if the output rate R_i satisfies the inequality [9, Ch. 3]

$$\begin{aligned} g_i^{\mathrm{MF}}\left(\{\mathbf{\Omega}_i\}_{i \in \mathcal{M}}\right) &\triangleq I\left(\mathbf{y}_i; \hat{\mathbf{y}}_i\right) \quad (3) \\ &= \log_2 \det\left(\mathbf{\Omega}_i + \mathbf{\Sigma}_{\mathbf{y}_i}\right) - \log_2 \det\left(\mathbf{\Omega}_i\right) \leq R_i. \end{aligned}$$

We now define as $f_e^i \geq 0$ the rate (in bits per channel use of the uplink) used to convey the compressed signal of RU i on edge e for $(i,e) \in \mathcal{N}_R \times \mathcal{E}_{\rm act}$. By the definition of the routing scheme, we have the following constraints on the flow variables f_e^i (see, e.g., [8])

$$f_e^i \geq R_i, \text{ for } i \in \mathcal{N}_R \text{ and } e \in \Gamma_O(i), \tag{4}$$

$$\sum_{e\in\Gamma_I(N_R+1)} f_e^i \geq R_i, \text{ for } i \in \mathcal{N}_R, \tag{5}$$

$$\sum_{i\in\mathcal{N}_R} f_e^i \leq \tilde{C}_e, \text{ for } e \in \mathcal{E}_{\rm act}, \tag{6}$$

$$\text{and} \sum_{e\in\Gamma_I(j)} f_e^i \geq \sum_{e\in\Gamma_O(j)} f_e^i, \text{ for } (j,i) \in \mathcal{N}_R \times \mathcal{N}_R, \tag{7}$$

As long as the constraints (3) and (4)-(7) are satisfied, the CU is able to recover the signals $\hat{\mathbf{y}}_i$, $i \in \mathcal{N}_R$, and an achievable sum-rate $R_{\rm sum}$ between the MSs and the CU is given as

$$\begin{aligned} R_{\rm sum} =& I\left(\mathbf{x}; \{\hat{\mathbf{y}}_i\}_{i\in\mathcal{N}_R}\right) = f^{\rm MF}\left(\{\mathbf{\Omega}_i\}_{i\in\mathcal{N}_R}\right) \\ \triangleq& \log_2\det\left(\mathbf{H}\mathbf{\Sigma}_{\mathbf{x}}\mathbf{H}^\dagger + \mathbf{I} + \mathbf{\Omega}\right) - \log_2\det\left(\mathbf{I}+\mathbf{\Omega}\right), \end{aligned} \tag{8}$$

with the definition $\mathbf{\Omega} = \text{diag}(\mathbf{\Omega}_1, \ldots, \mathbf{\Omega}_{N_R})$.

A. Problem Definition and Optimization

The compression strategies $\{\mathbf{\Omega}_i\}_{i\in\mathcal{N}_R}$ and the flow variables $\{f_e^i\}_{i\in\mathcal{N}_R, e\in\mathcal{E}_{\rm act}}$ can now be optimized with the goal of maximizing the sum-rate $R_{\rm sum}$ in (8) subject to the constraints (4)-(7) and (3). This problem is stated as

$$\underset{\substack{\{\mathbf{\Omega}_i \succeq \mathbf{0}, R_i \geq 0\}_{i\in\mathcal{N}_R}, \\ \{f_e^i \geq 0\}_{i\in\mathcal{N}_R, e\in\mathcal{E}_{\rm act}}}}{\text{maximize}} \quad f^{\rm MF}\left(\{\mathbf{\Omega}_i\}_{i\in\mathcal{N}_R}\right) \tag{9a}$$

$$\text{s.t.} \quad g_i^{\rm MF}\left(\{\mathbf{\Omega}_i\}_{i\in\mathcal{N}_R}\right) \leq R_i, \text{ for } i \in \mathcal{N}_R, \tag{9b}$$

$$(4)-(7). \tag{9c}$$

We note that the optimization (9) requires full channel state information (CSI).

The problem (9) with respect to the variables $\{\mathbf{\Omega}_i\}_{i\in\mathcal{N}_R}$ and $\{f_e^i \geq 0\}_{i\in\mathcal{N}_R, e\in\mathcal{E}_{\rm act}}$ is a difference-of-convex problem, since the functions $f^{\rm MF}(\{\mathbf{\Omega}_i\}_{i\in\mathcal{N}_R})$ and $g_i^{\rm MF}(\{\mathbf{\Omega}_i\}_{i\in\mathcal{N}_R})$ can be written as the difference of convex functions and all other constraints in (9c) are linear [10]. For difference-of-convex problems, the Majorization and Minimization (MM) algorithm provides an iterative procedure that is known to converge to a stationary point of the problem [10]. The detailed algorithm is described in [11, Algorithm 1].

IV. Decompress-Process-and-Recompress

The MF backhaul strategy studied in the previous section may incur a significant performance degradation when the RUs have a sufficiently large number of incoming edges. In fact, in this case, the bit rate obtained by multiplexing the signals received from the RUs in the previous layers is large and the backhaul capacity constraints may impose a critical performance bottleneck. In this section, we introduce a scheme that attempts to solve this problem via decompression at each

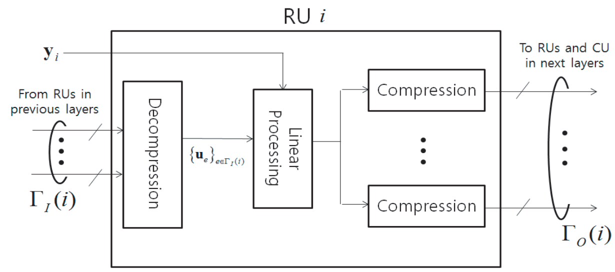

Figure 3. Illustration of the operation at a RU i in the "Decompress-Process-and-Recompress" (DPR) scheme studied in Sec. IV (Plain arrows "→" indicate baseband signals, while broken arrows "↛" denote bit streams).

RU and linear in-network processing of the decompressed signals and of the locally received signal. The key idea is that the processing step can reduce redundancy by properly combining the available (compressed) received signals. On the flip side, the processed signals need to be recompressed before they can be sent on the backhaul links. As discussed in [12] in the context of a cascade source coding problem, this recompression step introduces further distortion. The effect of this distortion must thus be counterbalanced by the advantages of in-network processing in order to make the strategy preferable to MF.

We now detail the DPR scheme and analyze its performance. As shown in Fig. 3, each RU i first decompresses the signals $\mathbf{u}_{e'}$ received on its incoming edges $e' \in \Gamma_I(i)$. Then, for each outgoing edge $e \in \Gamma_O(i)$, it processes the vector $\mathbf{r}_i$ that includes the decompressed signals $\mathbf{u}_{e'}$ for all edges $e' \in \Gamma_I(i)$ and the received baseband signal $\mathbf{y}_i$, namely

$$\mathbf{r}_i = [\mathbf{y}_i; \mathbf{u}_{e_1^i}; \cdots; \mathbf{u}_{e^i_{|\Gamma_I(i)|}}], \tag{10}$$

via a linear processing matrix $\mathbf{L}_e$. This produces a processed signal $\mathbf{L}_e\mathbf{r}_i$ for all outgoing edges $e \in \Gamma_O(i)$. We assume here that matrix $\mathbf{L}_e$ is square; the issue of dimensionality reduction via the use of "wide" matrices $\mathbf{L}_e$ is further discussed in Sec. V. Note that the matrix $\mathbf{L}_e$ can be written as

$$\mathbf{L}_e = [\mathbf{L}_e^{\rm rx}\, \mathbf{L}_e^{e_1^i}\, \cdots\, \mathbf{L}_e^{e^i_{|\Gamma_I(i)|}}], \tag{11}$$

where, by (10), the matrices $\mathbf{L}_e^{\rm rx} \in \mathbb{C}^{d_e \times n_{R,i}}$ and $\mathbf{L}_e^{e_j^i} \in \mathbb{C}^{d_e \times d_{e_j^i}}$ multiply the signals $\mathbf{y}_i$ and $\mathbf{u}_{e_j^i}$, respectively, for $j \in \{1, \ldots, |\Gamma_I(i)|\}$. Finally, RU i compresses the processed signal $\mathbf{L}_e\mathbf{r}_i$ at a rate of $\tilde{C}_e$ bits per channel use to produce the output signal $\mathbf{u}_e$ to be sent on the outgoing edge $e \in \Gamma_O(i)$.

As in the previous section, we adopt a Gaussian test channel, so that the signal $\mathbf{u}_e$ is given by

$$\mathbf{u}_e = \mathbf{L}_e\mathbf{r}_i + \mathbf{q}_e, \tag{12}$$

with quantization noise $\mathbf{q}_e \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_e)$, which is independent across the edge index e. Similar to (3), the signal $\mathbf{u}_e$ can be reliably transmitted to RU head(e) if the condition

$$\begin{aligned} g_e^{\rm DPR}\left(\{\mathbf{L}_e, \mathbf{\Omega}_e\}_{e\in\mathcal{E}_{\rm act}}\right) &\triangleq I\left(\mathbf{r}_i; \mathbf{u}_e\right) \\ &= \log_2\det\left(\mathbf{\Omega}_e + \mathbf{L}_e\mathbf{\Sigma}_{\mathbf{r}_i}\mathbf{L}_e^\dagger\right) - \log_2\det\left(\mathbf{\Omega}_e\right) \leq \tilde{C}_e \end{aligned} \tag{13}$$

is satisfied.

The CU performs joint decoding of the messages of all MSs based on the received signal $\mathbf{r}_{N_R+1}$, which can be written, similar to (10), as $\mathbf{r}_{N_R+1} = [\mathbf{u}_{e_1^{N_R+1}}; \cdots ; \mathbf{u}_{e_{|\Gamma_I(N_R+1)|}^{N_R+1}}]$. As a result, the sum-rate

$$R_{\text{sum}} = I\left(\mathbf{x}; \mathbf{r}_{N_R+1}\right) \tag{14}$$

is achievable between the MSs and the CU. The sum-rate (14) is characterized in the following lemma.

Lemma 1. *For any given routing strategy defined by the partition $\mathcal{V}_1, \ldots, \mathcal{V}_L$ with the active edges $\mathcal{E}_{\text{act}} = \{e_1, \ldots, e_{|\mathcal{E}_{\text{act}}|}\}$, the sum-rate R_{sum} in (14) is given by*

$$\begin{aligned} R_{\text{sum}} =& f^{\text{DPR}}\left(\{\mathbf{L}_e, \mathbf{\Omega}_e\}_{e\in\mathcal{E}_{\text{act}}}\right) \\ \triangleq& \log_2 \det\left(\mathbf{T}\mathbf{H}\mathbf{\Sigma}_{\mathbf{x}}\mathbf{H}^\dagger\mathbf{T}^\dagger + \mathbf{T}\mathbf{T}^\dagger + \tilde{\mathbf{T}}\mathbf{\Omega}\tilde{\mathbf{T}}^\dagger\right) \\ &- \log_2 \det\left(\mathbf{T}\mathbf{T}^\dagger + \tilde{\mathbf{T}}\mathbf{\Omega}\tilde{\mathbf{T}}^\dagger\right), \end{aligned} \tag{15}$$

where $\mathbf{\Omega} = \text{diag}(\mathbf{\Omega}_{e_1}, \ldots, \mathbf{\Omega}_{e_{|\mathcal{E}_{\text{act}}|}})$ and the matrices $\mathbf{T}$ and $\tilde{\mathbf{T}}$ are defined as

$$\mathbf{T} = \mathbf{C}\left(\mathbf{I}-\mathbf{F}\right)^{-1}\mathbf{E} \text{ and } \tilde{\mathbf{T}} = \mathbf{C}\left(\mathbf{I}-\mathbf{F}\right)^{-1}. \tag{16}$$

In (16), the matrix $\mathbf{C}$ is a $d_{N_R} \times \sum_{e\in\mathcal{E}_{\text{act}}} d_e$ block-matrix with the (i,j)-th block given by $\mathbf{C}_{e_i^{N_R+1}, e_j} \in \mathbb{C}^{d_{e_i^{N_R+1}} \times d_{e_j}}$ for $i \in \{1, \ldots, |\Gamma_I(N_R+1)|\}$ and $j \in \{1, \ldots, |\mathcal{E}_{\text{act}}|\}$; the matrix $\mathbf{F}$ is a $\sum_{e\in\mathcal{E}_{\text{act}}} d_e \times \sum_{e\in\mathcal{E}_{\text{act}}} d_e$ block-matrix with the (i,j)-th block given by $\mathbf{F}_{e_i,e_j} \in \mathbb{C}^{d_{e_i}\times d_{e_j}}$ for $i \in \{1, \ldots, |\mathcal{E}_{\text{act}}|\}$ and $j \in \{1, \ldots, |\mathcal{E}_{\text{act}}|\}$; the matrix $\mathbf{E}$ is a $\sum_{e\in\mathcal{E}_{\text{act}}} d_e \times \sum_{i\in\mathcal{N}_R} d_i$ block-matrix with the (i,j)-th block given by $\mathbf{E}_{e_i,j} \in \mathbb{C}^{d_{e_i}\times d_j}$ for $i \in \{1, \ldots, |\mathcal{E}_{\text{act}}|\}$ and $j \in \{1, \ldots, N_R\}$, where we have defined the matrices

$$\mathbf{C}_{e,e'} = \begin{cases} \mathbf{I}, & \text{if } e = e' \\ \mathbf{0}, & \text{otherwise} \end{cases}, \tag{17}$$

$$\mathbf{F}_{e,e'} = \begin{cases} \mathbf{L}_e^{e'}, & \text{if tail}(e) = \text{head}(e') \\ \mathbf{0}, & \text{otherwise} \end{cases}, \tag{18}$$

$$\text{and } \mathbf{E}_{e,j} = \begin{cases} \mathbf{L}_e^{\text{rx}}, & \text{if tail}(e) = j \\ \mathbf{0}, & \text{otherwise} \end{cases}. \tag{19}$$

Proof: The result follows by noting that the signal $\mathbf{r}_{N_R+1}$ received by the CU can be written as

$$\mathbf{r}_{N_R+1} = \mathbf{T}\mathbf{y} + \tilde{\mathbf{T}}\mathbf{q}, \tag{20}$$

with the quantization noise vector $\mathbf{q} = [\mathbf{q}_{e_1}; \cdots ; \mathbf{q}_{e_{|\mathcal{E}_{\text{act}}|}}] \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega})$. This can be proved by identifying the state-space equations and linear transfer functions as done in [7, Sec. III-A]. □

A. Problem Definition and Optimization

The problem of optimizing the variables $\{\mathbf{L}_e, \mathbf{\Omega}_e\}_{e\in\mathcal{E}_{\text{act}}}$ can be stated as

$$\begin{aligned} \underset{\{\mathbf{L}_e, \mathbf{\Omega}_e \succeq \mathbf{0}\}_{e\in\mathcal{E}_{\text{act}}}}{\text{maximize}} \quad & f^{\text{DPR}}\left(\{\mathbf{L}_e, \mathbf{\Omega}_e\}_{e\in\mathcal{E}_{\text{act}}}\right) \\ \text{s.t.} \quad & g_e^{\text{DPR}}\left(\{\mathbf{L}_e, \mathbf{\Omega}_e\}_{e\in\mathcal{E}_{\text{act}}}\right) \leq \tilde{C}_e, \ \text{for } e \in \mathcal{E}_{\text{act}}. \end{aligned} \tag{21}$$

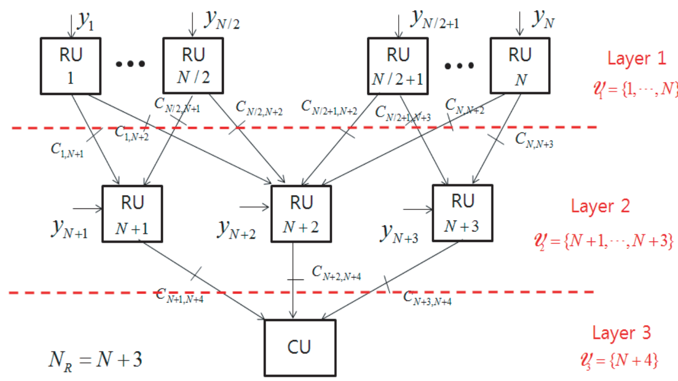

Figure 4. The backhaul network assumed for the simulations in Sec. V. All RUs have the same received SNR and are equipped with a single antenna.

The following proposition shows that, under the stated assumptions, we can fix the linear transformations $\mathbf{L}_e$ to be equal to an identity matrix, i.e., $\mathbf{L}_e = \mathbf{I}$ for all $e \in \mathcal{E}_{\text{act}}$, without loss of optimality.

Proposition 1. *For any solution $\{\mathbf{L}'_e, \mathbf{\Omega}'_e\}_{e\in\mathcal{E}_{\text{act}}}$ of problem (21), there exists another equivalent solution $\{\mathbf{L}''_e, \mathbf{\Omega}''_e\}_{e\in\mathcal{E}_{\text{act}}}$ with $\mathbf{L}''_e = \mathbf{I}$, in the sense that $f^{\text{DPR}}(\{\mathbf{L}'_e, \mathbf{\Omega}'_e\}_{e\in\mathcal{E}_{\text{act}}}) = f^{\text{DPR}}(\{\mathbf{L}''_e, \mathbf{\Omega}''_e\}_{e\in\mathcal{E}_{\text{act}}})$ and $g_e^{\text{DPR}}(\{\mathbf{L}'_e, \mathbf{\Omega}'_e\}_{e\in\mathcal{E}_{\text{act}}}) = g_e^{\text{DPR}}(\{\mathbf{L}''_e, \mathbf{\Omega}''_e\}_{e\in\mathcal{E}_{\text{act}}})$ for all $e \in \mathcal{E}_{\text{act}}$.*

Proof: See [11, Appendix I]. □

Using Proposition 1, the problem (21) can be reduced with no loss of optimality to an optimization solely with respect to the quantization noise covariances $\{\mathbf{\Omega}_e\}_{e\in\mathcal{E}_{\text{act}}}$. The mentioned optimization of (21) with $\mathbf{L}_e = \mathbf{I}$, $e \in \mathcal{E}_{\text{act}}$, can be seen to be a difference-of-convex problem, as introduced in Sec. III. Therefore, we can apply the MM approach [10] to find a stationary point of the problem as in Sec. III. The complete algorithm is described in [11, Algorithm 2].

V. Numerical Results

In this section, we demonstrate the performance of the backhaul communication schemes studied in the paper. Unless stated otherwise, we consider the backhaul network shown in Fig. 4 with a routing strategy described by the partition $\mathcal{V}_1 = \{1, \ldots, N\}$, $\mathcal{V}_2 = \{N+1, N+2, N+3\}$ and $\mathcal{V}_3 = \{N+4\}$ that leads to all edges being activated, i.e., $\mathcal{E} = \mathcal{E}_{\text{act}}$. We assume that all edges have the same backhaul capacity unless stated otherwise and set $T = D$ so that the effective capacity satisfies the equality $\tilde{C}_e = C_e$. It is also assumed that the elements of the channel matrix $\mathbf{H}_i$ are independent and identically distributed (i.i.d.) $\mathcal{CN}(0,1)$ variables for $i \in \mathcal{N}_R$ (Rayleigh fading). MSs and RUs are equipped with a single antenna and the signals $\mathbf{x}$ transmitted by MSs are distributed as $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, P_{\text{tx}}\mathbf{I})$, so that the transmitted power by each MS is given by P_{tx}.

Fig. 5 shows the average sum-rate versus the number N of RUs in layer 1 with $N_M = 4$ MSs, $P_{\text{tx}} = 0$ dB and backhaul capacity $C_e = 3$ bits/s/Hz except for RU $N+2$

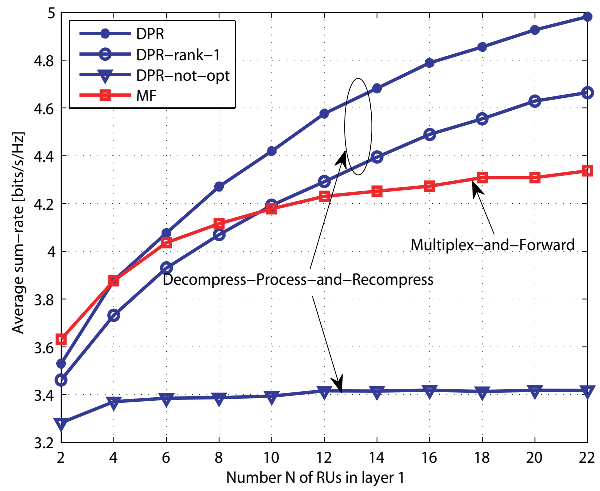

Figure 5. Average sum-rate versus the number N RUs in layer 1 with $N_M = 4$ MSs, $P_{\rm tx} = 0$ dB, $C = 3$ bits/s/Hz and RU $2N+2$ deactivated.

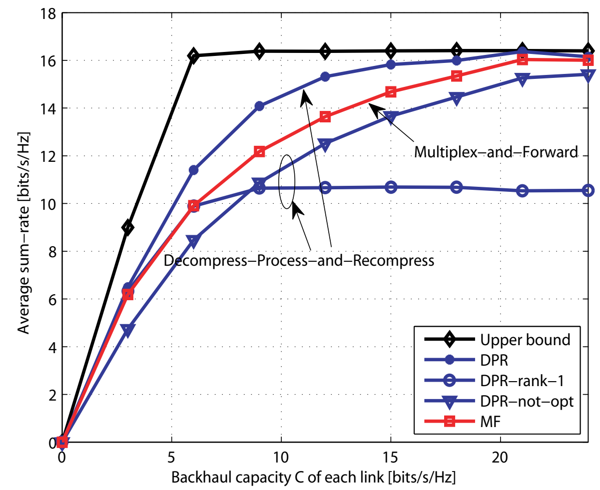

Figure 6. Average sum-rate versus the backhaul capacity C of each link with $N_M = 5$ MSs, $N = 8$ RUs in layer 1 and $P_{\rm tx} = 0$ dB.

which is assumed to be deactivated, i.e., $C_{N+2,N+4} = 0$. We compare the DPR scheme studied in Sec. IV with the MF scheme analyzed in Sec. III. We also consider for reference two suboptimal DPR schemes: in the first, the rank of matrices $\mathbf{L}_e$ in (11) is constrained to be one (labeled as "DPR-rank-1", see details in [11, Sec. IV-B]), while, in the second, the compression covariances are constrained to be equal to scaled identities, i.e., $\{\mathbf{\Omega}_e = c_e\mathbf{I}\}_{e\in\mathcal{E}_{\rm act}}$ (labeled as "DPR-not-opt"). It is first observed that the performance gain of the DPR scheme over MF becomes more pronounced as the number N of RUs in the first layer increases. This implies that, as the density of the RUs' deployment increases, it is desirable for each RU in layer 2 to perform in-network processing of the signals received from layer 1 in order to use the backhaul links to the CU more efficiently.

In Fig. 6, we plot the average sum-rate versus the backhaul capacity C of all edges with $N_M = 5$ MSs, $N = 8$ RUs in the first layer and $P_{\rm tx} = 0$ dB. For reference, we also plot an upper bound $R_{\rm UB}$ on the sum-rate achievable with Gaussian quantization noises, which is obtained using cut-set arguments [9]. We first observe from Fig. 6 that DPR outperforms MF in the regime of intermediate backhaul capacities C. It is also seen that both DPR and MF achieve the upper bound if the backhaul capacity C is large enough. Finally, we observe that, when the backhaul capacity is sufficiently large, limiting the rank of the baseband signals sent on the backhaul links (DPR-rank-1) leads to a significant performance loss.

VI. Concluding Remarks

In this work, we have studied efficient compression and routing strategies for the backhaul of uplink C-RAN systems with a multihop backhaul topology. We have first presented a baseline scheme in which each RU forwards the bit streams received from the connected RUs without any processing. Since this strategy may suffer from a significant performance degradation in the presence of a dense deployment of RUs, we have introduced a scheme in which each RU decompresses the received bit streams and performs linear in-network processing of the decompressed signals. Additional analysis concerning the robust system design in the presence of imperfect channel state information, multiple control units and the role of side information for decompression can be found in [11].

References

[1] Ericsson AB, Huawei Technologies, NEC Corporation, Alcatel Lucent and Nokia Siemens Networks, "Common public radio interface (CPRI); interface specification," CPRI specification v5.0, Sep. 2011.

[2] China Mobile, "C-RAN: the road towards green RAN," White Paper, ver. 2.5, China Mobile Research Institute, Oct. 2011.

[3] T. Biermann, L. Scalia, C. Choi, W. Kellerer and H. Karl, "How backhaul networks influence the feasibility of coordinated multipoint in cellular networks," *IEEE Comm. Mag.*, vol. 51, no. 8, pp. 168-176, Aug. 2013.

[4] O. Simeone, N. Levy, A. Sanderovich, O. Somekh, B. M. Zaidel, H. V. Poor and S. Shamai (Shitz), "Cooperative wireless cellular systems: an information-theoretic view," *Foundations and Trends in Communications and Information Theory*, vol. 8, nos. 1-2, pp. 1-177, 2012.

[5] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," arxiv:1306.6659.

[6] W. Ni, R. P. Liu, I. B. Collings and X. Wang, "Indoor cooperative small cells over Ethernet," *IEEE Comm. Mag.*, vol. 51, no. 9, pp. 100-107, Sep. 2013.

[7] N. Goela and M. Gastpar, "Reduced-dimension linear transform coding of correlated signasl in networks," *IEEE Trans. Sig. Processing*, vol. 60, no. 6, pp. 3174-3187, Jun. 2012.

[8] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Networking*, vol. 11, no. 5, pp. 782-795, Oct. 2003.

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

[10] A. Beck and M. Teboulle, "Gradient-based algorithms with applications to signal recovery problems," in *Convex Optimization in Signal Processing and Communications*, Y. Eldar and D. Palomar, eds., pp. 42-88, Cambridge University Press. 2010.

[11] S.-H. Park, O. Simeone, O. Sahin and S. Shamai (Shitz), "Multihop backhaul compression for the uplink of cloud radio access networks," available at http://arXiv.org.

[12] P. Cuff, H.-I. Su and A. E. Gamal, "Cascade multiterminal source coding," in *Proc. IEEE Int'l Symp. on Inf. Theory (ISIT 2009)*, Seoul, Korea, pp. 1199-1203, Jun. 2009.