Performance Evaluation of Multiterminal Backhaul Compression for Cloud Radio Access Networks

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

***Abstract*—In cloud radio access networks (C-RANs), the baseband processing of the available macro- or pico/femto-base stations (BSs) is migrated to control units, each of which manages a subset of BS antennas. The centralized information processing at the control units enables effective interference management. The main roadblock to the implementation of C-RANs hinges on the effective integration of the radio units, i.e., the BSs, with the backhaul network. This work first reviews in a unified way recent results on the application of advanced multiterminal, as opposed to standard point-to-point, backhaul compression techniques. The gains provided by multiterminal backhaul compression are then confirmed via extensive simulations based on standard cellular models. As an example, it is observed that multiterminal compression strategies provide performance gains of more than 60% for both the uplink and the downlink in terms of the cell-edge throughput.**

***Index Terms*—Cloud radio access network, multiterminal backhaul compression, proportional fairness.**

I. Introduction

A promising architecture for next-generation wireless cellular systems prescribes the separation of localized and distributed radio units from remote and centralized information processing, or control, nodes. This architecture is often referred to as a *cloud radio access network* (C-RAN) [1][2]. The centralization of information processing afforded by C-RANs potentially enables effective interference management at the geographical scale covered by the distributed radio units. The main roadblock to the realization of this potential hinges on the effective integration of the wireless interface provided by the radio units with the backhaul network [3]. Current solutions, which are the object of various standardization efforts [4], prescribe the use of standard analog-to-digital conversion (ADC) techniques in the uplink and standard digital-to-analog conversion (DAC) techniques in the downlink. With these standard solutions, backhaul capacity limitations are known to impose a formidable bottleneck to the system performance (see, e.g., [5]).

In order to alleviate the performance bottleneck identified above, recent efforts by a number of wireless companies have targeted the design of more advanced *backhaul compression* schemes. These are based on various ad hoc combinations of ADC and DAC techniques and proprietary *point-to-point* compression algorithms (see, e.g., [1]). However, as it is well known from network information theory, point-to-point techniques generally fail to achieve the optimal performance in even the simplest multiterminal settings [6]. Recent works have hence explored the performance of *multiterminal*, as opposed to standard point-to-point, backhaul compression techniques for the uplink [7]-[10] and the downlink [11] of C-RAN systems. In this paper, we first review these works in Sec. III for the uplink and in Sec. IV for the downlink in a unified fashion. We then provide extensive simulation results based on standard cellular models [12] to lend evidence to the gains provided by multiterminal backhaul compression as compared to standard point-to-point techniques in Sec. V.

Notation: For random variables X, Y and Z, we adopt standard information-theoretic definitions for the mutual information $I(X;Y)$, conditional mutual information $I(X;Y|Z)$, differential entropy $h(X)$ and conditional differential entropy $h(X|Y)$ [6]. Given a sequence $X_1,\ldots,X_m$, we define a set $X_{\mathcal{S}}=\{X_j|j\in\mathcal{S}\}$ for a subset $\mathcal{S}\subseteq\{1,\ldots,m\}$. For random vectors $\mathbf{x}$ and $\mathbf{y}$, we define the following correlation matrices $\mathbf{\Sigma_x}=\mathbb{E}[\mathbf{x}\mathbf{x}^\dagger]$, $\mathbf{\Sigma_{x,y}}=\mathbb{E}[\mathbf{x}\mathbf{y}^\dagger]$ and $\mathbf{\Sigma_{x|y}}=\mathbb{E}[\mathbf{x}\mathbf{x}^\dagger|\mathbf{y}]$.

II. System Model

We consider the two-dimensional hexagonal cellular layout with 19 macro cells shown in Fig. 1. We assume that each macro-base station (BS) uses three sectorized antennas, and each pico-BS and mobile station (MS) uses a single omni-directional antenna. In each macro-cell, K MSs and N pico-BSs are uniformly distributed. Fig. 1 illustrates an example with $K=2$ MSs and $N=1$ pico-BS.

In a C-RAN system, the baseband processing of the available macro- or pico/femto-BSs is migrated to control units, each of which manages a subset of BS antennas. For example, in Fig. 1, a control unit manages the three sectors of cell 1 and hence the corresponding sectorial antennas of the three relevant macro-BSs and the available pico-BS. We refer to a subset of BS antennas connected to the same control unit, and to the corresponding covered area, as a *cluster*.

Every ith BS is connected to the corresponding control unit via a backhaul link with capacity C_i bps/Hz [3] where the normalization is done with respect to the bandwidth of the wireless uplink/downlink channels. For instance, if BS i communicates with the corresponding control unit at a date rate of 100 Mbps and the wireless uplink/downlink channels

978-1-4799-3001-2/14/$31.00 ©2014 IEEE

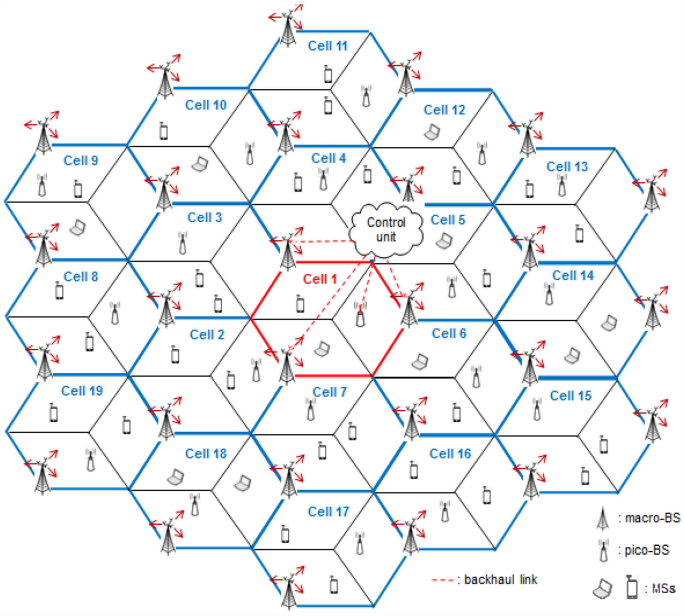


Figure 1. Two-dimensional hexagonal cellular layout with 19 macro hexagonal cells. Each macro BS has three sectorized antennas, while pico-BSs and MSs use omni-directional antennas. We are interested in the performance at macro cell 1 located at the center of the figure.

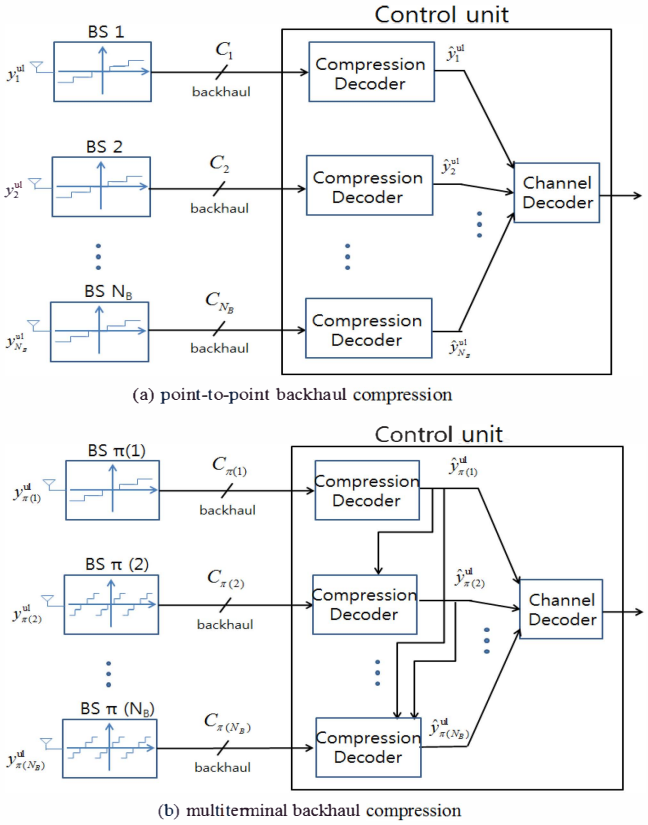


Figure 2. Backhaul compression and decompression for the uplink of C-RAN: (a) point-to-point compression; (b) multiterminal compression.

have a 10 MHz bandwidth, the normalized backhaul capacity is given as $C_i = 10$ bps/Hz.

According to the C-RAN principle, the data exchanged on the backhaul links between BSs and control units consists of compressed baseband signals [1]-[5]. Specifically, in the uplink, the baseband signal received by each BS is compressed and forwarded to the connected control unit, where decoding takes place. Instead, in the downlink, the baseband signals are produced and compressed by the control units, and then upconverted and transmitted by the BSs.

In the following, we detail the signal and channel model by focusing on one specific cluster, e.g., cell 1 in Fig. 1. For notational convenience, we index the BSs in the cluster as $1, 2, \ldots, N_B$ and the MSs in the cluster as $1, 2, \ldots, N_M$, and define the sets $\mathcal{N}_\mathcal{B} = \{1, \ldots, N_B\}$ and $\mathcal{N}_\mathcal{M} = \{1, \ldots, N_M\}$.

A. Uplink Channel

The signal y_i^{ul} received by BS i in the cluster under study in the uplink is given by

$$y_i^{\text{ul}} = \mathbf{h}_i^{\text{ul}\dagger}\mathbf{x}^{\text{ul}} + z_i^{\text{ul}}, \tag{1}$$

where $\mathbf{x}^{\text{ul}} = [x_1^{\text{ul}} \cdots x_{N_M}^{\text{ul}}]^T$ is the $n_M \times 1$ vector of symbols transmitted by all the N_M MSs in the cluster, with x_k^{ul} being the symbol transmitted by MS k; the noise $z_i^{\text{ul}} \sim \mathcal{CN}(0, \sigma^2_{z_i^{\text{ul}}})$ models thermal noise and the interference signals arising from the other clusters; and the channel vector $\mathbf{h}_i^{\text{ul}} \in \mathbb{C}^{N_M \times 1}$ from all the N_M MSs in the cluster toward BS i is given by $\mathbf{h}_i^{\text{ul}} = [h_{i,1}^{\text{ul}}\ h_{i,2}^{\text{ul}}\ \cdots\ h_{i,N_M}^{\text{ul}}]^T$ with $h_{i,k}^{\text{ul}}$ denoting the uplink channel response from the kth MS and to the ith BS. The signal x_k^{ul} is subject to the per-MS power constraint, which is stated as $\mathbb{E}[|x_k^{\text{ul}}|^2] \leq P_{M,k}$ for $k \in \mathcal{N}_\mathcal{M}$.

B. Downlink Channel

In the downlink, each MS k in the cluster under study receives a signal given as

$$y_k^{\text{dl}} = \mathbf{h}_k^{\text{dl}\dagger}\mathbf{x}^{\text{dl}} + z_k^{\text{dl}}, \tag{2}$$

where we have defined the aggregate transmit signal vector by all the N_B BSs in the cluster as $\mathbf{x}^{\text{dl}} = [x_1^{\text{dl}}, \ldots, x_{N_B}^{\text{dl}}]^T$ with x_i^{dl} denoting the signal transmitted by the ith BS; the additive noise $z_k^{\text{dl}} \sim \mathcal{CN}(0, \sigma^2_{z_k^{\text{dl}}})$ accounts for thermal noise and interference from the other clusters; and the channel vector $\mathbf{h}_k^{\text{dl}} \in \mathbb{C}^{N_B \times 1}$ from all the BSs in the cluster toward MS k is given as $\mathbf{h}_k^{\text{dl}} = [h_{k,1}^{\text{dl}}\ h_{k,2}^{\text{dl}}\ \cdots\ h_{k,N_B}^{\text{dl}}]^T$ with $h_{k,i}^{\text{dl}}$ denoting the downlink channel gain from BS i to MS k. Finally, we have the per-BS power constraints $\mathbb{E}[|x_i^{\text{dl}}|^2] \leq P_{B,i}$, for $i \in \mathcal{N}_\mathcal{B}$.

For both uplink and downlink, the channel vectors $\{\mathbf{h}_i^{\text{ul}}\}_{i\in\mathcal{N}_\mathcal{B}}$ and $\{\mathbf{h}_k^{\text{dl}}\}_{k\in\mathcal{N}_\mathcal{M}}$ remain constant for the entire coding block duration and are known to the corresponding control unit. As discussed in Sec. I, the main goal of this paper is to provide a realistic evaluation of the advantages of the multiterminal backhaul compression strategies proposed in [9] for the uplink and in [11] for the donwlink. In the next two sections, we review these strategies.

III. Multiterminal Compression for the Uplink of C-RAN

In the *uplink* of C-RAN, each MS k within the cluster under study encodes its message M_k to produce a transmitted signal $x_k^{\rm ul}$ for each channel use. This signal is taken from a conventional Gaussian codebook and is hence distributed as $x_k^{\rm ul} \sim \mathcal{CN}(0, P_k)$ where P_k satisfies the per-MS power constraint $P_k \leq P_{M,k}$. Note that, since the MSs cannot cooperate with each other, the transmitted signals $x_k^{\rm ul}$ are independent across the MS index k.

Each ith BS communicates with the control unit by providing the latter with a compressed version $\hat{y}_i^{\rm ul}$ of the received signal $y_i^{\rm ul}$. The control unit first decompresses the signals $\hat{y}_i^{\rm ul}$, $i \in \mathcal{N}_{\mathcal{B}}$, and then, based on all signals $\hat{y}_{\mathcal{N}_{\mathcal{B}}}^{\rm ul}$, decodes the MSs' messages. Using standard rate-distortion considerations, we express the compressed signal $\hat{y}_i^{\rm ul}$ as[1]

$$\hat{y}_i^{\rm ul} = y_i^{\rm ul} + q_i^{\rm ul}, \tag{3}$$

where the quantization noise $q_i^{\rm ul}$ is independent of the signal $y_i^{\rm ul}$ and distributed as $q_i^{\rm ul} \sim \mathcal{CN}(0, \omega_i^{\rm ul})$.

Point-to-Point Backhaul Compression [13]: In a conventional system, the control unit decompresses the descriptions $\hat{y}_{\mathcal{N}_{\mathcal{B}}}^{\rm ul}$ in parallel as shown in Fig. 2-(a). In this case, the signal $\hat{y}_i^{\rm ul}$ can be recovered at the control unit if the condition

$$I(y_i^{\rm ul}; \hat{y}_i^{\rm ul}) = \log_2\left(\omega_i^{\rm ul} + \sigma^2_{y_i^{\rm ul}}\right) - \log_2\left(\omega_i^{\rm ul}\right) \leq C_i \tag{4}$$

is satisfied where $\sigma^2_{y_i^{\rm ul}} = \mathbf{h}_i^{\rm ul\dagger}\mathbf{\Sigma}_{\mathbf{x}^{\rm ul}}\mathbf{h}_i^{\rm ul} + \sigma^2_{z_i^{\rm ul}}$ with $\mathbf{\Sigma}_{\mathbf{x}^{\rm ul}} = {\rm diag}(\{P_k\}_{k\in\mathcal{N}_{\mathcal{M}}})$ (see, e.g., [6, Ch. 3]).

Multiterminal Backhaul Compression [7]-[10]: Standard point-to-point compression does not leverage the statistical correlation among the signals $y_i^{\rm ul}$ received at different BSs. Based on this observation, distributed compression was proposed in [7] to utilize such correlation. Following [9][10], this can be done as follows. For a given ordering π of the BS indices, the control unit decompresses in the order $\hat{y}_{\pi(1)}^{\rm ul}, \hat{y}_{\pi(2)}^{\rm ul}, \ldots, \hat{y}_{\pi(N_{\mathcal{B}})}^{\rm ul}$ as shown in Fig. 2-(b). Therefore, when decompressing $\hat{y}_{\pi(i)}^{\rm ul}$, the control unit has already retrieved the signals $\hat{y}_{\pi(1)}^{\rm ul}, \ldots, \hat{y}_{\pi(i-1)}^{\rm ul}$. These signals can be hence treated as side information available at the decoder, namely the control unit, but not to the encoder, namely BS $\pi(i)$. As a result, using the Wyner-Ziv theorem [6, Ch. 3], the descriptions $\hat{y}_{\pi(i)}^{\rm ul}$ for $i \in \mathcal{N}_{\mathcal{B}}$ can be recovered at the control unit if the conditions

$$\begin{aligned} &I(y_{\pi(i)}^{\rm ul}; \hat{y}_{\pi(i)}^{\rm ul} | \hat{y}_{\{\pi(1),\ldots,\pi(i-1)\}}^{\rm ul}) = g_{\pi,i}^{\rm ul}(\mathbf{p}, \boldsymbol{\omega}) \\ &\triangleq \log_2\left(\omega_{\pi(i)}^{\rm ul} + \sigma^2_{y_{\pi(i)}^{\rm ul}|\hat{y}_{\{\pi(1),\ldots,\pi(i-1)\}}^{\rm ul}}\right) - \log_2\left(\omega_{\pi(i)}^{\rm ul}\right) \leq C_{\pi(i)} \end{aligned} \tag{5}$$

are satisfied, where we have defined vectors $\mathbf{p} = [P_1, \ldots, P_{N_M}]$ and $\boldsymbol{\omega} = [\omega_1, \ldots, \omega_{N_{\mathcal{B}}}]$, and the conditional variance $\sigma^2_{y_{\pi(i)}^{\rm ul}|\hat{y}_{\{\pi(1),\ldots,\pi(i-1)\}}^{\rm ul}}$ is given by

$$\sigma^2_{y_{\pi(i)}^{\rm ul}|\hat{y}_{\{\pi(1),\ldots,\pi(i-1)\}}^{\rm ul}} = \mathbf{h}_{\pi(i)}^{\rm ul\dagger}\mathbf{\Sigma}_{\mathbf{x}^{\rm ul}|\hat{y}_{\{\pi(1),\ldots,\pi(i-1)\}}^{\rm ul}}\mathbf{h}_{\pi(i)}^{\rm ul} + \sigma^2_{z_{\pi(i)}^{\rm ul}}, \tag{6}$$

with $\mathbf{\Sigma}_{\mathbf{x}^{\rm ul}|\hat{y}_{\{\pi(1),\ldots,\pi(i-1)\}}^{\rm ul}} = \mathbf{\Sigma}_{\mathbf{x}^{\rm ul}} - \mathbf{\Sigma}_{\mathbf{x}^{\rm ul},\hat{y}_{\{\pi(1),\ldots,\pi(i-1)\}}^{\rm ul}}\mathbf{\Sigma}^{-1}_{\hat{y}_{\{\pi(1),\ldots,\pi(i-1)\}}^{\rm ul}}\mathbf{\Sigma}^{\dagger}_{\mathbf{x}^{\rm ul},\hat{y}_{\{\pi(1),\ldots,\pi(i-1)\}}^{ul}}$. The matrices $\mathbf{\Sigma}_{\mathbf{x}^{\rm ul},\hat{y}_{\{\pi(1),\ldots,\pi(i-1)\}}^{\rm ul}}$ and $\mathbf{\Sigma}_{\hat{y}_{\{\pi(1),\ldots,\pi(i-1)\}}^{\rm ul}}$ are given by

$$\mathbf{\Sigma}_{\mathbf{x}^{\rm ul},\hat{y}_{\{\pi(1),\ldots,\pi(i-1)\}}^{\rm ul}} = \mathbf{\Sigma}_{\mathbf{x}^{\rm ul}}\mathbf{H}_{\pi,i-1}^{\rm ul\dagger}, \tag{7}$$

$$\begin{aligned} \text{and } \mathbf{\Sigma}_{\hat{y}_{\{\pi(1),\ldots,\pi(i-1)\}}^{\rm ul}} =& \mathbf{H}_{\pi,i-1}^{\rm ul}\mathbf{\Sigma}_{\mathbf{x}^{\rm ul}}\mathbf{H}_{\pi,i-1}^{\rm ul\dagger} \\ &+ {\rm diag}\left(\{\sigma^2_{z_{\pi(j)}^{\rm ul}} + \omega^2_{\pi(j)}\}_{j=1}^{i-1}\right), \end{aligned} \tag{8}$$

where we have defined $\mathbf{H}_{\pi,i-1}^{\rm ul} = [\mathbf{h}_{\pi(1)}^{\rm ul}, \ldots, \mathbf{h}_{\pi(i-1)}^{\rm ul}]^{\dagger}$.

We assume that the control unit performs single-user decoding of the messages $\{M_k\}_{k\in\mathcal{N}_{\mathcal{M}}}$ sent by MSs based on all the descriptions $\{\hat{y}_i^{\rm ul}\}_{i\in\mathcal{N}_{\mathcal{B}}}$, so that each message M_k is decoded by treating the interference signals $x_j^{\rm ul}$ for $j \neq k$ as noise (see [9] for the analysis with joint decoding of all MSs and [10] for successive interference cancellation). Under this assumption, the achievable rate R_k for MS k is given by

$$\begin{aligned} R_k^{\rm ul} =& I(x_k^{\rm ul}; \hat{y}_{\mathcal{N}_{\mathcal{B}}}^{\rm ul}) = f_k^{\rm ul}(\mathbf{p}, \boldsymbol{\omega}) \\ \triangleq& \log_2 \det\left({\rm diag}\left(\{\omega_i^{\rm ul}\}_{i\in\mathcal{N}_{\mathcal{B}}}\right) + \mathbf{\Sigma}_{\mathbf{y}_{\mathcal{N}_{\mathcal{B}}}^{\rm ul}}\right) \\ &- \log_2 \det\left({\rm diag}\left(\{\omega_i^{\rm ul}\}_{i\in\mathcal{N}_{\mathcal{B}}}\right) + \mathbf{\Sigma}_{\mathbf{y}_{\mathcal{N}_{\mathcal{B}}}^{\rm ul}|x_k^{\rm ul}}\right), \end{aligned} \tag{9}$$

where the conditional covariance $\mathbf{\Sigma}_{\mathbf{y}_{\mathcal{N}_{\mathcal{B}}}^{\rm ul}|x_{\mathcal{S}}^{\rm ul}}$ with $\mathcal{S} \subseteq \mathcal{N}_{\mathcal{M}}$ is given as

$$\mathbf{\Sigma}_{\mathbf{y}_{\mathcal{N}_{\mathcal{B}}}^{\rm ul}|x_{\mathcal{S}}^{\rm ul}} = \sum_{j\in\mathcal{N}_{\mathcal{M}}\setminus\mathcal{S}} P_j\tilde{\mathbf{h}}_j^{\rm ul}\tilde{\mathbf{h}}_j^{\rm ul\dagger} + {\rm diag}\left(\{\sigma^2_{z_i^{\rm ul}}\}_{i\in\mathcal{N}_{\mathcal{B}}}\right) \tag{10}$$

with $\tilde{\mathbf{h}}_k^{\rm ul} = [h_{1,k}^{\rm ul}, h_{2,k}^{\rm ul}, \ldots, h_{N_{\mathcal{B}},k}^{\rm ul}]^T$.

We are interested in evaluating the performance of the standard proportional-fair scheduler. This scheduler, at each time slot, select the power allocation $\mathbf{p}$ and the quantization noise powers $\boldsymbol{\omega}$ and the order π so as to maximize the weighted sum-rate

$$u^{\rm ul}(\mathbf{p}, \boldsymbol{\omega}) = \sum_{k\in\mathcal{N}_{\mathcal{M}}} f_k^{\rm ul}(\mathbf{p}, \boldsymbol{\omega})/\bar{R}_k^{\alpha}, \tag{11}$$

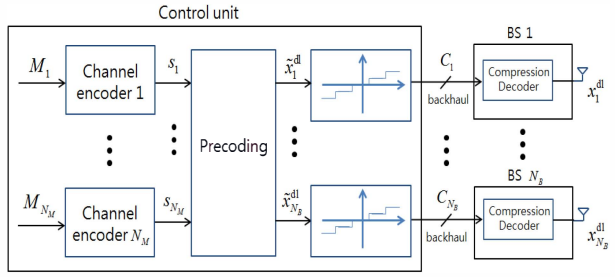
with $\alpha \geq 0$ being a fairness constant and $\bar{R}_k$ represents the average data rate of MS k until the previous time slot (see, e.g., [14]). After each time slot, the rate $\bar{R}_k$ is updated as $\bar{R}_k \leftarrow \beta\bar{R}_k + (1-\beta)R_k^{\rm ul}$ where $\beta \in [0,1]$ is a forgetting factor. We recall that increasing the constant α encourages fairness among the MSs, while the objective function reduces to the sum-rate when $\alpha = 0$. This problem is formulated as

$$\underset{\pi, \mathbf{p}\in\mathbb{R}_+^{N_M}, \boldsymbol{\omega}\in\mathbb{R}_+^{N_{\mathcal{B}}}}{\text{maximize}} \; u^{\rm ul}(\mathbf{p}, \boldsymbol{\omega}) \tag{12a}$$

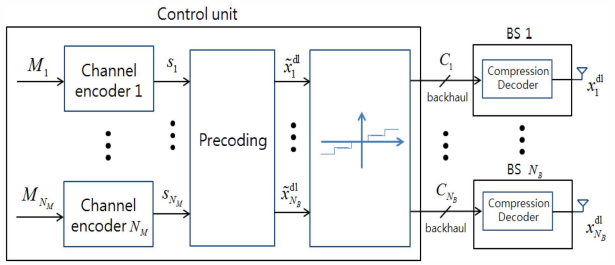
$$\text{s.t.} \quad g_{\pi,i}^{\rm ul}(\mathbf{p}, \boldsymbol{\omega}) \leq C_{\pi(i)}, \text{ for all } i \in \mathcal{N}_{\mathcal{B}}, \tag{12b}$$

$$P_k \leq P_{M,k}, \text{ for all } k \in \mathcal{N}_{\mathcal{M}}. \tag{12c}$$

[1] It is recalled that rate-distortion theory applies to vector quantizers of large dimension although the mathematical characterizations of the operation (such as (3)) and of the performance (such as (4) below) are given in terms of individual samples.



(a) point-to-point backhaul compression



(b) multiterminal backhaul compression

Figure 3. Backhaul compression and decompression for the downlink of C-RAN: (a) point-to-point compression; (b) multiterminal compression.

To tackle the non-convex problem (12), we propose a separate design of the power control variables $\mathbf{p}$ and the compression noise powers $\boldsymbol{\omega}$ for a fixed permutation π. Specifically, at Step 1, the power coefficients $\mathbf{p}$ are optimized assuming ideal backhaul links (i.e., $\omega_i^{\rm ul}=0$ for $i\in\mathcal{N}_{\mathcal{B}}$). This problem is equivalently stated as

$$\underset{R_k,\mathbf{p}\in\mathbb{R}_+^{N_M}}{\text{maximize}} \quad \sum_{k\in\mathcal{N}_{\mathcal{M}}} R_k/\bar{R}_k^{\alpha} \tag{13a}$$

$$\text{s.t.} \quad R_k \le f_k^{\rm ul}(\mathbf{p},\mathbf{0}),\ \text{for all } k\in\mathcal{N}_{\mathcal{M}}, \tag{13b}$$

$$P_k \le P_{M,k},\ \text{for all } k\in\mathcal{N}_{\mathcal{M}}. \tag{13c}$$

Albeit still non-convex, it is seen that the problem (13) belongs to the class of different-of-convex (DC) problems (see, e.g., [15]). Thus, we can leverage the iterative majorization minimization (MM) algorithm, which is known to converge to a locally optimal point of (13) (see, e.g., [15, Sec. 1.3.3]). The MM algorithm solves a sequence of convex problems obtained by linearizing the non-convex constraints (13b). With the so-obtained power variables $\mathbf{p}$, at Step 2, we optimize the quantization noise powers $\boldsymbol{\omega}$. It can be seen that the optimal quantization power $\omega_{\pi(i)}^{\rm ul}$, for fixed powers $\mathbf{p}$, is simply given by imposing equality in the backhaul constraint (12b), leading to

$$\omega_{\pi(i)}^{\rm ul}=\sigma^2_{y^{\rm ul}_{\pi(i)}|\hat{y}^{\rm ul}_{\{\pi(1),\ldots,\pi(i-1)\}}}/(2^{C_{\pi(i)}}-1) \tag{14}$$

for $i\in\mathcal{N}_{\mathcal{B}}$ with $\sigma^2_{y^{\rm ul}_{\pi(i)}|\hat{y}^{\rm ul}_{\{\pi(1),\ldots,\pi(i-1)\}}}$ given in (6).

IV. Multiterminal Compression for the Downlink of C-RAN

In the *downlink* of a C-RAN, the control unit first encodes each message M_k for MS $k\in\mathcal{N}_{\mathcal{M}}$ via a separate channel encoder, which produces a coded signal s_k for each channel use. Each coded symbol s_k is taken from a conventional Gaussian codebook and hence it is distributed as $s_k\sim\mathcal{CN}(0,1)$. The signals $\mathbf{s}=[s_1,\ldots,s_{N_M}]$ are further processed by the control unit in two stages, namely *precoding* and *compression*.

1. Precoding: In order to allow for interference management both across the MSs and among the data streams for the same MS, the signals in vector $\mathbf{s}$ are linearly precoded via multiplication of a complex matrix $\mathbf{A}\in\mathbb{C}^{N_{\mathcal{B}}\times N_M}$. The precoded data can be written as

$$\tilde{\mathbf{x}}^{\rm dl}=\mathbf{A}\mathbf{s}, \tag{15}$$

where the matrix $\mathbf{A}$ can be factorized as $\mathbf{A}=[\mathbf{a}_1\cdots\mathbf{a}_{N_M}]$ with $\mathbf{a}_k\in\mathbb{C}^{N_{\mathcal{B}}\times 1}$ denoting the precoding vector corresponding to MS k. The precoded data $\tilde{\mathbf{x}}^{\rm dl}$ can be written as $\tilde{\mathbf{x}}^{\rm dl}=[\tilde{x}_1^{\rm dl},\ldots,\tilde{x}_{N_{\mathcal{B}}}^{\rm dl}]^T$, where the signal $\tilde{x}_i^{\rm dl}$ is the precoded signal corresponding to the ith BS and is given as $\tilde{x}_i^{\rm dl}=\mathbf{e}_i^{\dagger}\mathbf{A}\mathbf{s}$ with the vector $\mathbf{e}_i\in\mathbb{C}^{N_{\mathcal{B}}\times 1}$ having all zero elements except for the ith element that contains 1.

2-(a). Point-to-Point Backhaul Compression [16]: Each precoded data stream $\tilde{x}_i^{\rm dl}$ for $i\in\mathcal{N}_{\mathcal{B}}$ must be compressed in order to allow the control unit to deliver it to the ith BS through the backhaul link of capacity C_i bps/Hz. Each ith BS then simply forwards the compressed signal $x_i^{\rm dl}$ obtained from the control unit. Using standard rate-distortion considerations, we adopt a Gaussian test channel to model the effect of compression on the backhaul link. In particular, we write the compressed signals $x_i^{\rm dl}$ to be transmitted by BS i as

$$x_i^{\rm dl}=\tilde{x}_i^{\rm dl}+q_i^{\rm dl}, \tag{16}$$

where the compression noise $q_i^{\rm dl}$ is modeled as a complex Gaussian noise. With conventional backhaul compression, as shown in Fig. 3-(a), the signal $\tilde{x}_i^{\rm dl}$ corresponding to different BSs are compressed separately, which leads to independent quantization noises $q_i^{\rm dl}$. Similar to the uplink, the compressed signal (16) can be transmitted to the ith BS if the condition

$$I\left(\tilde{\mathbf{x}}_i;\mathbf{x}_i\right)=\log_2\left(\mathbf{e}_i^{\dagger}\mathbf{A}\mathbf{A}^{\dagger}\mathbf{e}_i+\omega_{i,i}^{\rm dl}\right)-\log_2\left(\omega_{i,i}^{\rm dl}\right)\le C_i \tag{17}$$

is satisfied.

We now discuss the multiterminal backhaul compression strategies proposed in [11], and illustrated in Fig. 3-(b).

2-(b). Multiterminal Backhaul Compression [11]: The main idea of the multiterminal backhaul compression for the downlink is to control the effect of the additive quantization noises at the MSs by designing their correlation across the BSs within the cluster. This is made possible by *multivariate compression* [6, Ch. 7], which requires joint compression of all signals as in Fig. 3-(b). A successive compression implementation, which is dual to the successive decompression implementation of distributed source coding shown in Fig. 2-(b) for the uplink, is detailed in [11, Sec. IV-D].

To elaborate, we write the vector $\mathbf{x}^{\rm dl}=[x_1^{\rm dl},\ldots,x_{N_{\mathcal{B}}}^{\rm dl}]^T$ of compressed signals for all the BSs as

$$\mathbf{x}^{\rm dl}=\mathbf{A}\mathbf{s}+\mathbf{q}^{\rm dl}. \tag{18}$$

In (18), the compression noise $\mathbf{q}^{\rm dl} = [q_1^{\rm dl}, \ldots, q_{N_{\mathcal{B}}}^{\rm dl}]^T$ is modeled as a complex Gaussian vector $\mathbf{q}^{\rm dl} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}^{\rm dl})$, where the covariance matrix $\mathbf{\Omega}^{\rm dl}$ consists of elements $\omega_{i,j}^{\rm dl} = \mathbb{E}[q_i^{\rm dl} q_j^{\rm dl\dagger}]$ defining the correlation between the quantization noises of BS i and BS j.

Using the multivariate compression lemma in [6, Ch. 9], reference [11] shows that the signals $x_1^{\rm dl}, \ldots, x_{N_{\mathcal{B}}}^{\rm dl}$ obtained via the test channel (18) can be reliably transferred to the BSs on the backhaul links if the condition

$$g_{\mathcal{S}}^{\rm dl}\left(\mathbf{A}, \mathbf{\Omega}^{\rm dl}\right) \triangleq \sum_{i \in \mathcal{S}} h\left(x_i^{\rm dl}\right) - h\left(x_{\mathcal{S}}^{\rm dl} | \tilde{\mathbf{x}}^{\rm dl}\right) \tag{19}$$

$$= \sum_{i \in \mathcal{S}} \log_2(\mathbf{e}_i^\dagger \mathbf{A}\mathbf{A}^\dagger \mathbf{e}_i + \omega_{i,i}^{\rm dl}) - \log_2 \det(\mathbf{E}_{\mathcal{S}}^\dagger \mathbf{\Omega}^{\rm dl} \mathbf{E}_{\mathcal{S}}) \leq \sum_{i \in \mathcal{S}} C_i$$

is satisfied for all subsets $\mathcal{S} \subseteq \mathcal{N}_{\mathcal{B}}$, where the matrix $\mathbf{E}_{\mathcal{S}}$ is obtained by stacking the vectors $\mathbf{e}_i$ for $i \in \mathcal{S}$ horizontally. We observe that the inequalities (17) for standard point-to-point compression are obtained by substituting $\omega_{i,j}^{\rm dl} = 0$ into (19).

With the described precoding and compression operations and assuming that the interference signals are treated as noise at MSs, the achievable rate R_k for MS k is computed as

$$R_k = I\left(s_k; y_k^{\rm dl}\right) = f_k^{\rm dl}\left(\mathbf{A}, \mathbf{\Omega}^{\rm dl}\right) \tag{20}$$

$$\triangleq \log_2\left(\sigma_{z_k^{\rm dl}}^2 + \mathbf{h}_k^{\rm dl\dagger}\left(\mathbf{A}\mathbf{A}^\dagger + \mathbf{\Omega}^{\rm dl}\right)\mathbf{h}_k^{\rm dl}\right)$$

$$- \log_2\left(\sigma_{z_k^{\rm dl}}^2 + \mathbf{h}_k^{\rm dl\dagger}\left(\sum_{l \in \mathcal{N}_{\mathcal{M}} \setminus \{k\}} \mathbf{a}_l \mathbf{a}_l^\dagger + \mathbf{\Omega}^{\rm dl}\right)\mathbf{h}_k^{\rm dl}\right).$$

Similar to the uplink, our goal is to implement the proportional fairness scheduler, which requires to optimize the weighted sum-rate over the precoding matrix $\mathbf{A}$ and the quantization covariance matrix $\mathbf{\Omega}^{\rm dl}$, subject to the backhaul constraints (19) and the per-BS power constraints $P_{B,i}$. This problem can be formulated and solved following the same steps as for the uplink discussed in Sec. III by applying the MM algorithm on its epigraph form.

V. Performance Evaluation

In this section, we discuss the performance advantages of multiterminal backhaul compression for the uplink and downlink of C-RAN systems on a standard cellular model based on [12]. We focus on the performance evaluation in *macro-cell 1* in Fig. 1, which is served by the three sectorized antennas from the corresponding macro-BSs and by N pico-BSs. A control unit is connected to all BS antennas that serve cell 1 as in Fig. 1, which is to be hence considered as a cluster. The backhaul links to each macro-BS antenna and to each pico-BS have the capacities of $C_{\rm macro}$ and $C_{\rm pico}$ bps/Hz, respectively. All interference signals from other macro-cells, denoted by cell 2, cell 3,. . ., cell 19, are treated as independent noise signals. We used the system parameters suggested in [12, Tables 5.3.3-1, 5.3.4-1], and adopted the LTE rate model proposed in [17, Annex A]. We assume that the fairness is measured during T time slots in which the locations of pico-BSs and MSs are fixed and small-scale fading channels change independently from slot to slot.

As shown in [18], with frequency reuse factor $F = 1$, the advantages of intra-cluster cooperation are masked by the effects of the interference coming from the adjacent clusters. Thus, we consider the frequency reuse pattern with $F = 1/3$ proposed in [18] in which the available bandwidth is partitioned into three bands B_1, B_2 and B_3, which are allocated so as to minimize the resulting inter-cluster interference as illustrated in [18, Fig. 6]. As a result, cell 1 of interest suffers from the interference signals only from cells (8,10,12,14,16,18).

A. Uplink

In this subsection, we examine the advantage of the multiterminal compression scheme based on distributed source coding reviewed in Sec. III for the uplink of the C-RAN described above. In Fig. 4, the CDF of the sum-rate is plotted with $K = 5$ MSs, $(C_{\rm macro}, C_{\rm pico}) = (3, 1)$ bps/Hz and $\alpha = 0$. For the order π on the BS, we assume that the control unit first retrieves the signals compressed at the macro-BSs and then decompresses the signals received from the pico-BSs. It is observed that, as compared to standard point-to-point compression, multiterminal compression provides performance gains of 17%, 27% and 42% for $N = 5$, 10 and 20 pico-BSs, respectively, in terms of the 50%-ile sum-rate. Thus, the performance gain of the multiterminal compression is most pronounced when a large number of pico-BSs are located in the same cluster. This suggests that a sophisticated design of backhaul compression provides relevant gain if many radio units are concentrated in given areas.

In Fig. 5, we plot the cell-edge throughput, i.e., the 5%-ile rate, versus the average spectral efficiency. The curve is obtained by varying the fairness constant α in the utility function (11) (see, e.g., [13, Fig. 5]). We fix $N = 3$ pico-BSs, $K = 5$ MSs, $(C_{\rm macro}, C_{\rm pico}) = (9, 3)$ bps/Hz, $T = 10$ and $\beta = 0.5$. As we increase the constant α, the 5%-ile rate increases due to the enhanced fairness among the MSs. We observe that spectral efficiencies larger than 1.01 bps/Hz are not achievable with point-to-point compression, while they can be obtained with multiterminal compression. Moreover, it is seen that multiterminal compression provides 1.6x gain in terms of cell-edge throughput for spectral efficiency of 2.9 bps/Hz.

B. Downlink

In this subsection, we turn to the advantage of the multiterminal compression technique as described in Sec. IV for the downlink. Fig. 6 plots the cell-edge throughput versus the average spectral efficiency for $N = 3$ pico-BSs, $K = 5$ MSs, $(C_{\rm macro}, C_{\rm pico}) = (9, 3)$ bps/Hz, $T = 10$ and $\beta = 0.5$. As for the uplink, it is seen that spectral efficiencies larger than 1.05 bps/Hz are not achievable with point-to-point compression, while they can be obtained with multiterminal compression. Specifically, multiterminal compression provides about 2x gain in terms of cell-edge throughput for spectral efficiency of 1 bps/Hz.

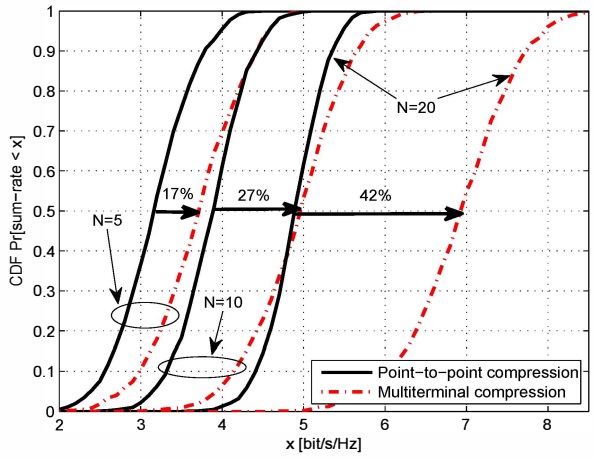


Figure 4. CDF of the sum-rate in the uplink C-RAN with parameters as in [12, Tables 5.3.3-1, 5.3.4-1], $K = 5$ MSs, $(C_{\text{macro}}, C_{\text{pico}}) = (3, 1)$ bps/Hz and $\alpha = 0$.

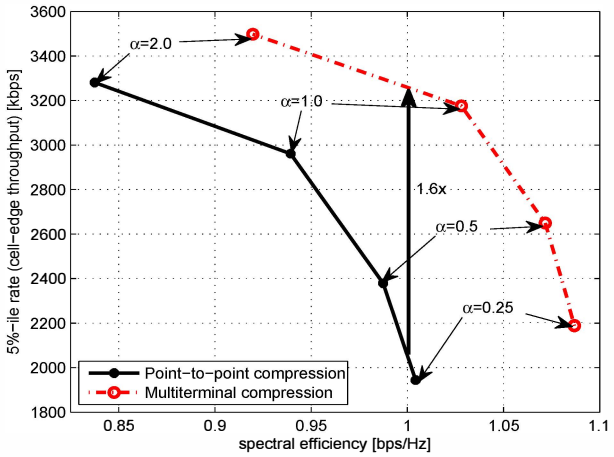


Figure 5. Cell-edge throughput, i.e., 5%-ile rate, versus the average spectral efficiency for various fairness constants α in the uplink C-RAN with $N = 3$ pico-BSs, $K = 5$ MSs, $(C_{\text{macro}}, C_{\text{pico}}) = (9, 3)$ bps/Hz, $T = 10$ and $\beta = 0.5$.

VI. Conclusion

In this work, we have studied the advantage of multiterminal backhaul compression techniques over standard point-to-point compression for the uplink and downlink of cloud radio access networks. The extensive simulations are based on standard cellular models and the results focused on performance metrics such as sum-rate, proportional-fairness utility and cell-edge throughput. As an example, we observed that multiterminal compression techniques provide performance gains of more than 60% for both the uplink and the downlink in terms of the cell-edge throughput.

References

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

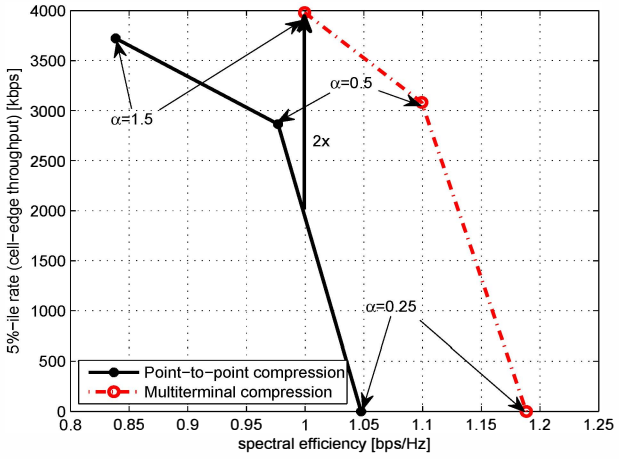


Figure 6. Cell-edge throughput versus the average spectral efficiency for various fairness constants α in the downlink C-RAN with $N = 1$ pico-BS, $K = 4$ MSs, $(C_{\text{macro}}, C_{\text{pico}}) = (3, 1)$ bps/Hz, $T = 5$ and $\beta = 0.5$.

[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] 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.

[5] Integrated Device Technology, Inc., "Front-haul compression for emerging C-RAN and small cell networks," Apr. 2013.

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

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

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

[9] S.-H. Park, O. Simeone, O. Sahin and S. Shamai (Shitz), "Robust and efficient distributed compression for cloud radio access networks," *IEEE Trans. Veh. Technology*, vol. 62, no. 2, pp. 692-703, Feb. 2013.

[10] L. Zhou and W. Yu, "Uplink multicell processing with limited backhaul via successive interference cancellation," in *IEEE Global Comm. Conf. (Globecom 2012)*, Anaheim, CA, Dec. 2012.

[11] S.-H. Park, O. Simeone, O. Sahin and S. Shamai (Shitz), "Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks," *IEEE Trans. Sig. Processing*, vol. 61, no. 22, pp. 5646-5658, Nov. 2013.

[12] 3GPP TR 36.931 ver. 9.0.0 Rel. 9, May 2011.

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

[14] P. Viswanth, D. Tse and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277-1294, Jun. 2002.

[15] 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.

[16] O. Simeone, O. Somekh, H. V. Poor and S. Shamai (Shitz), "Downlink multicell processing with limited-backhaul capacity," *EURASIP J. Adv. Sig. Proc.*, 2009.

[17] 3GPP TR 36.942 ver. 8.1.0 Rel. 8, Jan. 2009.

[18] L.-C. Wang and C.-J. Yeh, "3-cell network MIMO architectures with sectorization and fractional frequency reuse," *IEEE J. Sel. Areas Comm.*, vol. 29, no. 6, pp. 1185-1199, Jun. 2011.