Multivariate Backhaul Compression for the Downlink 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*—In the downlink of cloud radio access networks, a central encoder is connected to multiple multi-antenna base stations (BSs) via finite-capacity backhaul links. At the central encoder, precoding is followed by compression in order to produce the rate-limited bit streams delivered to each BS over the corresponding backhaul link. In current state-of-the-art schemes, the signals intended for different BSs are compressed independently. In contrast, this work proposes to leverage joint compression, also referred to as multivariate compression, of the signals for different BSs in order to better control the effect of the additive quantization noises at the mobile stations (MSs). The problem of maximizing the weighted sum-rate over precoding and compression strategies is formulated subject to power and backhaul capacity constraints. An iterative algorithm is proposed that achieves a stationary point of the problem. From numerical results, it is confirmed that the proposed joint precoding and compression strategy outperforms conventional approaches based on independent compression across the BSs.**

***Index Terms*—Cloud radio access network, constrained backhaul, precoding, multivariate compression, network MIMO.**

I. Introduction

In cloud radio access networks, the encoding/decoding functionalities of the base stations (BSs) are migrated to a central unit, which is connected to the BSs via backhaul links [1][2]. Therefore, a key issue in the design of cloud radio access networks is how to efficiently use the capacity-limited backhaul links connecting between the BSs and the central unit [3].

In the *uplink* of cloud radio access networks, each BS compresses its received signal for transmission to the central unit on its finite-capacity backhaul link. The central unit then performs joint decoding of all the mobile stations (MSs) based on all received compressed signals. Recent theoretical results have shown that *distributed compression* schemes [4] provide significant advantages over the conventional approach based on independent compression at the BSs. This is because the signals received by different BSs are statistically correlated [5]-[7], and hence distributed source coding can leverage the signals received from the other BSs as side information.

In the *downlink* of cloud radio access networks, the central encoder performs joint encoding and precoding of the messages intended for the MSs. Then, it independently compresses the produced baseband signals to be transmitted by each BS. These baseband signals are delivered via the backhaul links to the corresponding BSs, which simply upconvert and transmit them through their antennas. This system was studied in [8] assuming that the central encoder performs dirty-paper coding (DPC) [9] of all MSs' signals before compression.

In this work, we propose a novel approach for the compression on the backhaul links of cloud radio access networks in the downlink that can be seen as the counterpart of the distributed source coding strategy studied in [5]-[7] for the uplink. The key idea is that of allowing the quantization noise signals corresponding to different BSs to be correlated: by designing the correlation of the quantization noises across the BSs, it is possible to control the effects of the resulting quantization noise seen at the MSs. In order to create such correlation, we propose to jointly compress the baseband signals to be delivered over the backhaul links using so called *multivariate compression* [4, Ch. 9].

Multivariate compression allows compressed signals with correlated quantization noises to be produced at the expense of using larger quantization codebooks and hence larger output bit rates [4, Ch. 9]. For the application under study, this translates into more demanding backhaul capacity requirements. The benefits of correlation among the quantization noises at the BSs must then be weighted against the available backhaul capacity. Moreover, the backhaul requirements are also affected by the precoding strategy. Based on these observations, we tackle the problem of jointly optimizing the precoding matrix and the correlation matrix of the quantization noises across the BSs. The optimization is subject to the backhaul constraints resulting from multivariate compression [4, Ch. 9]. Numerical results illustrate the advantages offered by the proposed approach. Finally, a sequential, rather than joint, implementation of multivariate compression is proposed.

Notation: All logarithms are in base two unless specified. Given a sequence $X_1, \ldots, X_m$, we define a set $X_{\mathcal{S}} = \{X_j | j \in \mathcal{S}\}$ for a subset $\mathcal{S} \subseteq \{1, \ldots, m\}$. We also define the correlation matrices $\mathbf{\Sigma}_{\mathbf{x}} = \mathbb{E}[\mathbf{x}\mathbf{x}^\dagger]$, $\mathbf{\Sigma}_{\mathbf{x},\mathbf{y}} = \mathbb{E}[\mathbf{x}\mathbf{y}^\dagger]$ and $\mathbf{\Sigma}_{\mathbf{x}|\mathbf{y}} = \mathbb{E}[\mathbf{x}\mathbf{x}^\dagger | \mathbf{y}]$.

II. System Model

We consider the downlink of a cloud radio access network as illustrated in Fig. 1. In the system, a central encoder communicates to N_M MSs through N_B distributed BSs. The message M_k for each kth MS is uniformly distributed in the set $\{1, \ldots, 2^{nR_k}\}$, where n is blocklength and R_k is the

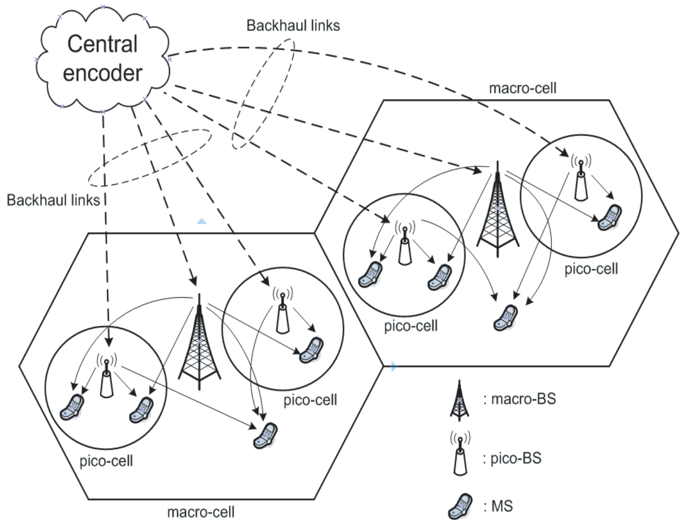

Figure 1. Downlink communication in a cloud radio access network in which there are N_B multi-antenna BSs and N_M multi-antenna MSs. The N_B BSs include both macro-BSs and pico/femto-BSs. The N_M MSs are distributed across all the cells.

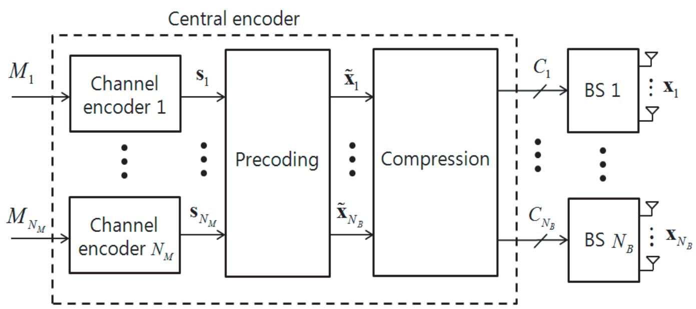

Figure 2. Illustration of the operation at the central encoder.

information rate of message M_k in bits per channel use (c.u.). Each MS k has $n_{M,k}$ receive antennas for $k = 1, \ldots, N_M$, and each BS i is equipped with $n_{B,i}$ antennas for $i = 1, \ldots, N_B$. Note that the BSs can be either macro-BSs or pico/femto-BSs and that the MSs are arbitrarily distributed across the cells. Each ith BS is connected to the central encoder via digital backhaul link with finite-capacity C_i bits per c.u. For notational convenience, we define $n_B = \sum_{i=1}^{N_B} n_{B,i}$ as the total number of transmitting antennas, $n_M = \sum_{k=1}^{N_M} n_{M,k}$ as the total number of receive antennas, and the sets $\mathcal{N_B} = \{1, \ldots, N_B\}$ and $\mathcal{N_M} = \{1, \ldots, N_M\}$.

As shown in Fig. 2, each message M_k is first encoded by a separate channel encoder, which produces a coded signal $\mathbf{s}_k$. The signal $\mathbf{s}_k \in \mathbb{C}^{r_k \times 1}$ corresponds to the $r_k \times 1$ vector of encoded symbols intended for the kth MS for a given c.u., and we have $r_k \leq n_{M,k}$. We assume that each coded symbol $\mathbf{s}_k$ is taken from a conventional Gaussian codebook so that we have $\mathbf{s}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. The signals $\mathbf{s}_1, \ldots, \mathbf{s}_{N_M}$ are further processed by the central encoder in two stages, namely *precoding* and *compression*. As is standard practice, precoding is used in order to control the interference between the data streams intended for the same MS and for different MSs. Compression is instead needed in order to produce the N_B rate-limited bit streams delivered to each BS over the corresponding backhaul link. Specifically, recall that each BS i receives up to C_i bits per c.u. on the backhaul link from the central encoder.

On the basis of the bits received on the backhaul links, each BS i produces a vector $\mathbf{x}_i \in \mathbb{C}^{n_{B,i} \times 1}$ for each c.u., which is the baseband signal to be transmitted from its $n_{B,i}$ antennas. We have the per-BS power constraints[1]

$$\mathbb{E}\left[||\mathbf{x}_i||^2\right] \leq P_i, \text{ for } i \in \mathcal{N_B}. \tag{1}$$

Assuming flat-fading channels, the signal $\mathbf{y}_k \in \mathbb{C}^{n_{M,k}}$ received by MS k is written as

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{z}_k, \tag{2}$$

where we have defined the aggregate transmit signal vector $\mathbf{x} = [\mathbf{x}_1^\dagger, \ldots, \mathbf{x}_{N_B}^\dagger]^\dagger$, the additive noise $\mathbf{z}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, and the channel matrix $\mathbf{H}_k \in \mathbb{C}^{n_{M,k} \times n_B}$ toward MS k as

$$\mathbf{H}_k = [\mathbf{H}_{k,1}\ \mathbf{H}_{k,2}\ \cdots\ \mathbf{H}_{k,N_B}], \tag{3}$$

with $\mathbf{H}_{k,i} \in \mathbb{C}^{n_{M,k} \times n_{B,i}}$ denoting the channel matrix from BS i to MS k. The channel matrices remain constant for the entire coding block duration. We assume that the central encoder has information about the global channel matrices $\mathbf{H}_k$ for all $k \in \mathcal{N_M}$ and that each MS k is only aware of the channel matrix $\mathbf{H}_k$. The BSs must also be informed about the compression codebooks used by the central encoder, as further detailed later.

Based on the definition given above and assuming single-user detection at each MS, the rates

$$R_k = I\left(\mathbf{s}_k; \mathbf{y}_k\right) \tag{4}$$

can be achieved for each MS $k \in \mathcal{N_M}$.

III. Proposed Approach and Problem Definition

In this section, we first propose a novel precoding-compression strategy based on multivariate compression for the downlink of a cloud radio access network. We then establish the problem definition.

A. Encoding Operation at the Central Encoder

As mentioned in the previous section, the operation at the central encoder can be represented by the block diagram in Fig. 2. Specifically, after channel encoding, the encoded signals $\mathbf{s} = [\mathbf{s}_1^\dagger, \ldots, \mathbf{s}_{N_M}^\dagger]^\dagger$ undergo precoding and compression, as detailed next.

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 $\mathbf{A} \in \mathbb{C}^{n_B \times n_M}$. The precoded data can be written as

$$\tilde{\mathbf{x}} = \mathbf{A}\mathbf{s}, \tag{5}$$

[1]The results in this paper can be immediately extended to the case with more general power constraints of the form $\mathbb{E}[\mathbf{x}^\dagger \mathbf{\Theta}_l \mathbf{x}] \leq \delta_l$ for $l \in \{1, \ldots, L\}$, where the matrix $\mathbf{\Theta}_l$ is a non-negative definite matrix (see, e.g., [10, Sec. I]).

where the matrix $\mathbf{A}$ can be factorized as

$$\mathbf{A} = [\mathbf{A}_1 \cdots \mathbf{A}_{N_M}], \tag{6}$$

with $\mathbf{A}_k \in \mathbb{C}^{n_B \times n_{M,k}}$ denoting the precoding matrix corresponding to MS k. The precoded data $\tilde{\mathbf{x}}$ can be written as $\tilde{\mathbf{x}} = [\tilde{\mathbf{x}}_1^\dagger, \ldots, \tilde{\mathbf{x}}_{N_B}^\dagger]^\dagger$, where the signal $\tilde{\mathbf{x}}_i$ is the $n_{B,i} \times 1$ precoded vector corresponding to the ith BS and given as

$$\tilde{\mathbf{x}}_i = \mathbf{E}_i^\dagger \mathbf{A}\mathbf{s}, \tag{7}$$

with the matrix $\mathbf{E}_i \in \mathbb{C}^{n_B \times n_{B,i}}$ having all zero elements except for the rows from $(\sum_{j=1}^{i-1} n_{B,j} + 1)$ to $(\sum_{j=1}^{i} n_{B,j})$ which contain an $n_{B,i} \times n_{B,i}$ identity matrix.

2. Compression: Each precoded data stream $\tilde{\mathbf{x}}_i$ for $i \in \mathcal{N}_\mathcal{B}$ must be compressed in order to allow the central encoder to deliver it to the ith BS through the backhaul link of capacity C_i bits per c.u. Each ith BS then simply transmits the compressed signal $\mathbf{x}_i$ obtained from the central encoder. Note that this implies that the BSs need not be aware of the channel codebooks and of the precoding matrix $\mathbf{A}$ used by the central encoder. Instead, they must be informed about the quantization codebooks selected by the central encoder.

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 $\mathbf{x}_i$ to be transmitted by BS i as[2]

$$\mathbf{x}_i = \tilde{\mathbf{x}}_i + \mathbf{q}_i, \tag{8}$$

where the compression noise $\mathbf{q}_i$ is modeled as a complex Gaussian vector distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{\Omega}_{i,i})$. Overall, the vector $\mathbf{x} = [\mathbf{x}_1^\dagger, \ldots, \mathbf{x}_{N_B}^\dagger]^\dagger$ of compressed signals for all the BSs is given by

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{q}, \tag{9}$$

where the compression noise $\mathbf{q} = [\mathbf{q}_1^\dagger, \ldots, \mathbf{q}_{N_B}^\dagger]^\dagger$ is modeled as a complex Gaussian vector distributed as $\mathbf{q} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega})$. The (i,j)-th block $\mathbf{\Omega}_{i,j} = \mathbb{E}[\mathbf{q}_i \mathbf{q}_j^\dagger]$ of the compression covariance $\mathbf{\Omega}$ defines the correlation between the quantization noises of BS i and BS j. Rate-distortion theory guarantees that compression codebooks can be found for any given covariance matrix $\mathbf{\Omega} \succeq \mathbf{0}$ under appropriate constraints imposed on the backhaul links' capacities. This aspect will be further discussed in Sec. III-B.

With the described precoding and compression operations, the achievable rate (4) for MS k is computed as

$$\begin{aligned} I(\mathbf{s}_k; \mathbf{y}_k) &= f_k(\mathbf{A}, \mathbf{\Omega}) \\ &\triangleq \log\det\left(\mathbf{I} + \mathbf{H}_k\left(\mathbf{A}\mathbf{A}^\dagger + \mathbf{\Omega}\right)\mathbf{H}_k^\dagger\right) \\ &\quad - \log\det\left(\mathbf{I} + \mathbf{H}_k\left(\sum_{l \in \mathcal{N}_\mathcal{M} \setminus \{k\}} \mathbf{A}_l \mathbf{A}_l^\dagger + \mathbf{\Omega}\right)\mathbf{H}_k^\dagger\right). \end{aligned} \tag{10}$$

Remark 1. If non-linear precoding via DPC [9] is deployed at the central encoder with a specific encoding permutation $\tilde{\pi} : \mathcal{N}_\mathcal{M} \rightarrow \mathcal{N}_\mathcal{M}$ of the MS indices $\mathcal{N}_\mathcal{M}$, the achievable rate $R_{\tilde{\pi}(k)}$ for MS $\tilde{\pi}(k)$ is given as $R_{\tilde{\pi}(k)} = I\left(\mathbf{s}_{\tilde{\pi}(k)}; \mathbf{y}_{\tilde{\pi}(k)} | \mathbf{s}_{\tilde{\pi}(1)}, \ldots, \mathbf{s}_{\tilde{\pi}(k-1)}\right)$ in lieu of (4) and can be calculated similar to (10).

[2] The test channel $\mathbf{x}_i = \mathbf{B}_i \tilde{\mathbf{x}}_i + \mathbf{q}_i$ is seemingly more general than (8), but this can be captured by adjusting the matrix $\mathbf{A}$ in (5).

B. Multivariate Backhaul Compression

As explained above, due to the fact that the BSs are connected to the central encoder via finite-capacity backhaul links, the precoded signals $\tilde{\mathbf{x}}_i$ in (7) for $i \in \mathcal{N}_\mathcal{B}$ are compressed before they are communicated to the BSs using the Gaussian test channels (8). In the conventional case in which the compression noise signals related to the different BSs are uncorrelated, i.e., $\mathbf{\Omega}_{i,j} = \mathbf{0}$ for all $i \neq j \in \mathcal{N}_\mathcal{B}$ as in [8], the signal $\mathbf{x}_i$ to be emitted from BS i can be reliably communicated from the central encoder to BS i if the condition

$$I(\tilde{\mathbf{x}}_i; \mathbf{x}_i) = \log\det\left(\mathbf{E}_i^\dagger \mathbf{A}\mathbf{A}^\dagger \mathbf{E}_i + \mathbf{\Omega}_{i,i}\right) - \log\det(\mathbf{\Omega}_{i,i}) \leq C_i \tag{11}$$

is satisfied for $i \in \mathcal{N}_\mathcal{B}$. This follows from standard rate-distortion theoretic arguments (see, e.g., [4, Ch. 3]). We emphasize that (11) is valid under the assumption that each BS i is informed about the quantization codebook used by the central encoder, as defined by the covariance matrix $\mathbf{\Omega}_{i,i}$.

In this paper, we instead propose to introduce correlation among the compression noise signals, i.e., to set $\mathbf{\Omega}_{i,j} \neq \mathbf{0}$ for $i \neq j$, in order to control the effect of the quantization noise at the MSs. In fact, introducing correlated quantization noises calls for joint, and not independent, compression of the precoded signals of different BSs. As seen, the family of compression strategies that produce descriptions with correlated compression noises is often referred to as *multivariate compression*. By choosing the test channel according to (9), we can leverage the multivariate covering lemma in [4, Ch. 9] to obtain sufficient conditions for the signal $\mathbf{x}_i$ to be reliably delivered to BS i for all $i \in \mathcal{N}_\mathcal{B}$. In Lemma 1, we use $\mathbf{E}_\mathcal{S}$ to denote the matrix obtained by stacking the matrices $\mathbf{E}_i$ for $i \in \mathcal{S}$ horizontally.

Lemma 1. *The signals $\mathbf{x}_1, \ldots, \mathbf{x}_{N_B}$ obtained via the test channel (9) can be reliably transferred to the BSs on the backhaul links if the condition*

$$\begin{aligned} g_\mathcal{S}(\mathbf{A}, \mathbf{\Omega}) &\triangleq \sum_{i \in \mathcal{S}} h(\mathbf{x}_i) - h(\mathbf{x}_\mathcal{S} | \tilde{\mathbf{x}}) \\ &= \sum_{i \in \mathcal{S}} \log\det\left(\mathbf{E}_i^\dagger \mathbf{A}\mathbf{A}^\dagger \mathbf{E}_i + \mathbf{\Omega}_{i,i}\right) - \log\det\left(\mathbf{E}_\mathcal{S}^\dagger \mathbf{\Omega} \mathbf{E}_\mathcal{S}\right) \\ &\leq \sum_{i \in \mathcal{S}} C_i \end{aligned} \tag{12}$$

is satisfied for all subsets $\mathcal{S} \subseteq \mathcal{N}_\mathcal{B}$.

Proof. The proof follows by applying the multivariate covering lemma in [4, Ch. 9]. □

Comparing (11) with (12) shows that the introduction of correlation among the quantization noises for different BSs leads to additional constraints on the backhaul link capacities.

C. Weighted Sum-Rate Maximization

Assuming the operation at the central encoder, BSs and MSs detailed above, we are interested in maximizing the weighted sum-rate $R_{\rm sum} = \sum_{k=1}^{N_M} w_k R_k$ subject to the backhaul constraints (12) over the precoding matrix $\mathbf{A}$ and the compression noise covariance $\mathbf{\Omega}$ for given weights $w_k \geq 0$, $k \in \mathcal{N}_{\mathcal{M}}$. This problem is formulated as

$$\underset{\mathbf{A},\,\mathbf{\Omega}\succeq\mathbf{0}}{\text{maximize}} \sum_{k=1}^{N_M} w_k f_k\left(\mathbf{A},\mathbf{\Omega}\right) \tag{13a}$$

$$\text{s.t.} \quad g_{\mathcal{S}}\left(\mathbf{A},\mathbf{\Omega}\right) \leq \sum_{i\in\mathcal{S}} C_i, \text{ for all } \mathcal{S} \subseteq \mathcal{N}_{\mathcal{B}}, \tag{13b}$$

$$\text{tr}\left(\mathbf{E}_i^{\dagger}\mathbf{A}\mathbf{A}^{\dagger}\mathbf{E}_i + \mathbf{\Omega}_{i,i}\right) \leq P_i, \text{ for all } i \in \mathcal{N}_{\mathcal{B}}. \tag{13c}$$

The condition (13b) corresponds to the backhaul constraints due to multivariate compression as introduced in Lemma 1, and the condition (13c) imposes the transmit power constraints (1). It is noted that the problem (13) is not easy to solve due to the non-convexity of the objective function $\sum_{k=1}^{N_M} w_k f_k\left(\mathbf{A},\mathbf{\Omega}\right)$ in (13a) and the functions $g_{\mathcal{S}}\left(\mathbf{A},\mathbf{\Omega}\right)$ in (13b) with respect to $(\mathbf{A},\mathbf{\Omega})$. In Sec. III-D, we will propose an algorithm to tackle the solution of problem (13).

D. MM Algorithm

Here, we discuss the optimization of the precoding matrix $\mathbf{A}$ and the compression covariance $\mathbf{\Omega}$ by solving problem (13). As mentioned, the optimization (13) is a non-convex problem. To tackle this issue, we first make a change of variable by defining the variables $\mathbf{R}_k \triangleq \mathbf{A}_k\mathbf{A}_k^{\dagger}$ for $k \in \mathcal{N}_{\mathcal{M}}$. The so obtained problem over the variables $\{\mathbf{R}_k\}_{k=1}^{N_M}$ and $\mathbf{\Omega}$ is still non-convex due to the second term in (10) and the first term in (12), which are concave in the variables $\{\mathbf{R}_k\}_{k=1}^{N_M}$ and $\mathbf{\Omega}$. However, we observe that this problem falls into the class of difference-of-convex (DC) problems [11] since both the negative of the objective function and the constraint function can be written as difference of convex functions. Thus, we can adopt the Majorization Minimization (MM) scheme [11], which solves a sequence of convex problems obtained by linearizing non-convex parts of the mentioned functions. It is known that the MM algorithm converges to a stationary point of the original non-convex problem (see, e.g., [11, Sec. 1.3.3]). The detailed algorithm is described in [12, Algorithm 1]. We observe that the same approach can be used to optimize compression and precoding with independent compression at the BSs by substituting (13b) with (11).

IV. Numerical Results

In this section, we present numerical results in order to investigate the advantage of the proposed approach based on multivariate compression as compared to the conventional approaches based on independent compression across the BSs. We assume that there is one MS active in each cell and we consider three cells, so that we have $N_B = N_M = 3$. Every BS is subject to the same power constraint P and has the same backhaul capacity C, i.e., $P_i = P$ and $C_i = C$ for $i \in \mathcal{N}_{\mathcal{B}}$.

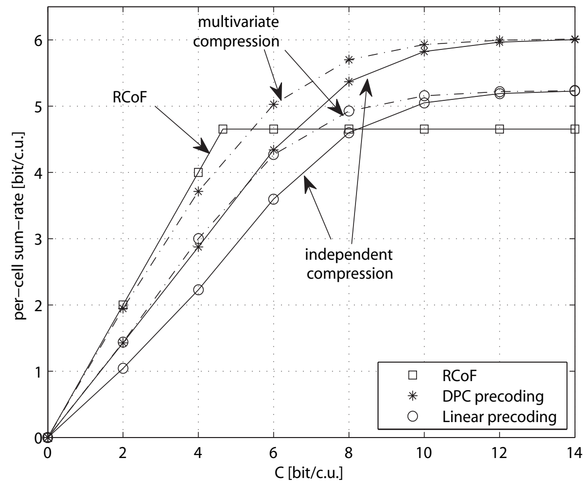

Figure 3. Per-cell sum-rate versus the backhaul capacity C for the circular Wyner model [13] with $P = 20$ dB and $g = 0.5$.

We start by considering as a benchmark the performance in a simple circulant Wyner model in Fig. 3. In this model, all MSs and BSs have a single antenna and the channel matrices $\mathbf{H}_{k,j}$ reduce to deterministic scalars given as $\mathbf{H}_{k,k} = 1$ for $k = 1, 2, 3$ and $\mathbf{H}_{k,j} = g \in [0, 1]$ for $j \neq k$ [13]. In the figure, we compare the performance of multivariate and independent compression applied to both linear precoding and DPC precoding (see Remark 1). We also show the performance of the reverse Compute-and-Forward (RCoF) scheme of [14]. It is observed that multivariate compression significantly outperforms the conventional independent compression strategy for both linear and DPC precoding. Moreover, RCoF in [14] remains the most effective approach in the regime of moderate backhaul C, although multivariate compression allows to compensate for most of the rate loss of standard DPC precoding in the low-backhaul regime. We finally observe that the lower saturation level of the rate achieved by RCoF for sufficiently large C is due to the integer constraints imposed on the function of the messages to be computed by the MSs [14].

In Fig. 4, we consider a more general MIMO fading channel, in which the elements of each channel matrix $\mathbf{H}_{k,i}$ are assumed to be i.i.d. complex Gaussian random variables with $\mathcal{CN}(0, 1)$. Moreover, each BS is assumed to use two transmit antennas while each MS is equipped with a single receive antenna. In the figure, the average sum-rate performance of the linear precoding and compression schemes is plotted versus the transmit power P with $C = 2$ bit/c.u. As a reference, we also plot the performance of a separate design of precoding and compression strategies, in which the precoding matrix $\mathbf{A}$ is fixed a priori to a standard sum-rate maximizing precoding scheme proposed in [15], and then the compression covariance $\mathbf{\Omega}$ is designed separately so as to maximize the sum-rate. It is seen that the gain of multivariate compression is more pronounced when each BS uses a larger transmit power. This

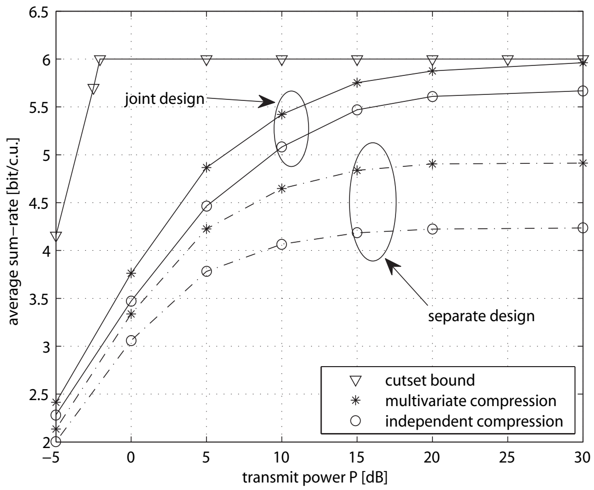

Figure 4. Average sum-rate versus the transmit power P for linear precoding with $C = 2$ bit/c.u..

implies that, as the received signal-to-noise ratio (SNR) increases, more efficient compression strategies are called for. In a similar vein, the importance of the joint design of precoding and compression is more significant when the transmit power is larger. Moreover, it is seen that multivariate compression is effective in partly compensating for the suboptimality of the separate design. Finally, we also plot the cutset bound, which is obtained as $\min\{R_{\text{full}}, 3C\}$, where R_{full} is the sum-capacity achievable when the BSs can fully cooperate under per-BS power constraints (see, e.g., [10, Sec. II] for the computation of R_{full}). It is seen that only the proposed joint design with multivariate compression approaches the cutset bound as the transmit power increases.

V. Concluding Remarks

In this work, we have studied the design of joint precoding and compression strategies for the downlink of cloud radio access networks where the BSs are connected to the central encoder via finite-capacity backhaul links. Specifically, we have proposed to leverage multivariate compression of the signals of different BSs in order to control the effect of the additive quantization noises at the MSs. Via numerical results, it was confirmed that the proposed approach based on multivariate compression outperforms the conventional approaches based on independent compression across the BSs.

The discussion in this work assumes that multivariate compression is carried out at the central processor by jointly compressing the signals for all BSs. However, as discussed in [12, Sec. IV-D], a simpler architecture based on successive steps of minimum mean squared error (MMSE) estimation and per-BS compression as in Fig. 5 can be often used as an alternative with no performance loss. In this architecture, the signals for different BSs are compressed one after the other according to a given order π. Moreover, compression takes place after an MMSE step with inputs given by the previously

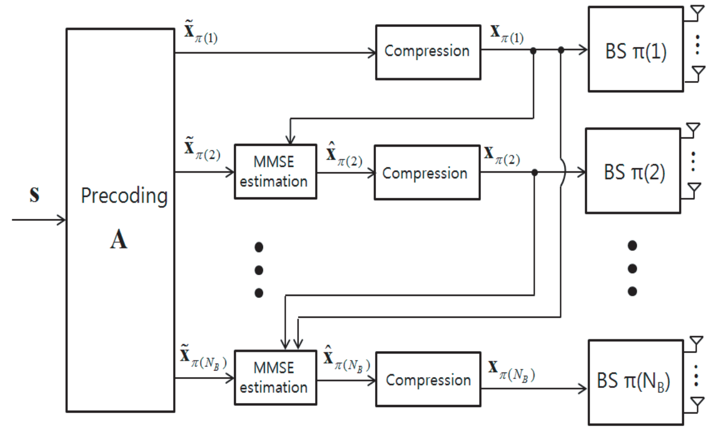

Figure 5. Sequential architecture for multivariate compression based on successive steps of MMSE estimation and per-BS compression.

compressed signals and the signal to be compressed. Robust design with respect to imperfect channel state information is also discussed in [12, Sec. V-D].

References

[1] J. Segel and M. Weldon, "Lightradio portfolio-technical overview," Technology White Paper 1, Alcatel-Lucent.
[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] A. E. Gamal and Y.-H. Kim, *Network information theory*, Cambridge University Press, 2011.
[5] 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.
[6] 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.
[7] L. Zhou and W. Yu, "Uplink multicell processing with limited backhaul via per-base-station successive interference cancellation," *IEEE. J. Sel. Areas Comm.*, vol. 31, no. 10, pp. 1981-1993, Oct. 2013.
[8] O. Simeone, O. Somekh, H. V. Poor and S. Shamai (Shitz), "Downlink multicell processing with limited-backhaul capacity," *EURASIP J. Adv. Sig. Proc.*, 2009.
[9] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439-441, May 1983.
[10] H. Huh, H. C. Papadopoulos and G. Caire, "Multiuser MISO transmitter optimization for intercell interference mitigation," *IEEE Trans. Sig. Proc.*, vol. 58, no. 8, pp. 4272-4285, Aug. 2010.
[11] 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.
[12] 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. Proc.*, vol. 61, no. 22, pp. 5646-5658, Nov. 2013.
[13] D. Gesbert, S. Hanly, H. Huang, S. Shamai (Shitz), O. Simeone and W. Yu, "Multi-cell MIMO cooperative networks: a new look at interference," *IEEE J. Sel. Areas Comm.*, vol. 28, no. 9, pp. 1380-1408, Dec. 2010.
[14] S.-N. Hong and G. Caire, "Compute-and-forward strategy for cooperative distributed antenna systems," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5227-5243, Sep. 2013.
[15] C. T. K. Ng and H. Huang, "Linear precoding in cooperative MIMO cellular networks with limited coordination clusters," *IEEE J. Sel. Areas Comm.*, vol. 28, no. 9, pp. 1446-1454, Dec. 2010.