Generalized Compression Strategy for the Downlink Cloud Radio Access Network

Pratik Patil and Wei Yu
Department of Electrical and Computer Engineering
University of Toronto, Toronto, ON, M5S 3G4, Canada
Emails: pratik.patil@mail.utoronto.ca, weiyu@comm.utoronto.ca

***Abstract*—This paper studies the downlink of a cloud radio access network (C-RAN), in which the base stations (BSs) are connected to a central processor (CP) via finite-capacity fronthaul links, modeled as a two-hop broadcast-relay network. We focus on a compression-based strategy in which the CP jointly encodes the signals to be broadcasted by the BSs, then compresses and sends these signals to the BSs through the fronthaul links. The paper characterizes an achievable rate region for a generalized compression strategy with Marton's multicoding for broadcasting and multivariate compression for fronthaul transmission. We then compare this rate region with the distributed decode-forward (DDF) scheme, which achieves the capacity of a general relay network to within a constant gap, and show that the difference lies in that DDF performs Marton's multicoding and multivariate compression jointly as opposed to successively as in the generalized compression strategy. A main result of this paper is that under an assumption of a *sum* fronthaul capacity constraint this difference is immaterial, so the successive encoding based compression strategy can already achieve the capacity region of the C-RAN to within a constant gap, where the gap is independent of the channel parameters and the power constraints at the BSs. For the special case of the Gaussian network, we further establish that under individual fronthaul constraints, the generalized compression strategy achieves to within a constant gap to the *sum* capacity of the C-RAN.**

A full version is accessible at arxiv.org/abs/1801.00394.

I. Introduction

Cloud radio access network (C-RAN) is a promising architecture for future cellular networks in which the base stations (BSs) are connected to a centralized processor (CP) through wired or wireless fronthaul links [1]. Information theoretically, the downlink C-RAN can be modeled as a broadcast-relay channel: the CP broadcasts the user messages through the BSs via the fronthaul links and the BSs act as relays for the users. This paper focuses on characterizing the theoretical achievable rate region of the downlink C-RAN with noiseless digital fronthaul links between the CP and BSs. Ideally with infinite fronthaul capacity links, downlink C-RAN model reduces to a multi-antenna broadcast channel for which cooperative beamforming combined with dirty-paper coding (DPC) [2] is known to be optimal. For the practical situation with finite fronthaul capacity links, the optimal coding strategy must combine both broadcasting and relaying, and is currently still an open problem. This paper makes progress by establishing the achievable rate region of a generalized compression strategy and showing that it is approximately optimal for the downlink C-RAN under certain conditions.

Several transmission strategies have been studied for the downlink C-RAN, including data-sharing [3], compression-based [4], reverse compute-forward [5], hybrid data-sharing and compresssion [6], that differ in how the fronthaul links are utilized and where the encoding of user messages is performed. We focus on the compression strategy in which the encoding is performed at the CP and the encoded analog signals are compressed and sent to the BSs to accommodate the fronthaul capacity constraints. The BSs then transmit the decompressed signals to the users. The aim of the paper is to understand the compression strategy from an information theoretical point of view.

As pointed earlier, with infinite fronthaul capacity, DPC achieves the capacity region for a Gaussian C-RAN. With finite fronthaul capacity, DPC and linear precoding schemes cannot be applied directly. A compressed version of DPC using independent compression across the BSs is introduced in [4]. The independent compression scheme can be further improved using a multivariate compression strategy across all the BSs [7] to better control effect of quantization noises by correlating them. The achievable rate expressions under linear beamforming and multivariate compression for the Gaussian C-RAN model are given in [7] and the corresponding achievable rate region using dirty paper coding followed by multivariate compression is given in [1]. This paper provides an achievable rate region of a general form of the compression strategy that includes Marton's multicoding followed by multivariate compression for a general discrete memoryless channel (DMC) in the second hop of C-RAN.

Can such generalized compression strategy approach the information theoretic capacity region of the C-RAN model? Toward answering this question, this paper draws inspiration from a new coding strategy named distributed decode-forward (DDF) [8] for broadcasting multiple messages over a general relay network, which has been shown to achieve the capacity region of the general Gaussian broadcast relay network to within a constant gap that is independent of the channel parameters and the power constraints and linear in the number of nodes in the network. When specialized to the downlink C-RAN model, the gap can be further improved from linear to logarithmic in the number of users and BSs [9]. Further, an improvement to the DDF strategy using a common message in Marton's coding is proposed in [10] for a 2-user 2-BS C-RAN model that allows for conferencing between the BSs.

This paper makes an observation that when specialized to C-RAN model, the DDF strategy resembles the generalized compression strategy, but with a crucial difference that instead of performing the compression *followed* by Marton's multicoding, DDF performs both the Marton's coding and multivariate compression *jointly* at the CP. As practical implementation for performing successive Marton's coding and multivariate compression would likely be much easier, we ask in this paper whether there are conditions under which the difference is immaterial. One of the main results of this paper is that under a sum fronthaul constraint, this is indeed true for a general DMC on the second hop. Furthermore, in the special case of Gaussian networks, we show that Marton's encoding followed by multivariate compression achieves the sum capacity of the downlink C-RAN to within a constant gap.

Rest of the paper is organized as follows. Section II provides the mathematical model for downlink C-RAN. Section III provides the achievable rate region for the generalized compresion strategy. Section IV specializes the DDF strategy for downlink C-RAN. In Section V, we compare two rate regions and provide conditions under which the two coincide. Finally, Section VI concludes the paper. We follow the notation of [2] throughout the paper. In addition, total correlation between a group of random variables indexed by a set $\mathcal{S}$ is defined as

$$T(X(\mathcal{S})) = \sum_{l \in \mathcal{S}} H(X_l) - H(X(\mathcal{S})). \tag{1}$$

II. System Model

Consider the downlink of a C-RAN comprising of a CP, L BSs, and K users. The CP communicates with BSs through noiseless fronthaul links of finite capacities. Let C_l denote the capacity of the fronthaul link from the central processor to BS l, $l \in \mathcal{L} := [1 : L]$. We assume a general DMC $p(y_1, \dots, y_K | x_1, \dots, x_L)$ between the BSs and the users. Let the intended message for user k be denoted by M_k, $k \in \mathcal{K} := [1 : K]$. A $(2^{nR_1}, \dots, 2^{nR_K}, n)$ code for the downlink C-RAN consists of a mapping at the CP from the K user messages $(m_1, \dots, m_K) \in [1 : 2^{nR_1}] \times \dots \times [1 : 2^{nR_K}]$ to L indices $(t_1, \dots, t_L) \in [1 : 2^{nC_1}] \times \dots \times [1 : 2^{nC_L}]$; encoders at the L BSs that map the index t_l to a codeword $x_l^n(t_l)$; decoders at the K users that estimate $\hat{m}_k$ based on the received signals y_k^n. The average probability of error is defined as $P_e^{(n)} = P\{\hat{m}_k \neq m_k \text{ for some } k \in \mathcal{K}\}$. A rate tuple $(R_1, \dots, R_K)$ is achievable if there exists a sequence of codes such that $\lim_{n \to \infty} P_e^{(n)} = 0$.

Of particular interest is the special case where the channel between the BSs and the users is a vector additive white Gaussian noise channel such that the received signal at user k is given by

$$Y_k = \sum_{l \in \mathcal{L}} h_{k,l} X_l + Z_k, \tag{2}$$

where Z_k are independent Gaussian noises with zero mean and variance σ^2. We assume all the BSs have an average power constraint of P without loss of generality. For simplicity, both the BSs and the users are assumed to be equipped with a single antenna in this paper.

III. Generalized Compression Strategy

The compression strategy has been extensively studied in the literature [4], [7]. The coding strategy involves two steps. First, the CP jointly encodes the user messages. Second, the encoded signals are compressed in order to accommodate them through the fronthaul links. Different options for joint encoding include linear beamforming strategies such as zero-forcing or regularized zero-forcing, or non-linear beamforming strategy such as DPC. Different options for compression include independent compression or multivariate compression.

The key point of this section is that these specific compression strategies previously studied in [4], [7] are special forms of a generalized compression strategy in which the joint encoding is performed via Marton's multicoding. This section characterizes the achievable rate region of such a generalized compression strategy assuming a general DMC on the second hop of C-RAN.

Theorem 1: A rate tuple $(R_1, \dots, R_K)$ is achievable for the downlink C-RAN using the compression strategy with Marton's multicoding followed by multivariate compression if

$$\sum_{k \in \mathcal{D}} R_k < \sum_{k \in \mathcal{D}} I(U_k; Y_k) - T(U(\mathcal{D})) \tag{3}$$

for all $\mathcal{D} \subseteq \mathcal{K}$ such that

$$\sum_{l \in \mathcal{S}} C_l > I(U(\mathcal{K}); X(\mathcal{S})) + T(X(\mathcal{S})) \tag{4}$$

for all $\mathcal{S} \subseteq \mathcal{L}$ for some distribution $p(u_1, \dots, u_K, x_1, \dots, x_L)$.

The set of inequalities (3) represents the achievable user rates using Marton's multicoding for broadcast channels. In linear beamforming, the U's are just the messages and are thus independent of each other. The advantage of using Marton's multicoding is to introduce correlation among U's for the possibility of increased rates. But doing so incurs a penalty that depends on the total correlation present among U's. DPC is an example of such Marton's coding.

The set of inequalities (4) represents the multivariate compression of $U(\mathcal{K})$ into X's that are transmitted by the BSs. If the BSs were co-located and can cooperate, the amount of quantization needed for compression is simply the first term $I(U(\mathcal{K}); X(\mathcal{S}))$. If the BSs are distributed and cannot cooperate, there's a penalty in terms of the correlation between the signals transmitted by the BSs.

The above achievability region has been presented at [11] and is subsequently generalized in [10] to the case with common information and BS cooperation when there are two BSs in the C-RAN.

IV. Distributed Decode-Forward

The main objective of this paper is to understand whether the generalized compression strategy can approximately achieve the capacity region of the C-RAN model. Toward this end, we examine the DDF strategy [8], which is a general coding scheme for broadcasting multiple messages over a general relay network that combines Marton's coding for the broadcast channel with partial decode-forward for the relay channel. The coding scheme involves using auxiliary random variables at each node in the network that implicitly carry information about the user messages. By specializing DDF to the C-RAN setup, we write down a succinct form of the achievable rate region using DDF that can be readily compared with the generalized compression strategy.

Theorem 2 ([8]): A rate tuple $(R_1, \dots, R_K)$ is achievable for the downlink C-RAN using the DDF strategy if

$$\sum_{k\in\mathcal{D}} R_k < \sum_{k\in\mathcal{D}} I(U_k;Y_k) + \sum_{l\in\mathcal{S}} C_l - T(U(\mathcal{D}), X(\mathcal{S})) \quad (5)$$

$$= \sum_{k\in\mathcal{D}} I(U_k;Y_k) - T(U(\mathcal{D})) + \sum_{l\in\mathcal{S}} C_l - I(U(\mathcal{D}); X(\mathcal{S})) - T(X(\mathcal{S})) \quad (6)$$

for all $\mathcal{D} \subseteq \mathcal{K}$ and $\mathcal{S} \subseteq \mathcal{L}$ for some distribution $p(u_1, \dots, u_K, x_1, \dots, x_L)$.

Comparing Theorem 2 with Theorem 1, we observe that DDF generalizes the compression strategy by jointly encoding the Marton's and compression codewords as opposed to successively as in the compression strategy. This allows the DDF strategy to consider distribution $p(u_1, \dots, u_K, x_1, \dots, x_L)$ that can violate the fronthaul constraint (4) in compression.

A key advantage of enlarging the allowable distributions to beyond the ones that explicitly satisfy the fronthaul constraints is that it permits a proof that the DDF strategy can achieve to within a constant gap to the cut-set bound of the general Gaussian broadcast relay channel [8]. The ingenious choice of $p(u_1, \dots, u_K|x_1, \dots, x_L)$ proposed in [8] that accomplishes this task is a distribution for $p(u_1, \dots, u_K|x_1, \dots, x_L)$ that tries to mimic the Gaussian channel distribution $p(y_1, \dots, y_K|x_1, \dots, x_L)$.

We now specialize the result of [8] to the C-RAN setup, where the second hop is a Gaussian network

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{Z}, \quad (7)$$

where $\mathbf{Y} = [Y_1, \dots, Y_K]^T$ are the received signals at the K users, $\mathbf{X} = [X_1, \dots, X_L]^T$ are the transmitted signals from the L BSs, $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_k]^T$ is the $K \times L$ channel matrix consisting of channel vectors $\mathbf{h}_1$ to $\mathbf{h}_k$ for users 1 to K, respectively, and $\mathbf{Z} = [Z_1, \dots, Z_K]^T \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ is the additive white Gaussian noise. The DDF strategy can be shown to achieve to within a constant gap to the cut-set outer bound by choosing $\mathbf{X}$ to be a vector of L independent Gaussian random variables $\mathcal{N}(0, P)$ and by choosing

$$\mathbf{U} = \mathbf{H}\mathbf{X} + \tilde{\mathbf{Z}}, \quad (8)$$

where $\tilde{\mathbf{Z}} \sim \mathcal{N}(0, \sigma^2 I)$ is independent of $\mathbf{Z}$. With this choice of $p(u_1, \dots, u_K|x_1, \dots, x_L)$, the DDF strategy achieves a rate region which is within a constant gap to the capacity region of C-RAN, where the gap is independent of the channel, the BS power constraints, and the fronthaul constraints [9].

A natural question at this point is whether we can use the generalized compression strategy to accomplish the same. The next section gives some partial answer in the affirmative but under specific conditions.

V. Compression versus DDF

As DDF generalizes the compression strategy, the achievable rate region of the compression strategy is a subset of the DDF region. This section asks the question of whether this subset inclusion is strict. The main result here is that under certain conditions the rate regions of the two strategies actually coincide. Specifically, we show that under a sum fronthaul constraint, the rate regions of the two strategies coincide for a general DMC on the second hop of C-RAN. As a second result of this section, we show that in the special case of Gaussian networks but under individual fronthaul constraint, the compression strategy achieves the sum capacity of C-RAN to within a constant gap. These results are useful, because successive Marton's coding and multivariate compression is much easier to implement in practice than DDF.

A. Sum Rate Under Sum Fronthaul Constraint

From (21) and (22), we have that achievable sum rate R^s_{DDF} for the DDF strategy under the sum fronthaul constraint C satisfies

$$R^s_{\text{DDF}} \le I(U_k;Y_k) - T(U(\mathcal{K})) + \min\left\{0, C - I(U(\mathcal{K}); X(\mathcal{L})) - T(X(\mathcal{L}))\right\}, \quad (9)$$

for some distribution $p(u_1, \dots, u_K, x_1, \dots, x_L)$. Similarly, from (19), the achievable sum rate R^s_{COM} using the compression strategy under the sum fronthaul constraint is given by

$$R^s_{\text{COM}} \le I(U_k;Y_k) - T(U(\mathcal{K})), \quad (10)$$

under some distribution $p(u_1, \dots, u_K, x_1, \dots, x_L)$ that satisfies

$$C \ge I(U(\mathcal{K}); X(\mathcal{L})) + T(X(\mathcal{L})), \quad (11)$$

from (20).

Theorem 3: Maximizing R^s_{DDF} over distributions $p(u_1, \dots, u_K, x_1, \dots, x_L)$ is equivalent to maximizing R^s_{COM} over distributions $p(u_1, \dots, u_K, x_1, \dots, x_L)$ that satisfy the sum fronthaul constraint given in (11).

Proof 1: We use proof by contradiction. Assume that the maximizing distribution $p(u_1, \dots, u_K, x_1, \dots, x_L)$ is such that $I(U(\mathcal{K}); X(\mathcal{L})) + T(X(\mathcal{L})) > C$. We show that we can alter any such distribution and produce a strictly better sum rate producing a contradiction. We remark that this proof is similar in spirit to the proof of equivalence between the unconstrained and constrained forms of compress-and-forward for the three-node relay channel, e.g., see Appendix 16C of [2]. Let us first rewrite the DDF sum rate R^s_{DDF} in a different form. It can be shown that we can write R^s_{DDF} equivalently as

$$R^s_{\text{DDF}} \le I(U(\mathcal{K}); X(\mathcal{L})) - \sum_{k=1}^{K} I(U_k; U_{k-1}, \dots, U_1, X(\mathcal{L})|Y_k) + \min\left\{0, C - I(U(\mathcal{K}); X(\mathcal{L})) - T(X(\mathcal{L}))\right\} \quad (12)$$

$$\le \min\Big\{I(U(\mathcal{K}); X(\mathcal{L})) - \sum_{k=1}^{K} I(U_k; U_{k-1}, \dots, U_1, X(\mathcal{L})|Y_k),$$

$$C-\sum_{k=1}^{K} I(U_k;U_{k-1},\ldots,U_1,X(\mathcal{L})|Y_k)-T(X(\mathcal{L}))\},$$

(13)

where, in this case, the expression $I(U_k;U_{k-1},\ldots,U_1,X(\mathcal{L})|Y_k)$ involves additional $U_{k-1},\ldots,U_1$ as we do not assume any assumptions on the distribution $p(u_1,\ldots,u_K,x_1,\ldots,x_L)$. With this equivalent way of writing the sum rate, under the assumption that $I(U(\mathcal{K});X(\mathcal{L}))+T(X(\mathcal{L}))>C$, we have

$$R^s_{\text{DDF}}=C-\sum_{k=1}^{K} I(U_k;U_{k-1},\ldots,U_1,X(\mathcal{L})|Y_k)-T(X(\mathcal{L})). \tag{14}$$

Now, we alter the distribution $p(u_1,\ldots,u_K,x_1,\ldots,x_L)$ in the following way. In the altered distribution, we keep $(X_1,\ldots,X_L)$ the same, but conditioned on $(X_1,\ldots,X_L)$, we set $(U'_1,\ldots,U'_K)=(U_1,\ldots,U_K)$ with probability p and and set it to 0 with probability $(1-p)$. Then, $(U'_1,\ldots,U'_K)\to(U_1,\ldots,U_K)\to(X_1,\ldots,X_L)\to(Y_1,\ldots,Y_K)$ form a Markov chain. With this altered distribution, we have that

$$\begin{aligned}&C-\sum_{k=1}^{K} I(U'_k;U'_{k-1},\ldots,U'_1,X(\mathcal{L})|Y_k)-T(X(\mathcal{L}))\\&>C-\sum_{k=1}^{K} I(U_k;U_{k-1},\ldots,U_1,X(\mathcal{L})|Y_k)-T(X(\mathcal{L})),\end{aligned} \tag{15}$$

because each of the terms $I(U'_k;U'_{k-1},\ldots,U'_1,X(\mathcal{L})|Y_k)$ is now reduced by $(1-p)$ fraction, while

$$\begin{aligned}&I(U'(\mathcal{K});X(\mathcal{L}))-I(U'_k;U'_{k-1},\ldots,U'_1,X(\mathcal{L})|Y_k)\\&<I(U(\mathcal{K});X(\mathcal{L}))-I(U_k;U_{k-1},\ldots,U_1,X(\mathcal{L})|Y_k),\end{aligned} \tag{16}$$

because the $I(U(\mathcal{K});X(\mathcal{L}))-I(U_k;U_{k-1},\ldots,U_1,X(\mathcal{L})|Y_k)$, which is positive, on whole decreases by a factor of $(1-p)$. As a consequence of this, we can observe from (13) that the resulting sum rate R^s_{DDF} is higher with the altered distribution, thus producing the required contradiction. Since the mutual information terms are continuous functions of p, continuing to decrease p, we also see that we can keep increasing the sum rate until the two terms in the minimum match, in which case, we have

$$C=I(U(\mathcal{K});X(\mathcal{L}))+T(X(\mathcal{L})), \tag{17}$$

i.e., the optimal distribution to maximize the sum rate must be such that the sum fronthaul constraint is met with equality, and the corresponding sum rate is given by

$$R^s_{\text{DDF}}=I(U_k;Y_k)-T(U(\mathcal{K})), \tag{18}$$

which is exactly the same as the compression strategy. In other words, there is no harm the sum rate due to performing the generalized compression strategy.

B. *Rate Region Under Sum Fronthaul Constraint*

Under a sum fronthaul constraint C, only the constraint for $\mathcal{S}=\mathcal{L}$ is active in (4) for the compression strategy. Therefore, the achievable rate tuples $(R_1,\ldots,R_K)$ satisfy

$$\sum_{k\in\mathcal{D}} R_k<\sum_{k\in\mathcal{D}} I(U_k;Y_k)-T(U(\mathcal{D})) \tag{19}$$

over all $\mathcal{D}\subseteq\mathcal{K}$ such that

$$C>I(U(\mathcal{K});X(\mathcal{L}))+T(X(\mathcal{L})) \tag{20}$$

for some distribution $p(u_1,\ldots,u_K,x_1,\ldots,x_L)$. Similarly, for the DDF strategy, the active constraints correspond to $\mathcal{S}=\mathcal{L}$ or $\mathcal{S}=\emptyset$. The achievable rate tuples $(R_1,\ldots,R_K)$ thus satisfy

$$\sum_{k\in\mathcal{D}} R_k<\sum_{k\in\mathcal{D}} I(U_k;Y_k)-T(U(\mathcal{D})) \tag{21}$$

$$\begin{aligned}\sum_{k\in\mathcal{D}} R_k<&\sum_{k\in\mathcal{D}} I(U_k;Y_k)-T(U(\mathcal{D}))\\&+C-I(U(\mathcal{D});X(\mathcal{L}))-T(X(\mathcal{L}))\end{aligned} \tag{22}$$

over all $\mathcal{D}\subseteq\mathcal{K}$ for some $p(u_1,\ldots,u_K,x_1,\ldots,x_L)$.

Definition 1: Let $\mathcal{R}^{\text{s}}_{\text{COM}}(C)$ denote the closure of the convex hull of achievable rate-sum-fronthaul tuples $(R_1,\ldots,R_K,C)$ using the generalized compression strategy satisfying (19)-(20) over all joint distributions $p(u_1,\ldots,u_K,x_1,\ldots,x_L)$ satisfying possibly input constraints on $(x_1,\ldots,x_L)$. Similarly, define the set $\mathcal{R}^{\text{s}}_{\text{DDF}}(C)$ for the DDF strategy satisfying (21) and (22) in respective manner.

Theorem 4: For the downlink C-RAN with a general DMC $p(y_1,\ldots,y_K|x_1,\ldots,x_L)$ in the second hop and a sum fronthaul constraint C, we have $\mathcal{R}^{\text{s}}_{\text{COM}}(C)=\mathcal{R}^{\text{s}}_{\text{DDF}}(C)$.

The proof makes use of the polymatroidal structure of the rate region to characterize all the corner points of the rate region achieved by the DDF strategy and construct appropriate time-shared compression strategies (by shutting down some users) to achieve all such corner points.

Since the DDF strategy is known to achieve the rate region of the C-RAN to within a constant gap for the Gaussian network, having the generalized compression rate region coincide with the DDF region under the sum fronthaul constraint immediately gives us the following corollary.

Corollary 1: For a Gaussian C-RAN under a sum fronthaul constraint C, the compression strategy achieves a rate region to within a constant gap to the capacity region, where the gap is independent of the channel, the BS power constraints, and the sum fronthaul constraint, and only depends on the number of BSs and users.

As a remark, we wonder whether the generalized compression and DDF rate regions coincide also under individual fronthaul constraints. While the answer is still not yet clear, we note here that the successive coding strategy of computing Marton's codewords first, then forming the compression codewords is not the only way to perform successive encoding. There is also the possibility of interleaving the encoding of U's and X's and it is perhaps necessary in general. However, the next section shows that if we only consider the sum rate, the two-step encoding of the generalized compression strategy indeed achieves the sum capacity of C-RAN to within a constant gap, even under individual fronthaul constraints, if we assume a Gaussian channel $p(y_1, \ldots, y_K | x_1, \ldots, x_L)$.

C. Sum Rate Under Individual Fronthaul Constraints

Consider the Gaussian C-RAN model specified in (7). Recall if we set $p(u_1, \ldots, u_K, x_1, \ldots, x_L)$ according to (8), the DDF strategy can be shown to achieve to within a constant gap to the capacity region. We show in this section that the sum rate achieved by the DDF strategy for the Gaussian C-RAN under the constant-gap distribution can also be achieved using the generalized compression strategy under the same set of fronthaul constraints.

The sum rate achieved by the DDF strategy is given by R that satisfies

$$R < \sum_{k \in \mathcal{K}} I(U_k; Y_k) + \sum_{l \in \mathcal{S}} C_l - T(U(\mathcal{K}), X(\mathcal{S})), \quad (23)$$

for all $\mathcal{S} \subseteq \mathcal{L}$ under some $p(u_1, \ldots, u_K, x_1, \ldots, x_L)$. The sum rate achieved by the generalized compression strategy is given by R that satisfies

$$R < \sum_{k \in \mathcal{K}} I(U_k; Y_k) - T(U(\mathcal{K})) \quad (24)$$

$$\sum_{l \in \mathcal{S}} C_l > I(U(\mathcal{K}); X(\mathcal{S})) - T(X(\mathcal{S})), \quad (25)$$

for all $\mathcal{S} \subseteq \mathcal{L}$ under some $p(u_1, \ldots, u_K, x_1, \ldots, x_L)$.

Definition 2: Consider the closure of the convex hull of achievable sum-rate-fronthaul tuples $(R, C_1, \ldots, C_L)$ for the C-RAN with the Gaussian channel model (7) using the DDF strategy as expressed in (23) under the constant-gap distribution (8) with the BS powers constrained by the power constraint P. Define $R^{\text{g}}_{\text{DDF}}$ to be the maximum sum rate under individual fronthaul constraints $(C_1, \ldots, C_L)$ in this set. Similarly, define $R^{\text{g}}_{\text{COM}}$ for the generalized compression strategy as expressed in (24)-(25) in respective manner.

Comparing the sum rate of DDF in (23) with the sum rate of generalized compression in (24)-(25), we clearly have $R^{\text{g}}_{\text{COM}} \leq R^{\text{g}}_{\text{DDF}}$. We show that actually $R^{\text{g}}_{\text{COM}} = R^{\text{g}}_{\text{DDF}}$. As a consequence, we have the following theorem.

Theorem 5: For a memoryless Gaussian channel on the second hop of C-RAN, the compression scheme achieves a sum rate to within a constant gap to the cut-set bound under arbitrary fronthaul constraints $(C_1, \ldots, C_L)$ where the gap is independent of the channel parameters, the BS power constraints, and the individual fronthaul constraints, and only depends on the number of BSs and users.

The proof uses the contra-polymatroidal structure of the fronthaul region to characterize the corner points for each fixed sum rate R under the DDF strategy. Using appropriate time-sharing schemes in the generalized compression strategy (by shutting down some BSs), we show that each such corner point is achievable in the compression strategy with a sum rate at least as large as R.

As a final remark, [9] shows that the gap between the achievable rate region and the cut-set bound for DDF can be refined in donwlink C-RAN to being logarithmic in the number of BSs and users, instead of being linear as in [8]. The refinement uses a slightly modified form of the constant-gap distribution. The equivalence result shown in this section also works for this modified constant-gap distribution. Thus, a similar refinement can be used to conclude that the compression strategy can achieve the sum capacity of the C-RAN network to within a constant gap which is a logarithmic in the number of BSs and users.

VI. Conclusion

This paper investigates the compression strategy for the downlink C-RAN from an information theoretic point of view. The paper first generalizes the existing compression strategies to include Marton's multicoding followed by multivariate compression for a C-RAN with a general DMC in the second hop. When compared with the DDF strategy specialized to the downlink C-RAN, it is observed that DDF is a generalization of the compression strategy where the Marton's multicoding and the multivariate compression are done jointly as opposed to successively as in the compression strategy. The paper then shows that under a sum fronthaul constraint, such generalization does not lead to higher rates. Furthermore, for the case of Gaussian network, the paper shows that the two-phase compression strategy achieves a sum rate that is within a constant gap to the cut-set bound. These results provide a justification for the practical choice of the two-phase compression strategy for the downlink C-RAN.

References

[1] O. Simeone, A. Maeder, M. Peng, O. Sahin, and W. Yu, "Cloud radio access network: Virtualizing wireless access for dense heterogeneous systems," *J. Commun. Networks*, vol. 18, no. 2, pp. 135–149, 2016.

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

[3] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access*, vol. 2, pp. 1326–1339, 2014.

[4] O. Simeone, O. Somekh, H. Poor, and S. Shamai, "Downlink multicell processing with limited-backhaul capacity," *EURASIP J. Advances Singal Process.*, Feb. 2009.

[5] S.-N. Hong and G. Caire, "Compute-and-forward strategies for cooperative distributed antenna systems," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5227–5243, 2013.

[6] P. Patil and W. Yu, "Hybrid compression and message-sharing strategy for the downlink cloud radio-access network," in *Inf. Theory Applicat. Workshop (ITA)*, Feb. 2014.

[7] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5646–5658, Nov. 2013.

[8] S. H. Lim, K. T. Kim, and Y. H. Kim, "Distributed decode-forward for relay networks," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4103–4118, July 2017.

[9] S. Ganguly and Y.-H. Kim, "On the capacity of cloud radio access networks," in *IEEE Int. Symp. Inf. Theory (ISIT)*, 2017.

[10] C. Wang, M. A. Wigger, and A. Zaidi, "On achievability for downlink cloud radio access networks with base station cooperation," *CoRR*, 2016. [Online]. Available: http://arxiv.org/abs/1610.09407

[11] W. Yu, "Cloud radio-access networks: Coding strategies, capacity analysis, and optimization techniques," in *IEEE Commun. Theory Workshop (CTW)*, 2016.