On Achievability for Downlink Cloud Radio Access Networks with Base Station Cooperation

Chien-Yi Wang, Michèle Wigger, *Senior Member, IEEE*, and
Abdellatif Zaidi, *Senior Member, IEEE*

***Abstract*—This work investigates the downlink of a cloud radio access network (C-RAN) in which a central processor communicates with two mobile users through two base stations (BSs). The BSs act as relay nodes and cooperate with each other through error-free rate-limited links. We develop and analyze two coding schemes for this scenario. The first coding scheme modifies the Liu-Kang scheme (to make it amenable to a rigorous analysis) and extends it to introduce common codewords and to apply for downlink C-RAN with BS-to-BS cooperation. This first coding scheme enables arbitrary correlation among the auxiliary codewords that are recovered by the BSs. We show that this scheme improves over previous schemes for various instances of Gaussian C-RAN channels. In particular, in many scearios, the scheme can better exploit the possibility of BS-to-BS cooperation than other schemes. The second coding scheme extends the distributed decode and forward (DDF) scheme by means of Gray-Wyner compression and by exploiting the cooperation links between BSs. In addition and as a separate extension, we provide an improved capacity approximation for the DDF strategy for the capacity of a general N-BS L-user C-RAN model in the memoryless Gaussian case.**

***Index Terms*—Broadcast relay networks, cloud radio access networks, compression, conferencing relays, data sharing, distributed decode–forward, Gaussian networks.**

I. Introduction

Cloud radio access networks (C-RANs) are promising candidates for fifth generation (5G) wireless communication networks. In a C-RAN, the base stations (BSs) are connected to a central processor through digital fronthaul links. Comprehensive surveys on C-RANs can be found in [1], [2]. The 2-BS 2-user case is depicted in Figure 1. Two common approaches for coding over downlink C-RANs are:

- *Data-sharing*: The central processor splits each message into independent submessages and conveys these independent submessages to one or multiple BSs. The BSs map the received submessages into codewords and transmit these codewords over the interference network. The mobile users decode their intended message parts by treating interference as noise. If there are N BSs, in general there can be up to $2^N - 1$ submessages, each of which is sent to a specific subset of BSs. Two special cases have been considered in the literature: Zakhour and Gesbert [3] studied the 2-BS 2-user case; and Dai and Yu [4] focused on BS clustering for general C-RANs wherein messages are sent as a whole to subsets of BSs and there is no message splitting.
- *Compression of Signals*: The central processor precalculates virtual channel inputs and describes compressed versions thereof over the rate-limited fronthaul links to the BSs. The BSs reconstruct the compressed signals and transmit them over the interference network. Compression schemes were first investigated by Park *et al.* [5] for the memoryless Gaussian case. More recently, Yu [6] showed that this scheme is actually subsumed by Lim et al.'s [7] *distributed decode and forward (DDF)* strategy when this latter is specialized to the C-RAN network.

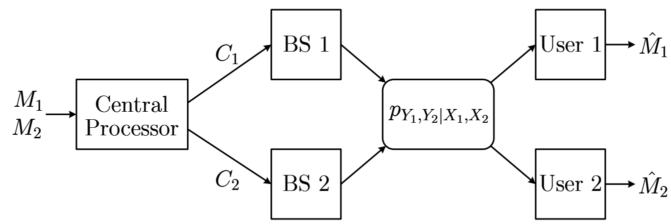


Fig. 1. Downlink C-RAN with 2 base stations and 2 mobile users.

A main distinguishing feature between the two approaches is that in data-sharing the BSs' transmit signals directly correspond to codewords, whereas under the compression approach the BSs send signals that are only compressed versions of such codewords.

Another interesting scheme is *reverse compute–forward* proposed by Hong and Caire [8], which uses nested lattice codes to perform precalculations in a finite field. The reverse compute–forward scheme can enhance the performance under the condition of weak fronthaul links, but it suffers from non-integer penalty and thus is less competitive than the first two approaches when the fronthaul links are strong.

In this paper, we study the downlink of a C-RAN with two BSs and two mobile users. Recently, Liu and Kang [9] proposed a coding scheme generalizing the data-sharing approach. In their scheme, the central processor maps the message pair (M_1, M_2) into "2-dimensional" Marton codewords: codewords U_1^n, U_2^n for message M_1 and codewords V_1^n, V_2^n for message M_2. It then describes codewords U_1^n, V_1^n to BS 1 and codewords U_2^n, V_2^n to BS 2, where the descriptions are obtained by enumerating all possible pairs of codewords (U_1^n, V_1^n) and (U_2^n, V_2^n). We view this scheme as a generalization of the

The work of C.-Y. Wang and M. Wigger has been supported by Huawei Technologies France SASU, under grant agreement YB2015120036.

Parts of the results have been presented at *IEEE WCNC 2018*, in San Francisco (CA), USA.

C.-Y. Wang and M. Wigger are with the Communications and Electronics Department, Telecom ParisTech, Université Paris-Saclay, Paris, France. Emails: {chien-yi.wang, michele.wigger}@telecom-paristech.fr

A. Zaidi is with the Université Paris-Est, Paris, France and the Mathematics and Algorithmic Sciences Lab., Huawei Technologies France, Boulogne-Billancourt, France. Email: abdellatif.zaidi@u-pem.fr

data-sharing approach, because the BSs transmit codewords and not compressed versions therefore. Liu and Kang showed that their scheme in some scenarios improves over all previous schemes [9]; their performance analysis is however flawed due to an erroneous application of the mutual covering lemma (at beginning of p. 1009 when analyzing $\Pr(\xi_0)$), and it is unclear whether the claimed performance is indeed achievable.

In this paper, we slightly modify the code construction of the Liu-Kang scheme and analyze the modified code. Even though a bit worse than the performance conjectured in [10], the modified scheme improves over all previously proposed schemes in some regimes. We also generalize the new scheme by introducing common codewords U_0^n, V_0^n that the central processor enumerates and describes to *both BSs*. Introducing such common codewords is of interest also for the more general scenario of *downlink C-RAN with BS cooperation* where BSs can communicate with each other over dedicated rate-limited cooperation links before communicating over the interference network with the mobile users. In practice such a BS-to-BS communication can take place over the traditional backhaul links that connect BSs. In fact, a prominent line of research advocates that performance of 5G systems can be improved by employing heterogeneous access techonologies [11]–[13]. The reasons include: practical difficulties for installing high-rate fiber-optic fronthaul links, BSs that are closer to each other than to cloud processors, heterogeneous traffic conditions, or economical considerations regarding installation and maintenance of the different access networks.

The more general setup with direct BS-to-BS cooperation is the main focus of this paper.

Finally, using the DDF scheme, this paper characterizes the capacity region of a general N-BS L-user C-RAN model under the memoryless Gaussian model to within a constant gap, independent of power, that is smaller than the general gap proved in [7].

The main contributions of this work can be summarized as follows:

1) We modify the Liu–Kang scheme [9] so as to be able to analyze it, and we introduce common codewords. We use the cooperation links to exchange parts of common codewords and to redirect private codewords for asymmetric link or channel conditions. This new *cooperative generalized data-sharing (G-DS) scheme* subsumes the data-sharing scheme proposed in [3]. (Sometimes when the capacities of the BS-to-BS cooperation links $C_{12} = C_{21} = 0$, we refer to our new scheme simply as G-DS.)
2) We introduce Gray-Wyner coding [14] to DDF for C-RAN and extend the scheme to the scenario with BS-to-BS cooperation. This new *cooperative generalized compression (G-compression) scheme* subsumes the previous compression schemes in [5] and in [6]. (When $C_{12} = C_{21} = 0$, we refer to this scheme also as G-compression scheme.)
3) Under the memoryless Gaussian model, we show that DDF for C-RAN with BS-to-BS cooperation achieves the capacity region of a downlink N-BS L-user C-RAN to within a gap of $\frac{L}{2} + \frac{\min\{N, L\log N\}}{2}$ bits per dimension. This improves the previous gap result in [7], which was $\frac{L+N}{2}$ bits per dimension.
4) We show that without BS-to-BS cooperation, under the memoryless Gaussian model, the G-DS scheme outperforms the G-compression scheme in the low-power regime and when the channel gain matrix is ill-conditioned. Furthermore, the cooperative G-DS scheme benefits more from BS-to-BS cooperation than the cooperative G-compression scheme and often improves over the latter when the cooperation rates are sufficiently large.

The paper is organized as follows. In Section II, we provide the problem formulation for the 2-BS 2-user case. Section III is devoted to the cooperative G-DS scheme, in which we describe the detailed coding scheme and consider three representative special cases and two examples with simpler network topologies. Section IV is devoted to the cooperative compression scheme. In this section, we describe the cooperative G-compression scheme and state the improved gap-result. Finally, in Section V we compare the cooperative G-DS and cooperative G-compression schemes through examples and evaluation for the memoryless Gaussian model. The lengthy proofs are deferred to appendices.

A. Notations

Random variables and their realizations are represented by uppercase letters (e.g., X) and lowercase letters (e.g., x), respectively. Matrices are represented by uppercase letters in sans-serif font (e.g., M) and vectors are in boldface font (e.g., $\mathbf{v}$). We use calligraphic symbols (e.g., $\mathcal{X}$) and the Greek letter Ω to denote sets. The probability distribution of a random variable X is denoted by p_X. Denote by $|\cdot|$ the cardinality of a set and by $\mathbb{1}\{\cdot\}$ the indicator function of an event. We denote $[a] := \{1, 2, \cdots, \lfloor a \rfloor\}$ for all $a \geq 1$, $X^k := (X_1, X_2, \cdots, X_k)$, and $X(\Omega) = (X_i : i \in \Omega)$. Throughout the paper, all logarithms are to the base two.

The usual notation for entropy, $H(X)$, and mutual information, $I(X;Y)$, is used. We follow the ϵ–δ notation in [15] and the robust typicality introduced in [16]: For $X \sim p_X$ and $\epsilon \in (0,1)$, the set of typical sequences of length k with respect to the probability distribution p_X and the parameter ϵ is denoted by $\mathcal{T}_\epsilon^{(k)}(X)$, which is defined as

$$\mathcal{T}_\epsilon^{(k)}(X) := \left\{ x^k \in \mathcal{X}^k : \left| \frac{\#(a|x^k)}{k} - p_X(a) \right| \leq \epsilon p_X(a), \forall a \in \mathcal{X} \right\}, \quad (1)$$

where $\#(a|x^k)$ is the number of occurrences of a in x^k. Finally, the total correlation among the random variables $X(\Omega)$ is defined as

$$\Gamma(X(\Omega)) := \sum_{i\in\Omega} H(X_i) - H(X(\Omega)). \quad (2)$$

II. Problem Statement

Consider the downlink 2-BS 2-user C-RAN with BS cooperation depicted in Figure 2. The network consists of one central processor, two BSs, and two mobile users. The central processor communicates with the two BSs through individual noiseless bit pipes of finite capacities. Denote by C_k the capacity of the link from the central processor to

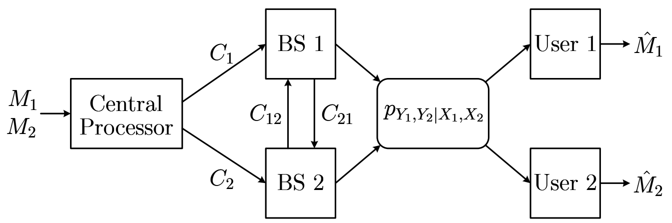


Fig. 2. Downlink C-RAN with BS cooperation: 2 base stations and 2 mobile users.

BS k. In addition, the two BSs can also communicate with each other through individual noiseless bit pipes of finite capacities. Denote by C_{kj} the capacity of the link from BS j to BS k. The network from the BSs to the mobile users is modeled as a discrete memoryless interference channel (DM-IC) $\langle \mathcal{X}_1 \times \mathcal{X}_2, p_{Y_1,Y_2|X_1,X_2}, \mathcal{Y}_1 \times \mathcal{Y}_2 \rangle$ that consists of four finite sets $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_1, \mathcal{Y}_2$ and a collection of conditional probability mass functions (pmf) $p_{Y_1,Y_2|X_1,X_2}$.

With the help of the two BSs, the central processor wants to communicate two messages M_1 and M_2 to users 1 and 2, respectively. Assume that M_1 and M_2 are independent and uniformly distributed over $[2^{nR_1}]$ and $[2^{nR_2}]$, respectively. In this paper, we restrict attention to information processing on a block-by-block basis. Each block consists of a sequence of n symbols. The entire communication is divided into three successive phases:

1) *Central processor to BSs:*
The central processor conveys two indices $(W_1, W_2) := f_0(M_1, M_2)$ to BS 1 and BS 2, respectively, where $f_0 : [2^{nR_1}] \times [2^{nR_2}] \to [2^{nC_1}] \times [2^{nC_2}]$ is the encoder of the central processor.
2) *BS-to-BS conferencing communication:*
BS 1 conveys an index $W_{21} := f_1(W_1)$ to BS 2, where $f_1 : [2^{nC_1}] \to [2^{nC_{21}}]$ is the conferencing encoder of BS 1. BS 2 conveys an index $W_{12} := f_2(W_2)$ to BS 1, where $f_2 : [2^{nC_2}] \to [2^{nC_{12}}]$ is the conferencing encoder of BS 2.
3) *BSs to mobile users:*
BS 1 transmits a sequence $X_1^n := g_1(W_1, W_{12})$ over the DM-IC, where $g_1 : [2^{nC_1}] \times [2^{nC_{12}}] \to \mathcal{X}_1^n$ is the channel encoder of BS 1. BS 2 transmits a sequence $X_2^n := g_2(W_2, W_{21})$ over the DM-IC, where $g_2 : [2^{nC_2}] \times [2^{nC_{21}}] \to \mathcal{X}_2^n$ is the channel encoder of BS 2.

Upon receiving the sequence $Y_\ell^n \in \mathcal{Y}_\ell^n$, mobile user $\ell \in \{1, 2\}$ finds an estimate $\hat{M}_\ell := d_\ell(Y_\ell^n)$ of message M_ℓ, where $d_\ell : \mathcal{Y}_\ell^n \to [2^{nR_\ell}]$ is the decoder of user ℓ. The collection of the encoders f_0, f_1, f_2, g_1, g_2 and the decoders d_1, d_2 constitute a $(2^{nR_1}, 2^{nR_2}, n)$ code.

The average error probability is defined as

$$\mathsf{P}_e^{(n)} := \mathbb{P}\left(\bigcup_{\ell=1}^{2} \{\hat{M}_\ell \neq M_\ell\} \right). \tag{3}$$

A rate pair (R_1, R_2) is said achievable if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes such that $\lim_{n\to\infty} \mathsf{P}_e^{(n)} = 0$. The capacity region is the closure of the set of achievable rate pairs.

Finally, we remark that using the discretization procedure [15, Section 3.4.1] and appropriately introducing input costs, our developed results for DM-ICs can be adapted to the Gaussian interference channel with constrained input power. The input–output relation of this channel is

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}, \tag{4}$$

where $X_k \in \mathbb{R}$ is the channel input from BS k, Y_ℓ is the channel output observed at user ℓ, $g_{\ell k} \in \mathbb{R}$ is the channel gain from BS k to user ℓ, and (Z_1, Z_2) are i.i.d. $\mathcal{N}(0, 1)$ and each BS has to satisfy an average power constraint P, i.e., $\frac{1}{n}\sum_{i=1}^{n} x_{ki}^2 \leq P$ for all $k \in \{1, 2\}$.

III. Cooperative Generalized Data-Sharing Scheme

Before describing our new scheme, it is instructive to briefly review the encoding of the data-sharing scheme.

A. Preliminary: Data-Sharing Scheme

The conventional data-sharing scheme follows from a *rate-splitting* approach: Each message M_ℓ is split into three independent submessages $M_{\ell 0}$, $M_{\ell 1}$, and $M_{\ell 2}$, where $\ell \in \{1, 2\}$. The central processor sends the private messages (M_{1k}, M_{2k}) to BS k, where $k \in \{1, 2\}$, and the common messages (M_{10}, M_{20}) to both BSs. The BSs map the received submessages into codewords, i.e., $m_{1j} \to u_j^n$ and $m_{2j} \to v_j^n$, for all $j \in \{0, 1, 2\}$, and each BS $k \in \{1, 2\}$ applies a symbol-by-symbol mapping $x_k(u_0, v_0, u_k, v_k)$ to map the codewords $(U_0^n, V_0^n, U_k^n, V_k^n)$ into channel inputs X_k^n. In the data-sharing scheme, codewords are generated according to the joint pmf

$$p_{U_0,U_1,U_2,V_0,V_1,V_2} = p_{U_0,V_0} \prod_{j=1}^{2} p_{U_j,V_j|U_0,V_0}. \tag{5}$$

Our aim is to develop a coding scheme that allows to exploit more general joint pmfs $p_{U_0,U_1,U_2,V_0,V_1,V_2}$. To this end, we follow the proposition by Liu and Kang [9]: Each message, instead of being split into three independent parts, is now represented by a set of auxiliary index tuples. A priori, auxiliary indices refer to codewords of independently generated codebooks, but through joint typicality tests they indicate coordinated codewords and thus ensure a more general joint distribution on the set of transmitted codewords than in (5).

B. Performance

First, let us give a high-level summary of the cooperative G-DS scheme. The encoding is based on *multicoding*. We fix a joint pmf $p_{U_0,V_0,U_1,V_1,U_2,V_2}$ and independently generate six codebooks U_j, V_j, $j \in \{0, 1, 2\}$, from the marginals p_{U_j}, p_{V_j}, $j \in \{0, 1, 2\}$, respectively. For $j \in \{0, 1, 2\}$, the codebook U_j contains $2^{nR_{\mathrm{u}j}}$ codewords and the codebook V_j contains $2^{nR_{\mathrm{v}j}}$ codewords. Each message $m_1 \in [2^{nR_1}]$ is associated with a unique bin $\mathcal{B}(m_1)$ of index tuples $(k_0, k_1, k_2) \in [2^{R_{\mathrm{u}0}}] \times [2^{R_{\mathrm{u}1}}] \times [2^{R_{\mathrm{u}2}}]$, which are indices to the codebooks $\mathsf{U}_0, \mathsf{U}_1, \mathsf{U}_2$, respectively. Similarly, each message $m_2 \in [2^{nR_2}]$ is associated with a unique bin $\mathcal{B}(m_2)$ of index tuples $(\ell_0, \ell_1, \ell_2) \in$

$[2^{R_{\mathsf{v}0}}] \times [2^{R_{\mathsf{v}1}}] \times [2^{R_{\mathsf{v}2}}]$, which are indices to the independently generated codebooks $\mathsf{V}_0, \mathsf{V}_1, \mathsf{V}_2$, respectively. Then, given (m_1, m_2), we apply joint typicality encoding to find index tuples $(k_0, k_1, k_2) \in \mathcal{B}(m_1)$ and $(\ell_0, \ell_1, \ell_2) \in \mathcal{B}(m_2)$ such that $(U_0^n(k_0), U_1^n(k_1), U_2^n(k_2), V_0^n(\ell_0), V_1^n(\ell_1), V_2^n(\ell_2))$ are jointly typical.

Remark 1: In addition to including the common auxiliaries U_0 and V_0, which was already mentioned in [9], the main difference of our proposed scheme from the Liu–Kang scheme is that we do not enumerate the jointly typical pairs $(U_1^n(k_1), U_2^n(k_2))$ and $(V_1^n(\ell_1), V_2^n(\ell_2))$, which renders the analysis of the success probability of finding jointly typical tuples $(U_1^n(k_1), U_2^n(k_2), V_1^n(\ell_1), V_2^n(\ell_2))$ difficult. ◊

The next step is to convey $(k_0, \ell_0, k_1, \ell_1)$ to BS 1 and $(k_0, \ell_0, k_2, \ell_2)$ to BS 2. By taking advantage of the following facts, we can reduce the conventional sum rate $R_{\mathsf{u}0} + R_{\mathsf{v}0} + R_{\mathsf{u}j} + R_{\mathsf{v}j}$, $j \in \{1, 2\}$:

1) *Correlated index tuples*
The index tuple to be sent represents certain jointly typical codewords. As long as U_0, V_0, U_j, V_j are not mutually independent, some members of $[2^{nR_{\mathsf{u}0}}] \times [2^{nR_{\mathsf{v}0}}] \times [2^{nR_{\mathsf{u}j}}] \times [2^{nR_{\mathsf{v}j}}]$ will never be used. Thus, instead of sending $(k_0, \ell_0, k_j, \ell_j)$ separately, we can enumerate all jointly typical codewords and simply convey an enumeration index.
2) *Opportunity of exploiting the cooperation links*
In the presence of cooperation links, the BSs do not need to learn all the information directly over the link from the central processor, but can learn part of it over the cooperation link.

Finally, user 1 applies joint typicality decoding to recover (k_0, k_1, k_2) and then the message m_1 can be uniquely identified. Similarly, user 2 applies joint typicality decoding to recover (ℓ_0, ℓ_1, ℓ_2) and then the message m_2 can be uniquely identified.

The achieved rate region of the cooperative G-DS scheme is presented in the following theorem.

Theorem 1: A rate pair (R_1, R_2) is achievable for the downlink 2-BS 2-user C-RAN with BS cooperation if there exist some rates $R_{\mathsf{u}j}, R_{\mathsf{v}j} \geq 0$, $j \in \{0, 1, 2\}$, some joint pmf $p_{U_0,V_0,U_1,V_1,U_2,V_2}$, and some functions $x_k(u_0, v_0, u_k, v_k)$, $k \in \{1, 2\}$, such that for all $\Omega_{\mathsf{u}}, \Omega_{\mathsf{v}} \subseteq \{0, 1, 2\}$ satisfying $|\Omega_{\mathsf{u}}| + |\Omega_{\mathsf{v}}| \geq 2$, the following rate constraints (6)–(11) hold:

$$\mathbb{1}\{|\Omega_{\mathsf{u}}| = 3\}R_1 + \mathbb{1}\{|\Omega_{\mathsf{v}}| = 3\}R_2 < \sum_{i\in\Omega_{\mathsf{u}}} R_{\mathsf{u}i} + \sum_{j\in\Omega_{\mathsf{v}}} R_{\mathsf{v}j} - \Gamma(U(\Omega_{\mathsf{u}}), V(\Omega_{\mathsf{v}})); \quad (6)$$

for all non-empty $\Omega_{\mathsf{u}}, \Omega_{\mathsf{v}} \subseteq \{0, 1, 2\}$,

$$\sum_{i\in\Omega_{\mathsf{u}}} R_{\mathsf{u}i} < I(U(\Omega_{\mathsf{u}}); U(\Omega_{\mathsf{u}}^c), Y_1) + \Gamma(U(\Omega_{\mathsf{u}})), \quad (7)$$

$$\sum_{j\in\Omega_{\mathsf{v}}} R_{\mathsf{v}j} < I(V(\Omega_{\mathsf{v}}); V(\Omega_{\mathsf{v}}^c), Y_2) + \Gamma(V(\Omega_{\mathsf{v}})); \quad (8)$$

and

$$\sum_{i\in\{0,1\}} R_{\mathsf{u}i} + \sum_{j\in\{0,1\}} R_{\mathsf{v}j} < C_1 + C_{12} + \Gamma(U_0, V_0, U_1, V_1), \quad (9)$$

$$\sum_{i\in\{0,2\}} R_{\mathsf{u}i} + \sum_{j\in\{0,2\}} R_{\mathsf{v}j} < C_2 + C_{21} + \Gamma(U_0, V_0, U_2, V_2), \quad (10)$$

$$\sum_{i=0}^{2} R_{\mathsf{u}i} + \sum_{j=0}^{2} R_{\mathsf{v}j} < C_1 + C_2 + \Gamma(U_0, V_0, U_1, V_1) + \Gamma(U_0, V_0, U_2, V_2) - \Gamma(U_0, V_0). \quad (11)$$

Unfortunately, the rate region in Theorem 1 is hard to evaluate. Besides, we find it insightful to learn the effects of different code components. Thus, now we present three corollaries to Theorem 1 where we restrict the correlation structure:

1) Corollary 1: $U_j = V_j = \emptyset$ and $R_{\mathsf{u}j} = R_{\mathsf{v}j} = 0$, $j \in \{1, 2\}$. Each of the two messages M_1 and M_2 is encoded into a common codeword U_0^n and V_0^n, which are then transmitted simultaneously by both BSs. Thus, here the two BSs *fully cooperate* in their transmission to the mobile users.
2) Corollary 2: $p_{U_0,V_0,U_1,V_1,U_2,V_2} = \prod_{j=0}^{2} p_{U_j} p_{V_j}$. This is equivalent to splitting messages M_1, M_2 into three parts, one common and two private parts, and to independently transmitting the two common parts by both BSs and each of the private parts by only one of the BSs.
3) Corollary 3: $U_0 = V_0 = \emptyset$ and $R_{\mathsf{u}0} = R_{\mathsf{v}0} = 0$. In this third corollary there are no cloud center codewords.

In all the corollaries, the auxiliaries $(R_{\mathsf{u}j}, R_{\mathsf{v}j} : j \in \{0, 1, 2\})$ are eliminated through the Fourier–Motzkin elimination.[1] We remark that the first two correlation structures can also be realized through the rate-splitting approach mentioned in Section III-A.

Corollary 1 (Scheme I): A rate pair (R_1, R_2) is achievable for the downlink 2-BS 2-user C-RAN with BS cooperation if

$$R_1 < I(U_0; Y_1), \quad (12)$$
$$R_2 < I(V_0; Y_2), \quad (13)$$
$$R_1 + R_2 < I(U_0; Y_1) + I(V_0; Y_2) - I(U_0; V_0), \quad (14)$$
$$R_1 + R_2 < \min\{C_1 + C_{12}, C_2 + C_{21}, C_1 + C_2\}, \quad (15)$$

for some joint pmf p_{U_0,V_0} and some functions $x_k(u_0, v_0)$, $k \in \{1, 2\}$.

Corollary 2 (Scheme II): A rate pair (R_1, R_2) is achievable for the downlink 2-BS 2-user C-RAN with BS cooperation if

$$R_1 < C_1 + C_{12} + I(U_2; Y_1|U_0, U_1), \quad (16)$$
$$R_1 < C_2 + C_{21} + I(U_1; Y_1|U_0, U_2), \quad (17)$$
$$R_1 < I(U_0, U_1, U_2; Y_1), \quad (18)$$
$$R_2 < C_1 + C_{12} + I(V_2; Y_2|V_0, V_1), \quad (19)$$
$$R_2 < C_2 + C_{21} + I(V_1; Y_2|V_0, V_2), \quad (20)$$
$$R_2 < I(V_0, V_1, V_2; Y_2), \quad (21)$$
$$R_1 + R_2 < C_1 + C_2, \quad (22)$$
$$R_1 + R_2 < C_1 + C_{12} + I(U_2; Y_1|U_0, U_1) + I(V_2; Y_2|V_0, V_1), \quad (23)$$
$$R_1 + R_2 < C_2 + C_{21} + I(U_1; Y_1|U_0, U_2) + I(V_1; Y_2|V_0, V_2), \quad (24)$$
$$R_1 + 2R_2 < C_1 + C_2 + C_{12} + C_{21} + I(V_1, V_2; Y_2|V_0), \quad (25)$$

[1]In this paper, all Fourier–Motzkin eliminations are performed using the software developed by Gattegno, *et al.* [17].

$$2R_1 + R_2 < C_1 + C_2 + C_{12} + C_{21} + I(U_1, U_2; Y_1|U_0), \quad (26)$$

$$2R_1 + 2R_2 < C_1 + C_2 + C_{12} + C_{21} + I(U_1, U_2; Y_1|U_0) + I(V_1, V_2; Y_2|V_0), \quad (27)$$

for some joint pmf $\prod_{j=0}^{2} p_{U_j} p_{V_j}$ and some functions $x_k(u_0, v_0, u_k, v_k)$, $k \in \{1, 2\}$.

When applied to the memoryless Gaussian model (4), Corollary 2 with $C_{12} = C_{21} = 0$ recovers the rate region of the scheme of Zakhour and Gesbert [3, Proposition 1].

Corollary 3 (Scheme III): A rate pair (R_1, R_2) is achievable for the downlink 2-BS 2-user C-RAN with BS cooperation if

$$R_1 < C_1 + C_{12} + I(U_2; U_1, Y_1) - I(U_2; U_1, V_1), \quad (28)$$

$$R_1 < C_2 + C_{21} + I(U_1; U_2, Y_1) - I(U_1; U_2, V_2), \quad (29)$$

$$R_1 < I(U_1, U_2; Y_1) + \min \left\{ \begin{array}{l} 0, \\ I(V_1; V_2, Y_2) - I(V_1; U_1, U_2), \\ I(V_2; V_1, Y_2) - I(V_2; U_1, U_2) \end{array} \right\}, \quad (30)$$

$$R_2 < C_1 + C_{12} + I(V_2; V_1, Y_2) - I(V_2; U_1, V_1), \quad (31)$$

$$R_2 < C_2 + C_{21} + I(V_1; V_2, Y_2) - I(V_1; U_2, V_2), \quad (32)$$

$$R_2 < I(V_1, V_2; Y_2) + \min \left\{ \begin{array}{l} 0, \\ I(U_2; U_1, Y_1) - I(U_2; V_1, V_2), \\ I(U_1; U_2, Y_1) - I(U_1; V_1, V_2) \end{array} \right\}, \quad (33)$$

$$R_1 + R_2 < I(U_1, U_2; Y_1) + I(V_1, V_2; Y_2) - I(U_1, U_2; V_1, V_2), \quad (34)$$

$$R_1 + R_2 < C_1 + C_2 - I(U_1, V_1; U_2, V_2), \quad (35)$$

$$R_1 + R_2 < C_1 + C_{12} - I(U_1, V_1; U_2, V_2) + \min \left\{ \begin{array}{l} I(U_2; U_1, Y_1) + I(V_2; V_1, Y_2) \\ \quad -I(U_2; V_2), \\ 2I(U_2; U_1, Y_1) + I(V_1, V_2; Y_2) \\ \quad -I(U_2; V_1) - I(U_2; V_2) \\ \quad +I(V_1; V_2), \\ I(U_1, U_2; Y_1) + 2I(V_2; V_1, Y_2) \\ \quad -I(U_1; V_2) - I(U_2; V_2) \\ \quad +I(U_1; U_2) \end{array} \right\}, \quad (36)$$

$$R_1 + R_2 < C_2 + C_{21} - I(U_1, V_1; U_2, V_2) + \min \left\{ \begin{array}{l} I(U_1; U_2, Y_1) + I(V_1; V_2, Y_2) \\ \quad -I(U_1; V_1), \\ 2I(U_1; U_2, Y_1) + I(V_1, V_2; Y_2) \\ \quad -I(U_1; V_1) - I(U_1; V_2) \\ \quad +I(V_1; V_2), \\ I(U_1, U_2; Y_1) + 2I(V_1; V_2, Y_2) \\ \quad -I(U_1; V_1) - I(U_2; V_1) \\ \quad +I(U_1; U_2) \end{array} \right\}, \quad (37)$$

for some joint pmf p_{U_1,V_1,U_2,V_2} and some functions $x_k(u_k, v_k)$, $k \in \{1, 2\}$, such that

$$I(U_1; V_1) < I(U_1; U_2, Y_1) + I(V_1; V_2, Y_2), \quad (38)$$

$$I(U_2; V_2) < I(U_2; U_1, Y_1) + I(V_2; V_1, Y_2), \quad (39)$$

$$I(U_1; V_2) < I(U_1; U_2, Y_1) + I(V_2; V_1, Y_2), \quad (40)$$

$$I(U_2; V_1) < I(U_2; U_1, Y_1) + I(V_1; V_2, Y_2). \quad (41)$$

C. Examples

Now let us consider two special cases with simpler topologies.

Example 1 (1 BS and 2 users): The downlink 1-BS 2-user C-RAN can be considered as a special case of the downlink 2-BS 2-user C-RAN with $C_2 = C_{12} = C_{21} = 0$ and $p_{Y_1,Y_2|X_1,X_2} = p_{Y_1,Y_2|X_1}$. We fix a joint pmf $p_{U,V}$ and substitute $(U_1, V_1) = (U, V)$, $U_j = V_j = \emptyset$, and $R_{\mathsf{u}j} = R_{\mathsf{v}j} = 0$, $j \in \{0, 2\}$, in Theorem 1. Then, after removing $R_{\mathsf{u}1}$ and $R_{\mathsf{v}1}$ by the Fourier–Motzkin elimination, we have the following corollary.

Corollary 4: A rate pair (R_1, R_2) is achievable for the downlink 1-BS 2-user C-RAN if there exist some pmf $p_{U,V}$ and some function $x_1(u, v)$ such that

$$R_1 < I(U; Y_1), \quad (42)$$

$$R_2 < I(V; Y_2), \quad (43)$$

$$R_1 + R_2 < I(U; Y_1) + I(V; Y_2) - I(U; V), \quad (44)$$

$$R_1 + R_2 < C_1. \quad (45)$$

Thus, the achieved rate region is essentially Marton's inner bound [18] with the additional constraint (45) due to the fact that the digital link is of finite capacity. ◊

Example 2 (2 BSs and 1 user): The downlink 2-BS 1-user C-RAN is a class of *diamond networks* [10], [19], which can be considered as a special case of the downlink 2-BS 2-user C-RAN by setting $R_2 = 0$. We fix a joint pmf p_{U,X_1,X_2} and substitute $(U_0, U_1, U_2) = (U, X_1, X_2)$, $V_j = \emptyset$, and $R_{\mathsf{v}j} = 0$, $j \in \{0, 1, 2\}$, in Theorem 1. Then, after removing $R_{\mathsf{u}0}$, $R_{\mathsf{u}1}$, and $R_{\mathsf{u}2}$ by the Fourier–Motzkin elimination, we have the following corollary.

Corollary 5: Any rate R_1 is achievable for the downlink 2-BS 1-user C-RAN with BS cooperation if there exists some pmf p_{U,X_1,X_2} such that

$$R_1 < \min \left\{ \begin{array}{l} C_1 + C_2 - I(X_1; X_2|U), \\ C_1 + C_{12} + I(X_2; Y_1|U, X_1), \\ C_2 + C_{21} + I(X_1; Y_1|U, X_2), \\ I(X_1, X_2; Y_1), \\ \frac{1}{2}[C_1 + C_2 + C_{12} + C_{21} \\ \quad +I(X_1, X_2; Y_1|U) - I(X_1; X_2|U)] \end{array} \right\}. \quad (46)$$

Remark 2: Considering diamond networks with an orthogonal broadcast channel, the proposed G-DS scheme recovers the achievability results in [10, Theorem 2] and [19, Theorem 1]. It is shown in [19] that the achievability is optimal when the second hop is the binary-adder multiple-access channel, i.e., $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$, $\mathcal{Y}_1 = \{0, 1, 2\}$, and $Y_1 = X_1 + X_2$. Furthermore, the proposed cooperative G-DS scheme recovers the achievability result in [20, Theorem 2] in which cooperation between relays is also included in the network model. ◊

D. Coding Scheme

Codebook generation: Fix a joint pmf $p_{U_0,V_0,U_1,V_1,U_2,V_2}$ and functions $x_j(u_0, v_0, u_j, v_j)$, $j \in \{1, 2\}$. Randomly and independently generate sequences

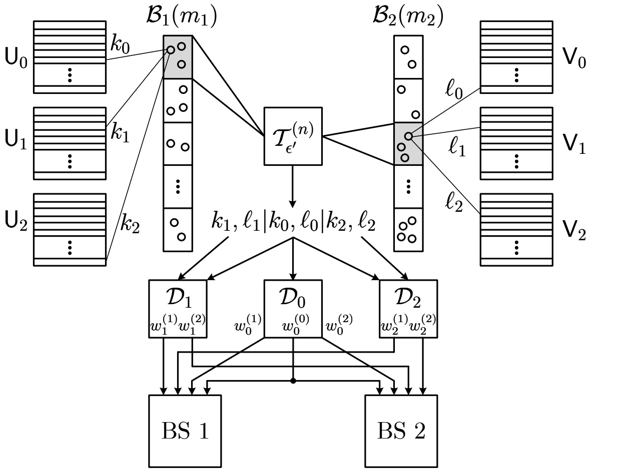


Fig. 3. Illustration of the encoding operation at the central processor in the G-DS scheme.

- $u_0^n(k_0)$, each according to $\prod_{i=1}^n p_{U_0}(u_{0i})$, for $k_0 \in [2^{nR_{\mathsf{u}0}}]$;
- $u_1^n(k_1)$, each according to $\prod_{i=1}^n p_{U_1}(u_{1i})$, for $k_1 \in [2^{nR_{\mathsf{u}1}}]$;
- $u_2^n(k_2)$, each according to $\prod_{i=1}^n p_{U_2}(u_{2i})$, for $k_2 \in [2^{nR_{\mathsf{u}2}}]$;
- $v_0^n(\ell_0)$, each according to $\prod_{i=1}^n p_{V_0}(v_{0i})$, for $\ell_0 \in [2^{nR_{\mathsf{v}0}}]$;
- $v_1^n(\ell_1)$, each according to $\prod_{i=1}^n p_{V_1}(v_{1i})$, for $\ell_1 \in [2^{nR_{\mathsf{v}1}}]$;
- $v_2^n(\ell_2)$, each according to $\prod_{i=1}^n p_{V_2}(v_{2i})$, for $\ell_2 \in [2^{nR_{\mathsf{v}2}}]$.

Next, we generate dictionaries:

$$\mathcal{D}_0 = \{(k_0, \ell_0) \in [2^{nR_{\mathsf{u}0}}] \times [2^{nR_{\mathsf{v}0}}] \colon \\ (u_0^n(k_0), v_0^n(\ell_0)) \in \mathcal{T}_{\epsilon'}^{(n)}\}, \quad (47)$$

and for all $(k_0, \ell_0) \in \mathcal{D}_0$:

$$\mathcal{D}_1(k_0, \ell_0) = \\ \{(k_1, \ell_1) \in [2^{nR_{\mathsf{u}1}}] \times [2^{nR_{\mathsf{v}1}}] \colon \\ (u_1^n(k_1), v_1^n(\ell_1)) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, V_1 | u_0^n(k_0), v_0^n(\ell_0))\}, \quad (48)$$

$$\mathcal{D}_2(k_0, \ell_0) \\ \{(k_2, \ell_2) \in [2^{nR_{\mathsf{u}2}}] \times [2^{nR_{\mathsf{v}2}}] \colon \\ (u_2^n(k_2), v_2^n(\ell_2)) \in \mathcal{T}_{\epsilon'}^{(n)}(U_2, V_2 | u_0^n(k_0), v_0^n(\ell_0))\}. \quad (49)$$

Every index tuple in the dictionaries is assigned a unique reference label by means of the functions

$$\delta_0 \colon \mathcal{D}_0 \to \{1, \ldots, |\mathcal{D}_0|\} \quad (50)$$

and for all $(k_0, \ell_0) \in \mathcal{D}_0$:

$$\delta_1(\cdot | k_0, \ell_0) \colon \mathcal{D}_1(k_0, \ell_0) \to \{1, \ldots, |\mathcal{D}_1(k_0, \ell_0)|\}, \quad (51)$$

$$\delta_2(\cdot | k_0, \ell_0) \colon \mathcal{D}_2(k_0, \ell_0) \to \{1, \ldots, |\mathcal{D}_2(k_0, \ell_0)|\}. \quad (52)$$

Let δ_0^{-1}, $\delta_1^{-1}(\cdot | k_0, \ell_0)$, and $\delta_2^{-1}(\cdot | k_0, \ell_0)$ denote the corresponding inverse maps.

Finally, we randomly and independently assign an index $m_1(k_0, k_1, k_2)$ to each index tuple $(k_0, k_1, k_2) \in [2^{nR_{\mathsf{u}0}}] \times [2^{nR_{\mathsf{u}1}}] \times [2^{nR_{\mathsf{u}2}}]$ according to a uniform pmf over $[2^{nR_1}]$. Similarly, we randomly and independently assign an index $m_2(\ell_0, \ell_1, \ell_2)$ to each index tuple $(\ell_0, \ell_1, \ell_2) \in [2^{nR_{\mathsf{v}0}}] \times [2^{nR_{\mathsf{v}1}}] \times [2^{nR_{\mathsf{v}2}}]$ according to a uniform pmf over $[2^{nR_2}]$. We refer to each subset of index tuples with the same index m_j as a bin $\mathcal{B}_j(m_j)$, $j \in \{1, 2\}$.

Central Processor: Upon seeing (m_1, m_2), the central processor finds $(k_0, k_1, k_2) \in \mathcal{B}_1(m_1)$ and $(\ell_0, \ell_1, \ell_2) \in \mathcal{B}_2(m_2)$ such that

$$(u_0^n(k_0), u_1^n(k_1), u_2^n(k_2), v_0^n(\ell_0), v_1^n(\ell_1), v_2^n(\ell_2)) \in \mathcal{T}_{\epsilon'}^{(n)}. \quad (53)$$

If there is more than one such tuple, choose an arbitrary one among them. If no such tuple exists, choose $(k_0, k_1, k_2, \ell_0, \ell_1, \ell_2) = (1, 1, 1, 1, 1, 1)$. Then, the central processor splits $\delta_0(k_0, \ell_0)$ into three subindices $w_0^{(0)}$, $w_0^{(1)}$, and $w_0^{(2)}$ of rates R_{00}, R_{01}, and R_{02}, respectively. Also, for $j \in \{1, 2\}$, the central processor splits $\delta_j(k_j, \ell_j | k_0, \ell_0)$ into two subindices $w_j^{(1)}$ and $w_j^{(2)}$ of rates R_{j1} and R_{j2}, respectively. Finally, the central processor sends the index tuple $(w_0^{(0)}, w_0^{(1)}, w_1^{(1)}, w_2^{(1)})$ to BS 1 and $(w_0^{(0)}, w_0^{(2)}, w_1^{(2)}, w_2^{(2)})$ to BS 2. The encoding operation at the central processor is illustrated in Figure 3.

Basestations: BS 1 forwards $(w_0^{(1)}, w_2^{(1)})$ to BS 2 over the cooperation link. BS 2 forwards $(w_0^{(2)}, w_1^{(2)})$ to BS 1 over the cooperation link. Both BSs apply the inverse mapping δ_0^{-1} to the received triple $(w_0^{(0)}, w_0^{(1)}, w_0^{(2)})$ to recover the common indices (k_0, ℓ_0):

$$(k_0, \ell_0) = \delta_0^{-1}(w_0^{(0)}, w_0^{(1)}, w_0^{(2)}). \quad (54)$$

Then, BS $j \in \{1, 2\}$, applies the inverse mapping $\delta_j^{-1}(\cdot | k_0, \ell_0)$ to the obtained $(w_j^{(1)}, w_j^{(2)})$ to recover its private indices (k_j, ℓ_j):

$$(k_j, \ell_j) = \delta_j^{-1}(w_j^{(1)}, w_j^{(2)} | k_0, \ell_0). \quad (55)$$

Finally, BS j transmits the symbol $x_{ji}(u_{0i}(k_0), v_{0i}(\ell_0), u_{ji}(k_j), v_{ji}(\ell_j))$ at each time $i \in [n]$.

Mobile users: Let $\epsilon > \epsilon'$. User 1 declares that $\hat{m}_1$ is sent if it is the unique message such that for some $(k_0, k_1, k_2) \in \mathcal{B}_1(\hat{m}_1)$ it holds that

$$(u_0^n(k_0), u_1^n(k_1), u_2^n(k_2), y_1^n) \in \mathcal{T}_{\epsilon}^{(n)}; \quad (56)$$

otherwise it declares an error. User 2 declares that $\hat{m}_2$ is sent if it is the unique message such that for some $(\ell_0, \ell_1, \ell_2) \in \mathcal{B}_2(\hat{m}_2)$ it holds that

$$(v_0^n(\ell_0), v_1^n(\ell_1), v_2^n(\ell_2), y_2^n) \in \mathcal{T}_{\epsilon}^{(n)}; \quad (57)$$

otherwise it declares an error.

Analysis of error probability: Let (M_1, M_2) be the messages and let $(K_0, K_1, K_2, L_0, L_1, L_2)$ be the indices chosen at the encoder. In order to have a lossless transmission over the digital links,

$$R_{00} + R_{01} + R_{11} + R_{21} \leq C_1, \quad (58)$$

$$R_{00} + R_{02} + R_{12} + R_{22} \leq C_2, \quad (59)$$

$$R_{01} + R_{21} \leq C_{21}, \quad (60)$$

$$R_{02} + R_{12} \leq C_{12}. \quad (61)$$

Also, we note that

$$R_{00} + R_{01} + R_{02} = \log |\mathcal{D}_0|, \quad (62)$$

$$R_{11} + R_{12} = \log |\mathcal{D}_1(K_0, L_0)|, \quad (63)$$

$$R_{21} + R_{22} = \log |\mathcal{D}_2(K_0, L_0)|. \quad (64)$$

Thus, after applying Fourier-Motzkin elimination to remove R_{00} and (R_{j1}, R_{j2}), $j \in \{0,1,2\}$, from (58)–(61)

$$\log|\mathcal{D}_0| + \log|\mathcal{D}_1(K_0,L_0)| \le C_1 + C_{12}, \tag{65}$$
$$\log|\mathcal{D}_0| + \log|\mathcal{D}_2(K_0,L_0)| \le C_2 + C_{21}, \tag{66}$$
$$\log|\mathcal{D}_0| + \log|\mathcal{D}_1(K_0,L_0)| + \log|\mathcal{D}_2(K_0,L_0)| \le C_1 + C_2. \tag{67}$$

We denote by $\mathcal{A}$ the intersection of the random events (65), (66), and (67). From Lemma 1 proved in Appendix A, the random event $\mathcal{A}$ happens with high probability as $n \to \infty$ if

$$C_1 + C_{12} \ge R_{\mathsf{u}0} + R_{\mathsf{v}0} - I(U_0;V_0) + R_{\mathsf{u}1} + R_{\mathsf{v}1} - I(U_1;V_1) - I(U_0,V_0;U_1,V_1), \tag{68}$$
$$C_2 + C_{21} \ge R_{\mathsf{u}0} + R_{\mathsf{v}0} - I(U_0;V_0) + R_{\mathsf{u}2} + R_{\mathsf{v}2} - I(U_2;V_2) - I(U_0,V_0;U_2,V_2), \tag{69}$$
$$C_1 + C_2 \ge R_{\mathsf{u}0} + R_{\mathsf{v}0} - I(U_0;V_0) + R_{\mathsf{u}1} + R_{\mathsf{v}1} - I(U_1;V_1) - I(U_0,V_0;U_1,V_1) + R_{\mathsf{u}2} + R_{\mathsf{v}2} - I(U_2;V_2) - I(U_0,V_0;U_2,V_2). \tag{70}$$

Besides the error event $\mathcal{A}^c$, the decoding at User 1 fails if one or more of the following events occur:

$$\mathcal{E}_{\mathsf{s}} = \{(U_0^n(k_0), U_1^n(k_1), U_2^n(k_2), V_0^n(\ell_0), V_1^n(\ell_1), V_2^n(\ell_2)) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } (k_0,k_1,k_2) \in \mathcal{B}_1(M_1), (\ell_0,\ell_1,\ell_2) \in \mathcal{B}_2(M_2)\}, \tag{71}$$
$$\mathcal{E}_{\mathsf{d}0} = \{(U_0^n(K_0), U_1^n(K_1), U_2^n(K_2), Y_1^n) \notin \mathcal{T}_{\epsilon}^{(n)}\}, \tag{72}$$
$$\mathcal{E}_{\mathsf{d}1} = \{(U_0(K_0), U_1^n(k_1), U_2^n(K_2), Y_1^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } k_1 \ne K_1\}, \tag{73}$$
$$\mathcal{E}_{\mathsf{d}2} = \{(U_0(K_0), U_1^n(K_1), U_2^n(k_2), Y_1^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } k_2 \ne K_2\}, \tag{74}$$
$$\mathcal{E}_{\mathsf{d}3} = \{(U_0(K_0), U_1^n(k_1), U_2^n(k_2), Y_1^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } k_1 \ne K_1, k_2 \ne K_2\}, \tag{75}$$
$$\mathcal{E}_{\mathsf{d}4} = \{(U_0(k_0), U_1^n(K_1), U_2^n(K_2), Y_1^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } k_0 \ne K_0\}, \tag{76}$$
$$\mathcal{E}_{\mathsf{d}5} = \{(U_0(k_0), U_1^n(k_1), U_2^n(K_2), Y_1^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } k_0 \ne K_0, k_1 \ne K_1\}, \tag{77}$$
$$\mathcal{E}_{\mathsf{d}6} = \{(U_0(k_0), U_1^n(K_1), U_2^n(k_2), Y_1^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } k_0 \ne K_0, k_2 \ne K_2\}, \tag{78}$$
$$\mathcal{E}_{\mathsf{d}7} = \{(U_0(k_0), U_1^n(k_1), U_2^n(k_2), Y_1^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } k_0 \ne K_0, k_1 \ne K_1, k_2 \ne K_2\}. \tag{79}$$

Thus, the average error probability for M_1 is upper bounded as

$$\mathbb{P}(\{\hat{M}_1 \ne M_1\}) \le \mathbb{P}(\mathcal{E}_{\mathsf{s}}) + \mathbb{P}(\mathcal{A}^c) + \mathbb{P}(\mathcal{E}_{\mathsf{d}0} \cap \mathcal{E}_{\mathsf{s}}^c \cap \mathcal{A}) + \sum_{i=1}^{7} \mathbb{P}(\mathcal{E}_{di}). \tag{80}$$

From Lemma 2 proved in Appendix B, the term $\mathbb{P}(\mathcal{E}_{\mathsf{s}})$ tends to zero as $n \to \infty$ if

$$\sum_{i\in\Omega_{\mathsf{u}}} R_{\mathsf{u}i} + \sum_{j\in\Omega_{\mathsf{v}}} R_{\mathsf{v}j} > \mathbb{1}\{\Omega_{\mathsf{u}} = \{0,1,2\}\}R_1 + \mathbb{1}\{\Omega_{\mathsf{v}} = \{0,1,2\}\}R_2 + \Gamma(U(\Omega_{\mathsf{u}}), V(\Omega_{\mathsf{v}})), \tag{81}$$

for all $\Omega_{\mathsf{u}}, \Omega_{\mathsf{v}} \subseteq \{0,1,2\}$ such that $|\Omega_{\mathsf{u}}| + |\Omega_{\mathsf{v}}| \ge 2$.

Next, due to the codebook construction and the conditional typicality lemma [15, p. 27], $\mathbb{P}(\mathcal{E}_{\mathsf{d}0} \cap \mathcal{E}_{\mathsf{s}}^c \cap \mathcal{A})$ tends to zero as $n \to \infty$. Finally, using the joint typicality lemma [15, p. 29], $\sum_{i=1}^{7} \mathbb{P}(\mathcal{E}_{\mathsf{d}i})$ tends to zero as $n \to \infty$ if

$$R_{\mathsf{u}1} < I(U_1;U_0,U_2,Y_1) - \delta(\epsilon), \tag{82}$$
$$R_{\mathsf{u}2} < I(U_2;U_0,U_1,Y_1) - \delta(\epsilon), \tag{83}$$
$$R_{\mathsf{u}1} + R_{\mathsf{u}2} < I(U_1,U_2;U_0,Y_1) + I(U_1;U_2) - \delta(\epsilon), \tag{84}$$
$$R_{\mathsf{u}0} < I(U_0;U_1,U_2,Y_1) - \delta(\epsilon), \tag{85}$$
$$R_{\mathsf{u}0} + R_{\mathsf{u}1} < I(U_0,U_1;U_2,Y_1) + I(U_0;U_1) - \delta(\epsilon), \tag{86}$$
$$R_{\mathsf{u}0} + R_{\mathsf{u}2} < I(U_0,U_2;U_1,Y_1) + I(U_0;U_2) - \delta(\epsilon), \tag{87}$$
$$R_{\mathsf{u}0} + R_{\mathsf{u}1} + R_{\mathsf{u}2} < I(U_0,U_1,U_2;Y_1) + I(U_0;U_1,U_2) + I(U_1;U_2) - \delta(\epsilon). \tag{88}$$

The average error probability for M_2 can be bounded in a similar manner and then we have the additional rate conditions

$$R_{\mathsf{v}1} < I(V_1;V_0,V_2,Y_2) - \delta(\epsilon), \tag{89}$$
$$R_{\mathsf{v}2} < I(V_2;V_0,V_1,Y_2) - \delta(\epsilon), \tag{90}$$
$$R_{\mathsf{v}1} + R_{\mathsf{v}2} < I(V_1,V_2;V_0,Y_2) + I(V_1;V_2) - \delta(\epsilon), \tag{91}$$
$$R_{\mathsf{v}0} < I(V_0;V_1,V_2,Y_2) - \delta(\epsilon), \tag{92}$$
$$R_{\mathsf{v}0} + R_{\mathsf{v}1} < I(V_0,V_1;V_2,Y_2) + I(V_0;V_1) - \delta(\epsilon), \tag{93}$$
$$R_{\mathsf{v}0} + R_{\mathsf{v}2} < I(V_0,V_2;V_1,Y_2) + I(V_0;V_2) - \delta(\epsilon), \tag{94}$$
$$R_{\mathsf{v}0} + R_{\mathsf{v}1} + R_{\mathsf{u}2} < I(V_0,V_1,V_2;Y_2) + I(V_0;V_1,V_2) + I(V_1;V_2) - \delta(\epsilon). \tag{95}$$

Finally, the theorem is established by letting ϵ tend to zero.

IV. Cooperative Generalized Compression Scheme

This section is devoted to compression-based schemes, where our contributions are as follows:

- For the downlink 2-BS 2-user C-RAN, we introduce Gray-Wyner coding [14] to the DDF scheme [7] for C-RAN [6] and extend the resulting scheme to BS-to-BS cooperation. Also, we derive a single-letter rate region for general discrete memoryless channels on the second hop.
- We show that under the memoryless Gaussian model, the original DDF scheme for C-RAN with BS-to-BS cooperation achieves within a constant gap (independent

of transmission power) from the capacity region. This constant gap is smaller than the constant gap provided in [7] for general relay networks.

A. Performance and Coding Scheme

We start with a high-level summary of the proposed cooperative compression scheme. The encoding is based on superposition coding and multicoding. Each message m_j, $j \in \{1,2\}$, is associated with a set of independently generated codewords $U_j^n(m_j, \ell_j)$ of size $2^{n\tilde{R}_j}$. Then, we introduce Gray-Wyner source coding [14]. That means, we generate three codebooks $\mathsf{X}_0, \mathsf{X}_1, \mathsf{X}_2$ using superposition coding: the codebook X_0 contains the cloud centers $X_0^n(k_0)$ and the codebooks X_1 and X_2 contain the satellite codewords $X_1^n(k_1|k_0)$ and $X_2^n(k_2|k_0)$, respectively.

Given (m_1, m_2), we apply joint typicality encoding to find an index tuple $(k_0, k_1, k_2, \ell_1, \ell_2)$ such that $(U_1^n(m_1,\ell_1), U_2^n(m_2,\ell_2), X_0^n(k_0), X_1^n(k_1|k_0), X_2^n(k_2|k_0))$ are jointly typical. In words, we jointly apply Marton channel coding on the messages m_1 and m_2 and Gray-Wyner source coding on the resulting codewords.

The next step is to convey (k_0, k_1) to BS 1 and (k_0, k_2) to BS 2, during which the cooperation links are used to reduce the workload of the digital links from the central processor to the BSs. Finally, each user $j \in \{1,2\}$ applies joint typicality decoding to recover the auxiliary $U_j^n(m_j, \ell_j)$ and thus can recover the desired message m_j.

Theorem 2: A rate pair (R_1, R_2) is achievable for the downlink 2-BS 2-user C-RAN with BS cooperation if

$$R_1 < I(U_1;Y_1) + \min\left\{\begin{array}{l} 0, \\ C_1 + C_{12} - I(U_1; X_0, X_1), \\ C_2 + C_{21} - I(U_1; X_0, X_2) \end{array}\right\}, \quad (96)$$

$$R_2 < I(U_2;Y_2) + \min\left\{\begin{array}{l} 0, \\ C_1 + C_{12} - I(U_2; X_0, X_1), \\ C_2 + C_{21} - I(U_2; X_0, X_2) \end{array}\right\}, \quad (97)$$

$$R_1 + R_2 < I(U_1;Y_1) + I(U_2;Y_2) - I(U_1;U_2) + \min\left\{\begin{array}{l} 0, \\ C_1 + C_{12} - I(U_1, U_2; X_0, X_1), \\ C_2 + C_{21} - I(U_1, U_2; X_0, X_2), \\ C_1 + C_2 - I(U_1, U_2; X_0, X_1, X_2) \\ -I(X_1; X_2|X_0) \end{array}\right\}, \quad (98)$$

$$\begin{aligned} 2R_1 + R_2 < {} & I(U_1;Y_1) + I(U_2;Y_2) - I(U_1;U_2) \\ & + C_1 + C_2 + C_{12} + C_{21} \\ & - I(U_1, U_2; X_0, X_1, X_2) - I(X_1; X_2|X_0), \\ & + I(U_1;Y_1) - I(U_1; X_0), \end{aligned} \quad (99)$$

$$\begin{aligned} R_1 + 2R_2 < {} & I(U_1;Y_1) + I(U_2;Y_2) - I(U_1;U_2) \\ & + C_1 + C_2 + C_{12} + C_{21} \\ & - I(U_1, U_2; X_0, X_1, X_2) - I(X_1; X_2|X_0), \\ & + I(U_2;Y_2) - I(U_2; X_0), \end{aligned} \quad (100)$$

$$\begin{aligned} 2R_1 + 2R_2 < {} & I(U_1;Y_1) + I(U_2;Y_2) - I(U_1;U_2) \\ & + C_1 + C_2 + C_{12} + C_{21} \\ & - I(U_1, U_2; X_0, X_1, X_2) - I(X_1; X_2|X_0) \\ & + I(U_1;Y_1) + I(U_2;Y_2) \\ & - I(U_1;U_2) - I(U_1, U_2; X_0), \end{aligned} \quad (101)$$

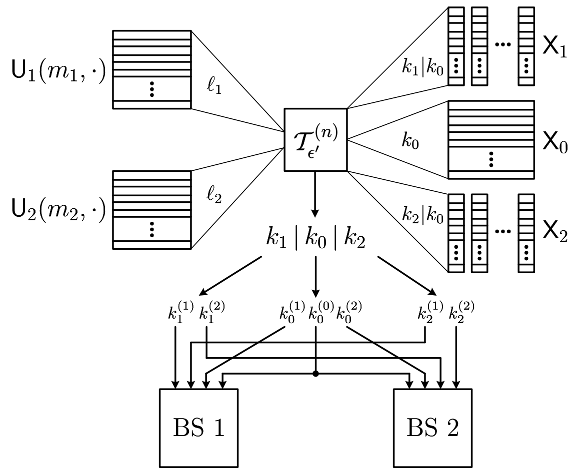


Fig. 4. Illustration of the encoding operation at the central processor in the G-Compression scheme.

for some joint pmf p_{U_1,U_2,X_0,X_1,X_2}.

Proof: We describe and analyze a coding scheme achieving the desired rate region.

Codebook generation: Fix a joint pmf p_{U_1,U_2,X_0,X_1,X_2}. For $j \in \{1,2\}$, randomly and independently generate sequences $u_j^n(m_j,\ell_j)$, according to $\prod_{i=1}^n p_{U_j}(u_{ji})$, for $(m_j,\ell_j) \in [2^{nR_j}] \times [2^{n\tilde{R}_j}]$. Randomly and independently generate sequences $x_0^n(k_0)$, according to $\prod_{i=1}^n p_{X_0}(x_{0i})$, for $k_0 \in [2^{nR'_0}]$. Finally, for $j \in \{1,2\}$, randomly and independently generate sequences $x_j^n(k_j|k_0)$, each according to $\prod_{i=1}^n p_{X_j|X_0}(x_{ji}|x_{0i}(k_0))$, for $(k_0,k_j) \in [2^{nR'_0}] \times [2^{nR'_j}]$.

Central processor: Upon seeing (m_1, m_2), the central processor finds an index tuple $(k_0, k_1, k_2, \ell_1, \ell_2)$ such that

$$\begin{aligned} &(u_1^n(m_1,\ell_1), u_2^n(m_2,\ell_2), \\ &x_0^n(k_0), x_1^n(k_1|k_0), x_2^n(k_2|k_0)) \in \mathcal{T}_{\epsilon'}^{(n)}. \end{aligned} \quad (102)$$

If there is more than one such tuple, choose an arbitrary one among them. If no such tuple exists, choose $(k_0, k_1, k_2, \ell_1, \ell_2) = (1,1,1,1,1)$. Then, the central processor splits k_0 into three subindices $k_0^{(0)}$, $k_0^{(1)}$, and $k_0^{(2)}$ of rates R'_{00}, R'_{01}, and R'_{02}, respectively. Also, for $j \in \{1,2\}$, the central processor splits k_j into two subindices $k_j^{(1)}$ and $k_j^{(2)}$ of rates R'_{j1} and R'_{j2}, respectively. Finally, the central processor sends the index tuple $(k_0^{(0)}, k_0^{(1)}, k_1^{(1)}, k_2^{(1)})$ to BS 1 and $(k_0^{(0)}, k_0^{(2)}, k_1^{(2)}, k_2^{(2)})$ to BS 2. The encoding operation at the central processor is illustrated in Figure 4.

Base stations: BS 1 forwards $(k_0^{(1)}, k_2^{(1)})$ to BS 2 over the cooperation link. BS 2 forwards $(k_0^{(2)}, k_1^{(2)})$ to BS 1 over the cooperation link. Thus, BS $j \in \{1,2\}$ learns the value of (k_0, k_j) and transmits $x_j^n(k_j|k_0)$.

Mobile users: Let $\epsilon > \epsilon'$. For $j \in \{1,2\}$, upon seeing y_j^n, user j finds the unique pair $(\hat{m}_j, \hat{\ell}_j)$ such that

$(u_j^n(\hat{m}_j, \hat{\ell}_j), y_j^n) \in \mathcal{T}_\epsilon^{(n)}$ and declares that $\hat{m}_j$ is sent; otherwise it declares an error.

Analysis of Error Probability: Let (M_1, M_2) be the messages and let $(K_0, K_1, K_2, L_1, L_2)$ be the indices chosen at the central processor. In order to have a lossless transmission over the digital links, it requires that

$$R'_{00} + R'_{01} + R'_{11} + R'_{21} \leq C_1, \tag{103}$$
$$R'_{00} + R'_{02} + R'_{12} + R'_{22} \leq C_2, \tag{104}$$
$$R'_{01} + R'_{21} \leq C_{21}, \tag{105}$$
$$R'_{02} + R'_{12} \leq C_{12}. \tag{106}$$

Note that $R'_{00} + R'_{01} + R'_{02} = R'_0$ and $R'_{j1} + R'_{j2} = R'_j$, $j \in \{1, 2\}$.

Assuming the above conditions are satisfied, the decoding at user 1 fails if one or more of the following events occur:

$$\begin{aligned} \mathcal{E}_0 = \{&(U_1^n(M_1, \ell_1), U_2^n(M_2, \ell_2), \\ &X_0^n(k_0), X_1^n(k_1|k_0), X_2^n(k_2|k_0)) \notin \mathcal{T}_{\epsilon'}^{(n)} \\ &\text{for all } (k_0, k_1, k_2, \ell_1, \ell_2)\}, \end{aligned} \tag{107}$$
$$\mathcal{E}_1 = \{(U_1^n(M_1, L_1), Y_1^n) \notin \mathcal{T}_\epsilon^{(n)}\}, \tag{108}$$
$$\mathcal{E}_2 = \{(U_1^n(m_1, \ell_1), Y_1^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } (m_1, \ell_1) \neq (M_1, L_1)\}. \tag{109}$$

The average error probability for M_1 is upper bounded as

$$\mathbb{P}(\{\hat{M}_1 \neq M_1\}) \leq \mathbb{P}(\mathcal{E}_0) + \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_0^c) + \mathbb{P}(\mathcal{E}_2). \tag{110}$$

By extending [15, Lemma 14.1, p. 351], it can be shown that the term $\mathbb{P}(\mathcal{E}_0)$ tends to zero as $n \to \infty$ if $\tilde{R}_1 + \tilde{R}_2 > I(U_1; U_2) + \delta(\epsilon')$ and

$$\begin{aligned} R'_0 &+ \sum_{k \in \mathcal{S}} R'_k + \sum_{j \in \mathcal{D}} \tilde{R}_j \\ &> I(U(\mathcal{D}); X_0, X(\mathcal{S})) + \mathbb{1}\{\mathcal{S} = \{1,2\}\} I(X_1; X_2|X_0) \\ &+ \mathbb{1}\{\mathcal{D} = \{1,2\}\} I(U_1; U_2) + \delta(\epsilon'), \end{aligned} \tag{111}$$

for all $\mathcal{D}, \mathcal{S} \subseteq \{1, 2\}$. Next, due to the codebook construction and the conditional typicality lemma [15, p. 27], $\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_0^c)$ tends to zero as $n \to \infty$. Finally, using the joint typicality lemma [15, p. 29], $\mathbb{P}(\mathcal{E}_2)$ tends to zero as $n \to \infty$ if

$$R_1 + \tilde{R}_1 < I(U_1; Y_1) - \delta(\epsilon). \tag{112}$$

The average error probability for M_2 can be bounded in a similar manner and then we have the additional rate condition

$$R_2 + \tilde{R}_2 < I(U_2; Y_2) - \delta(\epsilon). \tag{113}$$

Using the Fourier–Motzkin elimination to project out $\tilde{R}_1$, $\tilde{R}_2$, and R'_{0j}, R'_j, $j \in \{0, 1, 2\}$, we obtain the rate conditions in Theorem 2. Finally, the theorem is established by letting $\epsilon \to 0$. ∎

B. Improved Constant-Gap Result

Since downlink C-RAN is a special instance of memoryless broadcast relay networks, the DDF scheme achieves any point in the capacity region of an N-BS L-user C-RAN to within a gap of $(1+N+L)/2$ bits per dimension under the memoryless Gaussian model [7, Corollary 8]. The following theorem tightens this gap for downlink C-RANs. The proof is deferred to Appendix C.

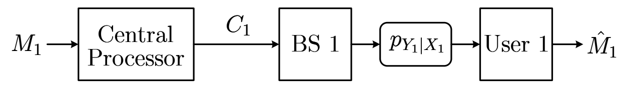


Fig. 5. The system considered in Example 3.

Theorem 3: Consider the downlink of any N-BS L-user C-RAN with BS cooperation. Under the memoryless Gaussian model, the DDF scheme for broadcast achieves within $\frac{L}{2} + \frac{\min\{N, L \log N\}}{2}$ bits per dimension from the capacity region.

A tighter gap might be obtained by considering our extension of DDF that includes also cloud center codewords.

V. Comparison and Numerical Evaluations

In this section, we evaluate and compare our cooperative G-DS scheme and our cooperative T-compression scheme at hand of some examples. We start with examples without BS-to-BS cooperation. In the first example, the G-DS scheme (as well as the data-sharing scheme) is optimal, whereas the G-compression scheme is strictly suboptimal. In the second example the opposite is true. In the second part of this section, we provide numerical results for the memoryless Gaussian model.

A. Examples

Example 3 (One BS and One User): Consider the special case with only one BS and one user, as depicted in Figure 5. (Our model reduces to this scenario when the DM-IC is of the form $p_{Y_1,Y_2|X_1,X_2} = p_{Y_1,Y_2|X_1}$ and when $C_2 = R_2 = 0$.) Decode-and-forward [21] is optimal in this special case and rate R_1 is achievable whenever

$$R_1 < \min\left\{C_1, \max_{p_{X_1}} I(X_1; Y_1)\right\}. \tag{114}$$

Furthermore, compress-and-forward [21] is also optimal since the first hop is noiseless. To see this, notice that for this simple setup the rate achieved with compress-and-forward is $R_1 \leq I(\tilde{X}_1; \hat{Y}_1|X_1)$ where $\tilde{X}_1^n$ describes the bits sent over the fronthaul link and where the pmf $p_{\tilde{X}_1 X_2 \hat{Y}_1}$ has to decompose as $p_{\tilde{X}_1} p_{X_1} p_{\hat{Y}_1|\tilde{X}_1 X_1}$ and satisfy $C_1 \leq H(\tilde{X}_1)$ and $I(X_1; Y_1) \geq I(\tilde{X}_1; \hat{Y}_1|X_1)$. Picking $\hat{Y}_1 = \tilde{X}_1$ uniform over $\min\{C_1, \max_{p_{X_1}} I(X_1; Y_1)\}$ independent of X_1 and picking X_1 to achieve capacity over the channel from the BS to the mobile user, establishes the desired achievability result. This performance is also recovered by the G-DS scheme; see Corollary 4 specialized to $R_2 = 0$ and the choice of auxiliaries $V = \emptyset$ and $X_1 = U$.

The compression scheme and the DDF scheme for broadcast achieve any rate R_1 that satisfies:

$$R_1 < I(U_1; Y_1), \tag{115}$$
$$\begin{aligned} R_1 &< C_1 + I(U_1; Y_1) - I(U_1; X_1) \\ &= C_1 - I(U_1; X_1|Y_1), \end{aligned} \tag{116}$$

for some pmf p_{U_1,X_1} s.t. $U_1 \multimap\!\!- X_1 \multimap\!\!- Y_1$ form a Markov chain.

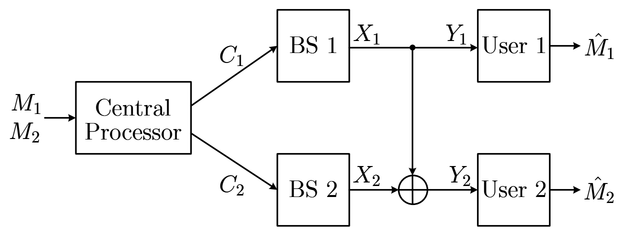


Fig. 6. The system considered in Example 4.

If the second hop is deterministic, i.e., Y_1 is a deterministic function of X_1, then the compression scheme with $U_1 = Y_1$ achieves capacity. However, if the second hop is not deterministic, then setting $U_1 = Y_1$ violates the Markov condition $U_1 -\!\!\circ\!\!- X_1 -\!\!\circ\!\!- Y_1$. In general, compression schemes are suboptimal. To see this, consider a discrete memoryless channel satisfying $p_{Y_1|X_1}(y_1|x_1) < 1$ for all $(x_1, y_1) \in \mathcal{X}_1 \times \mathcal{Y}_1$, i.e., for all inputs $x_1 \in \mathcal{X}_1$, the output Y_1 is not a deterministic function of x_1. Then, for every pmf p_{X_1}, the corresponding joint pmf p_{X_1,Y_1} is indecomposable.[2] Next, let us assume that $0 < C_1 < \max_{p_{X_1}} I(X_1; Y_1)$. Now we show that the compression scheme is not capacity achieving by contradiction.

If the compression scheme is capacity achieving, then it holds that the capacity-achieving distribution p_{U_1,X_1} satisfies that $I(U_1; X_1|Y_1) = 0$, i.e., $U_1 -\!\!\circ\!\!- Y_1 -\!\!\circ\!\!- X_1$ form a Markov chain. However, since $U_1 -\!\!\circ\!\!- X_1 -\!\!\circ\!\!- Y_1$ also form a Markov chain, the indecomposability of the joint pmf p_{X_1,Y_1} implies that the capacity-achieving distribution p_{U_1,X_1} satisfies that U_1 is independent of (X_1, Y_1) (see [22, Problem 16.25, p. 392]) and thus $I(U_1; Y_1) = 0$, which contradicts that the joint pmf p_{U_1,X_1} achieves the capacity $C_1 > 0$. ♢

Example 4 (Z-Interference Channel): Consider the setup without BS-to-BS cooperation where $C_1 = C_2 = 1$, $C_{12} = C_{21} = 0$, $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$, $Y_1 = X_1$, and $Y_2 = X_1 \oplus X_2$. The system is depicted in Figure 6. Now we show that the rate pair $(R_1, R_2) = (1, 1)$, which is on the boundary of the capacity region, is achievable by the G-compression scheme but not by the G-DS scheme.

The following scheme achieves the desired rate pair $(R_1, R_2) = (1, 1)$. Fix a blocklength n and denote by $B_\ell^n := (B_{\ell,1}, \dots, B_{\ell,n})$ the n-bits representation of M_ℓ, $\ell \in \{1, 2\}$. The central processor sends all bits B_1^n to BS 1, and it sends the x-or bits $B_\oplus^n := (B_{1,1} \oplus B_{2,1}, \dots, B_{1,n} \oplus B_{2,n})$ to BS 2. BS 1 sends inputs $X_1^n = B_1^n$ over the DM-IC and BS 2 sends inputs $X_2^n = B_\oplus^n$.

The same performance is achieved by the compression scheme when the auxiliaries (U_1, U_2) are chosen i.i.d. Bernoulli(1/2), and $X_0 = \emptyset$, $X_1 = U_1$, and $X_2 = U_1 \oplus U_2$.

Now let us investigate the G-DS scheme. We consider the following relaxed conditions, where the inequalities do not need to be strict:

$$\begin{aligned} R_1 + R_2 &\overset{(a)}{\le} I(U_0, U_1, U_2; Y_1) + I(V_0, V_1, V_2; Y_2) \\ &\quad - I(U_0, U_1, U_2; V_0, V_1, V_2), \qquad (117) \\ 2R_1 + R_2 &\overset{(b)}{\le} C_1 + C_2 + C_{12} + C_{21} + I(U_1, U_2; Y_1|U_0) \\ &\quad - I(U_1, V_1; U_2, V_2|U_0, V_0), \qquad (118) \end{aligned}$$

where (a) follows by combining (6), (7), and (8) with $\Omega_\mathsf{u} = \Omega_\mathsf{v} = \{0, 1, 2\}$ and (b) follows by combining two times of (6) with $(\Omega_\mathsf{u}, \Omega_\mathsf{v}) = (\{0, 1, 2\}, \{0, 1, 2\})$ and $(\Omega_\mathsf{u}, \Omega_\mathsf{v}) = (\{0, 1, 2\}, \emptyset)$, (7) with $\Omega_\mathsf{u} = \{1, 2\}$, (9), and (10).

If $(R_1, R_2) = (1, 1)$ is achievable by the G-DS scheme, then there must exist a joint pmf $p_{U_0,V_0,U_1,V_1,U_2,V_2}$ and functions $x_k(u_0, v_0, u_k, v_k)$, $k \in \{1, 2\}$, such that

1) $I(U_0, U_1, U_2; V_0, V_1, V_2) = 0$;
2) $I(U_1, V_1; U_2, V_2|U_0, V_0) = 0$;
3) $I(U_0, U_1, U_2; Y_1) = 1$;
4) $I(V_0, V_1, V_2; Y_2) = 1$;
5) $I(U_1, U_2; Y_1|U_0) = 1$.

However, the above constraints cannot be satisfied simultaneously. To see this, let us assume that the first four conditions hold, which imply that

1) (U_0, U_1, U_2) is independent of (V_0, V_1, V_2);
2) the Markov chains $U_1 -\!\!\circ\!\!- U_0 -\!\!\circ\!\!- U_2$ and $V_1 -\!\!\circ\!\!- V_0 -\!\!\circ\!\!- V_2$ hold; and
3) $H(X_1|U_0, U_1, U_2) = H(X_1 \oplus X_2|V_0, V_1, V_2) = 0$.

Thus,

$$\begin{aligned} I(X_1; V_0, V_1|U_0, U_1) &= I(X_1, U_0, U_1; V_0, V_1) \qquad (119) \\ &\le I(X_1, U_0, U_1, U_2; V_0, V_1) \qquad (120) \\ &\overset{(a)}{=} I(U_0, U_1, U_2; V_0, V_1) = 0, \qquad (121) \end{aligned}$$

where (a) follows since $H(X_1|U_0, U_1, U_2) = 0$. Since X_1 is a function of (U_0, V_0, U_1, V_1) by construction, we have $H(X_1|U_0, U_1) = H(X_1|U_0, V_0, U_1, V_1) = 0$, i.e., X_1 is a function of (U_0, U_1). Finally, it holds that

$$\begin{aligned} 0 &= H(X_1 \oplus X_2|V_0, V_1, V_2) \qquad (122) \\ &\ge H(X_1 \oplus X_2|U_0, U_2, V_0, V_1, V_2) \qquad (123) \\ &= H(X_1|U_0, U_2, V_0, V_1, V_2) \qquad (124) \\ &\overset{(a)}{=} H(X_1|U_0), \qquad (125) \end{aligned}$$

where (a) follows since X_1 is a function of (U_0, U_1); (U_0, U_1, U_2) is independent of (V_0, V_1, V_2); and $U_1 -\!\!\circ\!\!- U_0 -\!\!\circ\!\!- U_2$ form a Markov chain. From all above we obtain that constraint 5) cannot be satisfied since Y_1 is a function of U_0, which concludes that the G-DS scheme cannot achieve the rate pair $(1, 1)$. ♢

B. Numerical Evaluation for the Memoryless Gaussian Model

In this subsection, we compare the achieved sum rates of the various coding schemes under the memoryless Gaussian model. For simplicity, we consider the symmetric case, i.e.,

$$\begin{aligned} C_1 &= C_2 = C \\ C_{12} &= C_{21} = C_{\text{coop}} \\ g_{11} &= g_{22} = 1 \\ |g_{12}| &= |g_{21}|. \end{aligned}$$

[2] A joint pmf $p_{X,Y}$ is said to be *indecomposable* [22, Problem 15.12, p. 345] if there are no functions f and g with respective domains $\mathcal{X}$ and $\mathcal{Y}$ so that 1) $\mathbb{P}(f(X) = g(Y)) = 1$ and 2) $f(X)$ takes at least two values with non-zero probability.

Then, the achievable sum rate $R_1 + R_2$ can be upper bounded using the cut-set bound as

$$R_1 + R_2 < \min\{2C, R^\star_{\mathsf{sum}}\}, \tag{126}$$

where $R^\star_{\mathsf{sum}}$ denotes the optimal sum rate assuming $C = \infty$, which can be computed by evaluating the corresponding Gaussian MIMO broadcast channel. We will use the cut-set bound (126) as a reference for comparison. To date, no better converse bound is known, except when there is only a single mobile user [19].

We first consider scenarios without BS-to-BS cooperation, i.e., $C_{\mathsf{coop}} = 0$. Here, we are mainly interested in regimes where the G-DS scheme outperforms the G-compression scheme and the reverse compute–forward. Evaluating the G-DS scheme directly is challenging, so we evaluate the special cases with restricted correlation structures and then apply time sharing on them. To summarize, we evaluate the following schemes

1) G-DS scheme I, II, and III (Corollaries 1, 2, and 3),
2) G-Compression scheme (Theorem 2), and
3) reverse compute–forward with power allocation [8].

Now let us specify our (sub-optimal) choice of auxiliary random variables for the various schemes.

Let $\mathbf{S}^{(k)}$ be a 2×1 jointly Gaussian random vector with zero-mean entries and covariance matrix $\mathsf{K}^{(k)}$, for $k \in \{1, 2\}$. We assume that $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ are independent. For notational convenience, we denote $\mathbf{g}_2 = \begin{bmatrix} g_{21} & g_{22} \end{bmatrix}$.

1) G-DS Scheme I: $U_0 = \mathbf{S}^{(1)}$, $V_0 = \mathbf{S}^{(2)} + \mathsf{A}\mathbf{S}^{(1)}$, and $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \mathbf{S}^{(1)} + \mathbf{S}^{(2)}$, where

$$\mathsf{A} = \mathsf{K}^{(2)} \mathbf{g}_2^T \left(1 + \mathbf{g}_2 \mathsf{K}^{(2)} \mathbf{g}_2^T\right)^{-1} \mathbf{g}_2. \tag{127}$$

Note that $X_k = U_k + V_k$, $k \in \{1, 2\}$. We optimize over the covariance matrices $\mathsf{K}^{(1)}$ and $\mathsf{K}^{(2)}$ that satisfy the average power constraints.

2) G-DS Scheme II: The random variables $(U_0, V_0, U_1, V_1, U_2, V_2)$ are i.i.d. $\mathcal{N}(0, 1)$ and $X_1 = a_1 U_0 + a_2 V_0 + a_3 U_1 + a_4 V_1$ and $X_2 = b_1 U_0 + b_2 V_0 + b_3 U_2 + b_4 V_2$ for some $a_j, b_j \in \mathbb{R}$, $j \in \{1, 2, 3, 4\}$. We optimize over the coefficients $(a_j, b_j : j \in \{1, 2, 3, 4\})$ that satisfy the average power constraints.

3) G-DS Scheme III:

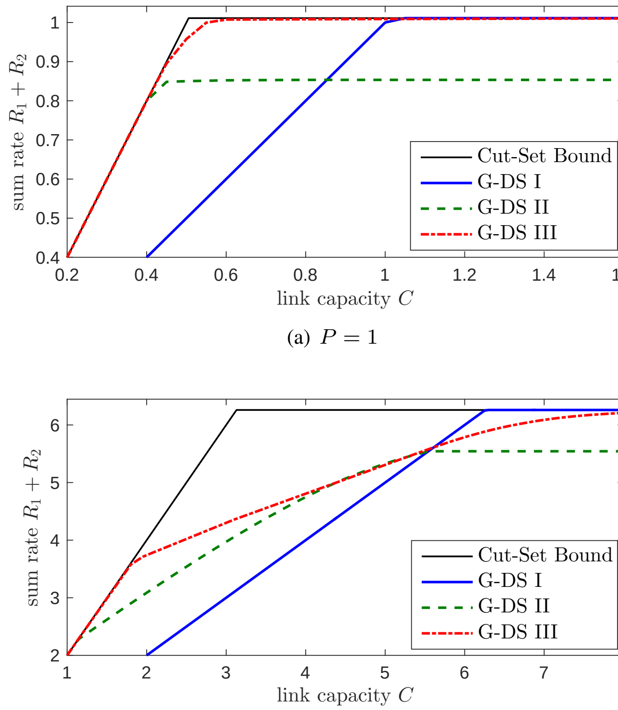
$$\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} = \mathbf{S}^{(1)}, \tag{128}$$

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \mathbf{S}^{(2)} + \mathsf{B}\mathbf{S}^{(1)}, \tag{129}$$

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = (\mathsf{I} + \mathsf{B})\mathbf{S}^{(1)} + \mathbf{S}^{(2)}, \tag{130}$$

where I is the 2×2 identity matrix and B is a 2×2 real-valued matrix. Note that $X_k = U_k + V_k$, $k \in \{1, 2\}$.[3]

[3]We remark that since the BSs do not have full information about $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$, setting $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \mathbf{S}^{(1)} + \mathbf{S}^{(2)}$ is not allowed because the resulting X_k is not a function of (U_k, V_k), $k \in \{1, 2\}$.



(a) $P = 1$

(b) $P = 100$

Fig. 7. Achieved sum-rates of the G-DS schemes I, II, and III under the symmetric memoryless Gaussian model. Here $C_{\mathsf{coop}} = 0$ and $\left[\begin{smallmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right]$.

We optimize over the covariance matrices $\mathsf{K}^{(1)}$, $\mathsf{K}^{(2)}$ and the precoding matrix B that satisfy the average power constraints.

4) G-Compression: $U_1 = \mathbf{S}^{(1)}$, $U_2 = \mathbf{S}^{(2)} + \mathsf{C}\mathbf{S}^{(1)}$, and $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \mathbf{S}^{(1)} + \mathbf{S}^{(2)} + \mathbf{W}$, where

$$\mathsf{C} = \mathsf{K}^{(2)} \mathbf{g}_2^T \left(1 + \mathbf{g}_2 (\mathsf{K}^{(2)} + \mathsf{K}^{(\mathsf{w})}) \mathbf{g}_2^T\right)^{-1} \mathbf{g}_2, \tag{131}$$

and $\mathbf{W}$ is a 2×1 jointly Gaussian random vector with zero-mean entries and covariance matrix $\mathsf{K}^{(\mathsf{w})}$, independent of $(\mathbf{S}^{(1)}, \mathbf{S}^{(2)})$. Finally, we let X_0 be an $\mathcal{N}(0, 1)$ random variable such that X_0 and $(\mathbf{S}^{(1)}, \mathbf{S}^{(2)}, \mathbf{W})$ are jointly Gaussian. We optimize over the covariance matrices $\mathsf{K}^{(1)}$, $\mathsf{K}^{(2)}$, and $\mathsf{K}^{(\mathsf{w})}$ that satisfy the average power constraints and over the covariances of X_0 with each of $(\mathbf{S}^{(1)}, \mathbf{S}^{(2)}, \mathbf{W})$.

We first compare the G-DS schemes I–III where we numerically optimize over the parameters $\mathsf{B}, \mathsf{K}^{(1)}, \mathsf{K}^{(2)}$, $a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4$. In Figures 7 and 8, we fix $C_{\mathsf{coop}} = 0$ and $g_{12} = 0.5$ and consider $(P, g_{21}) \in \{1, 100\} \times \{0.5, -0.5\}$. From the evaluation results, we make the following observations and remarks for the considered setup:

- In general, the G-DS scheme I using only common codewords performs well in the strong-fronthaul regime, i.e., when C is large. By contrast, the G-DS scheme III using only private codewords performs well in the weak-fronthaul regime.

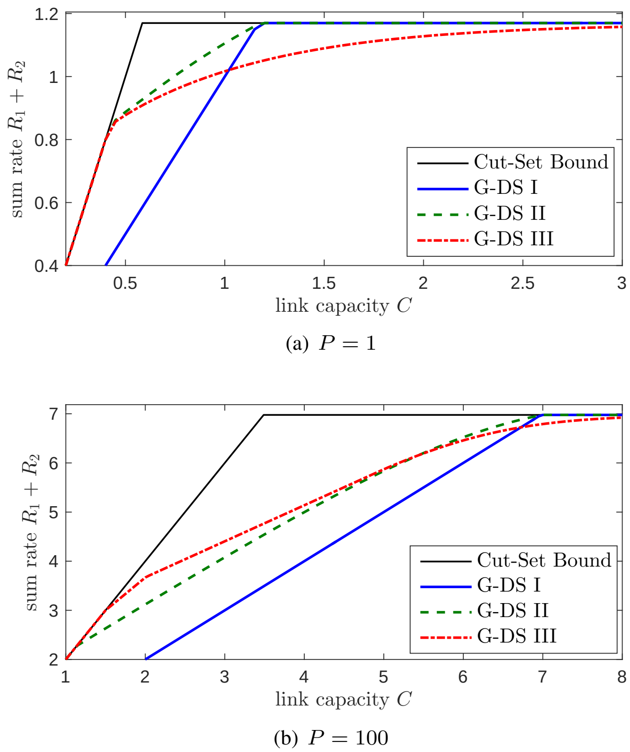


Fig. 8. Achieved sum-rates of the G-DS schemes I, II, and III under the symmetric memoryless Gaussian model. Here $C_{\text{coop}} = 0$ and $\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ -0.5 & 1 \end{bmatrix}$.

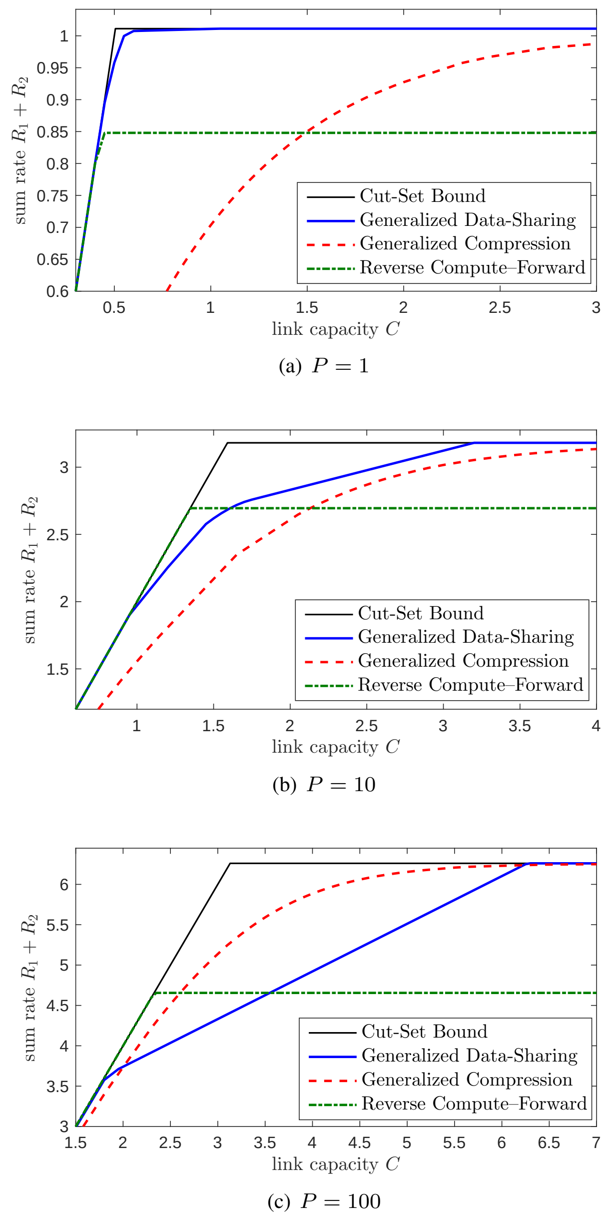


Fig. 9. Achieved sum-rates of the G-DS scheme, the G-compression scheme, and the reverse compute–forward scheme with power control under the symmetric memoryless Gaussian model. Here $C_{\text{coop}} = 0$, and $\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$.

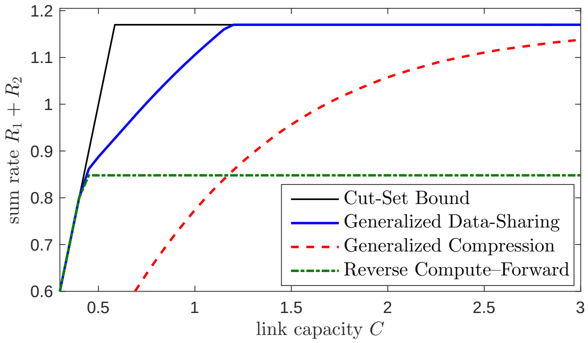
- Introducing correlation among codewords is useful. In fact, time sharing between the G-DS schemes I and III outperforms the G-DS scheme II for all values of link capacity C.
- The G-DS scheme III is more beneficial in the low-power regime, i.e., when P is small.
- For the channel matrix in Figures 8(a) and 8(b), the G-DS scheme II outperforms the G-DS schemes I and III for certain regimes of link capacity C.

We remark that the achieved sum rate of the G-DS scheme I can be simply expressed as $\min\{C + C_{\text{coop}}, 2C, R^{\star}_{\text{sum}}\}$. Thus, when $C_{\text{coop}} = 0$, the G-DS scheme I is optimal for the regime where $C \geq R^{\star}_{\text{sum}}$. It is however unclear whether including common codewords strictly improves the performance of the DS scheme. The presented numerical results cannot answer this question, because the complexity of exhaustively searching over all possible choices of the auxiliary random variables seems too large.

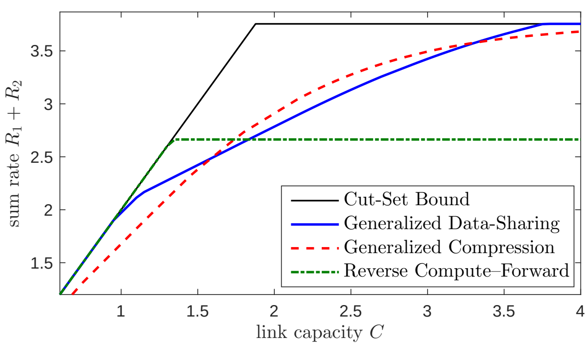
Next, we compare the G-DS scheme (time sharing among the G-DS schemes I, II, and III) with the G-compression scheme and the reverse compute–forward scheme. In Figures 9 and 10, we fix $g_{12} = 0.5$ and consider $(P, g_{21}) \in \{1, 10, 100\} \times \{0.5, -0.5\}$. From the evaluation results, we make the following observations and remarks for the considered setup:

- The G-DS scheme achieves the optimal sum rate when the link capacity C is relatively small or relatively large. The range of optimality depends on the power and the channel conditions. In general, in the low-power regime and/or when the channel gain matrix is ill-conditioned, the G-DS scheme has a more apparent advantage over the other two schemes.
- The G-compression scheme achieves a better performance in the high-power regime. As P increases, the G-compression scheme outperforms the other two schemes in the middle range of link capacity.
- The reverse compute–forward has a good performance when the link capacity C is relatively small, especially when P is large. However, the reverse compute–forward suffers from non-integer penalty and thus its achieved sum rate cannot reach $R^{\star}_{\text{sum}}$ even if the link capacity C is large.

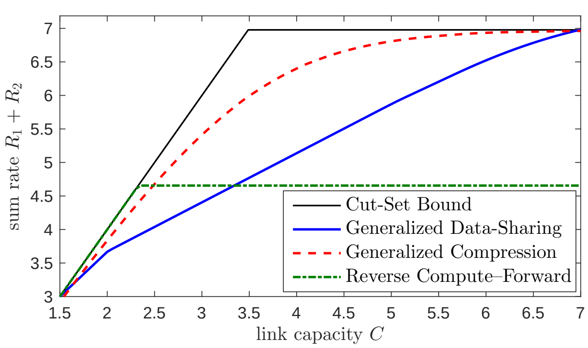
Finally, we consider BS cooperation, i.e., the case where



(a) $P = 1$



(b) $P = 10$



(c) $P = 100$

Fig. 10. Achieved sum-rates of the G-DS scheme, the G-compression scheme, and the reverse compute–forward scheme with power control under the symmetric memoryless Gaussian model. Here $C_{\text{coop}} = 0$, and $\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ -0.5 & 1 \end{bmatrix}$.

$C_{\text{coop}} > 0$.[4] Figure 11 plots the achieved sum rates for the case of $(P, g_{12}, g_{21}) = (100, 0.5, -0.5)$. It turns out that for the symmetric case, only the cooperative G-DS scheme can benefit from the cooperation links. In particular, as the link capacity C_{coop} increases to two, the cooperative G-DS scheme already outperforms the cooperative G-compression scheme for all values of C. Recall that the G-DS scheme I achieves the sum rate $\min\{C + C_{\text{coop}}, 2C, R^\star_{\textsf{sum}}\}$. Since the cut-set bound is $\min\{2C, R^\star_{\textsf{sum}}\}$, we see that increasing C_{coop} is beneficial when $R_1 + R_2 < C + C_{\text{coop}}$ is the dominating constraint. By contrast, for the symmetric case the cooperative

[4]We note that the reverse compute–forward has not been extended for the scenario with BS cooperation. We only include it here as a reference.

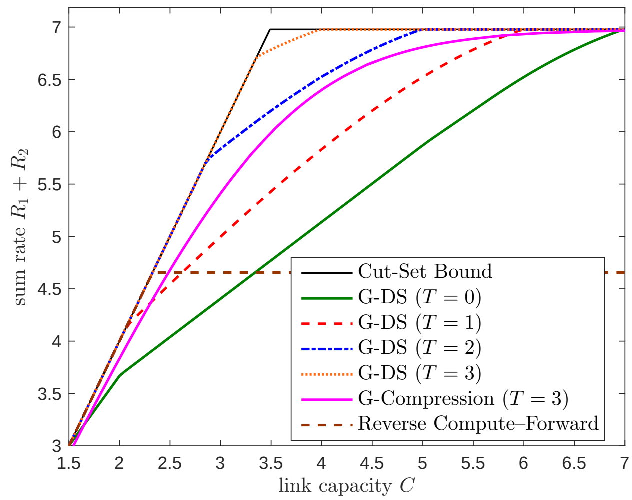


Fig. 11. Achieved sum-rates of the G-DS scheme, the G-Compression scheme, and the reverse compute–forward scheme with power control under the symmetric memoryless Gaussian model. Here $P = 100$ and $\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ -0.5 & 1 \end{bmatrix}$.

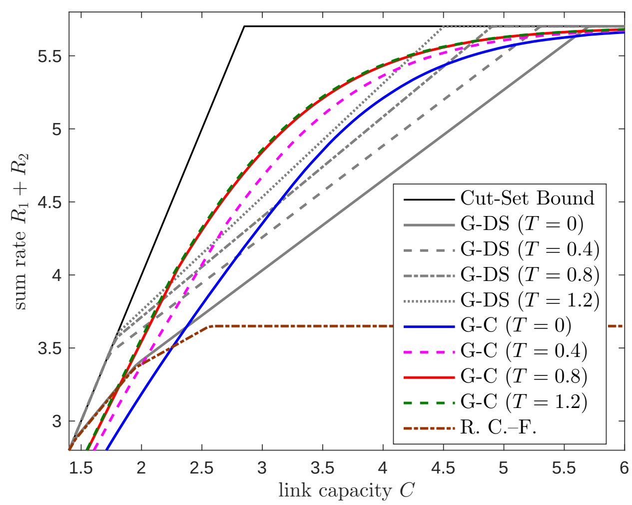


Fig. 12. Achieved sum-rates of the G-DS scheme, the G-Compression (G-C) scheme, and the reverse compute–forward scheme with power control (R. C.-F.) under the memoryless Gaussian model. Here $P = 100$ and $\begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0.25 \\ 1 & -0.25 \end{bmatrix}$.

G-compression scheme cannot benefit from the cooperation links because the dominating rate constraints do not involve C_{12} and C_{21}:

$$R_1 + R_2 < I(U_1; Y_1) + I(U_2; Y_2) - I(U_1; U_2) + \min\left\{ \begin{array}{l} 0\ ,\ C_1 + C_2 - I(X_1; X_2|X_0) \\ \qquad -I(U_1, U_2; X_0, X_1, X_2) \end{array} \right\}, \tag{132}$$

which can be rewritten as

$$R_1 + R_2 < I(U_1; Y_1) + I(U_2; Y_2) - I(U_1; U_2), \tag{133}$$

$$R_1 + R_2 < C_1 + C_2 - I(U_1; X_0, X_1, X_2|Y_1) - I(U_2; U_1, X_0, X_1, X_2|Y_2) - I(X_1; X_2|X_0). \tag{134}$$

If the channel gain matrix is asymmetric, the G-Compression scheme can benefit from the cooperation links, but the gain eventually saturates as C_{coop} increases, again due to the dominating constraint (132). Figure 12 plots the achieved sum rates for the case of $P = 100$ and $\left[\begin{smallmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1 & 0.25 \\ 1 & -0.25 \end{smallmatrix}\right]$. As can been seen, as C_{coop} increases from 0.8 to 1.2, there is little improvement for the cooperative G-compression scheme. By contrast, the G-DS scheme keeps benefiting from the cooperation links before coinciding with the cut-set bound, especially when the fronthaul link capacity C is large.

VI. Concluding Remarks and Future Directions

This paper presents new coding schemes for downlink C-RAN without and with BS-to-BS cooperation. Of particular interest is the (cooperative) G-DS scheme, which improves over all previous schemes for various channels and regimes. Moreover, this scheme can well exploit the possibility of BS-to-BS cooperation links. In the proposed schemes, the BS-to-BS cooperation links are used for rerouting information, which can be used to overcome bottlenecks in asymmetric fronthaul configurations or to free up resources on the fronthaul liks.

The proposed cooperative G-DS scheme introduces common codewords, for simplicity however without superposition coding. Extending the analysis in this paper to a superposition code with common cloud center codewords remains an open challenge. A similar challenge remains open also for the cooperative G-compression scheme which applies Gray-Wyner compression (i.e., superposition coding for compression) but not superposition coding for Marton's code.

Finally, again for simplicity, in our numerical simulations of our G-DS and G-compression schemes, we restricted to special cases and specific choices of the pmfs. Considering larger sets of pmfs of course will lead to better results and more accurate comparisons.

Acknowledgments

The authors are indebted to the associate editor and the anonymous reviewers for their valuable comments which helped improving this manuscript considerably.

Appendix A
Expected Size of Independently Generated Codebooks

The following lemma is a simple extension of [15, Problem 3.8, p. 73] (see also [23]).

Lemma 1: Let $(U, V, W) \sim p_{U,V,W}$. Let W^n be generated according to $\prod_{i=1}^n p_W(w_i)$. Consider two independently generated codebooks $\mathcal{C}_1 = \{U^n(1), \cdots, U^n(2^{nR_1})\}$ and $\mathcal{C}_2 = \{V^n(1), \cdots, V^n(2^{nR_2})\}$. The codewords of $\mathcal{C}_1$ are generated independently each according to $\prod_{i=1}^n p_U(u_i)$. The codewords of $\mathcal{C}_2$ are generated independently each according to $\prod_{i=1}^n p_V(v_i)$. Define the set

$$\mathcal{C} = \{(u^n, v^n) \in \mathcal{C}_1 \times \mathcal{C}_2 : (u^n, v^n, W^n) \in \mathcal{T}_\epsilon^{(n)}(U, V, W)\}. \tag{135}$$

Then, there exists $\delta(\epsilon) > 0$ that tends to zero as $\epsilon \to 0$ such that

$$\mathbb{E}[|\mathcal{C}|] \le 2^{n(R_1+R_2-I(U;V)-I(U,V;W)+\delta(\epsilon))}. \tag{136}$$

Proof:

$$\begin{aligned}
&\mathbb{E}[|\mathcal{C}|] \\
&= \sum_{m=1}^{2^{nR_1}} \sum_{\ell=1}^{2^{nR_2}} \mathbb{P}((U^n(m), V^n(\ell), W^n) \in \mathcal{T}_\epsilon^{(n)}) &(137)\\
&= 2^{n(R_1+R_2)} \mathbb{P}((U^n(1), V^n(1), W^n) \in \mathcal{T}_\epsilon^{(n)}) &(138)\\
&= 2^{n(R_1+R_2)} \sum_{(u^n,v^n,w^n)\in\mathcal{T}_\epsilon^{(n)}(U,V,W)} p_{U^n}(u^n) p_{V^n}(v^n) p_{W^n}(w^n) &(139)\\
&\le 2^{n(R_1+R_2)} \sum_{(u^n,v^n,w^n)\in\mathcal{T}_\epsilon^{(n)}(U,V,W)} 2^{-n(H(U)+H(V)-2\delta(\epsilon))} p_{W^n}(w^n) &(140)\\
&= 2^{n(R_1+R_2)} \sum_{w^n\in\mathcal{T}_\epsilon^{(n)}} p_{W^n}(w^n) |T_\epsilon^{(n)}(U, V|w^n)| \cdot 2^{-n(H(U)+H(V)-2\delta(\epsilon))} &(141)\\
&\le 2^{n(R_1+R_2)} 2^{n(H(U,V|W)+\delta(\epsilon))} 2^{-n(H(U)+H(V)-2\delta(\epsilon))} &(142)\\
&= 2^{n(R_1+R_2-I(U;V)-I(U,V;W)+3\delta(\epsilon))}. &(143)
\end{aligned}$$

■

Appendix B
Multivariate Covering Lemma with Non-Cartesian Product Sets

Lemma 2: Let $(U_0, U_1, U_2, V_0, V_1, V_2) \sim p_{U_0,U_1,U_2,V_0,V_1,V_2}$. For $j \in \{0, 1, 2\}$, randomly and independently generate sequences $U_j^n(k_j)$, $k_j \in [2^{nR_{\mathsf{u}j}}]$, each according to $\prod_{i=1}^n p_{U_j}(u_{ji})$. For $j \in \{0, 1, 2\}$, randomly and independently generate sequences $V_j^n(\ell_j)$, $\ell_j \in [2^{nR_{\mathsf{v}j}}]$, each according to $\prod_{i=1}^n p_{V_j}(v_{ji})$. Randomly and independently assign an index $m_1(k_0, k_1, k_2)$ to each index tuple $(k_0, k_1, k_2) \in [2^{nR_{\mathsf{u}0}}] \times [2^{nR_{\mathsf{u}1}}] \times [2^{nR_{\mathsf{u}2}}]$ according to a uniform pmf over $[2^{nR_1}]$. Randomly and independently assign an index $m_2(\ell_0, \ell_1, \ell_2)$ to each index tuple $(\ell_0, \ell_1, \ell_2) \in [2^{nR_{\mathsf{v}0}}] \times [2^{nR_{\mathsf{v}1}}] \times [2^{nR_{\mathsf{v}2}}]$ according to a uniform pmf over $[2^{nR_2}]$. Define for each sixtuple $(k_0, k_1, k_2, \ell_0, \ell_1, \ell_2)$

$$\begin{aligned}
&\tilde{\mathcal{E}}(k_0, k_1, k_2, \ell_0, \ell_1, \ell_2) \\
&= (U_0^n(k_0), U_1^n(k_1), U_2^n(k_2), V_0^n(\ell_0), V_1^n(\ell_1), V_2^n(\ell_2)) \notin \mathcal{T}_\epsilon^{(n)},
\end{aligned} \tag{144}$$

and for each pair $(m_1, m_2) \in [2^{nR_1}] \times [2^{nR_2}]$ the event

$$\mathcal{E}(m_1, m_2) = \bigcap_{\substack{(k_0,k_1,k_2)\in\mathcal{B}_1(m_1)\\(\ell_0,\ell_1,\ell_2)\in\mathcal{B}_2(m_2)}} \tilde{\mathcal{E}}(k_0, k_1, k_2, \ell_0, \ell_1, \ell_2). \tag{145}$$

Then, for each (m_1, m_2), there exists $\delta(\epsilon)$ that tends to zero as $\epsilon \to 0$ such that $\lim_{n\to\infty} \mathbb{P}(\mathcal{E}(m_1, m_2)) = 0$, if

$$\begin{aligned}
&\sum_{i\in\Omega_{\mathsf{u}}} R_{\mathsf{u}i} + \sum_{j\in\Omega_{\mathsf{v}}} R_{\mathsf{v}j} \\
&\quad > \mathbb{1}\{\Omega_{\mathsf{u}} = \{0, 1, 2\}\} R_1 + \mathbb{1}\{\Omega_{\mathsf{v}} = \{0, 1, 2\}\} R_2 \\
&\quad\quad + \Gamma(U(\Omega_{\mathsf{u}}), V(\Omega_{\mathsf{v}})),
\end{aligned} \tag{146}$$

for all $\Omega_{\mathsf{u}}, \Omega_{\mathsf{v}} \subseteq \{0,1,2\}$ such that $|\Omega_{\mathsf{u}}| + |\Omega_{\mathsf{v}}| \geq 2$.

Proof: The proof follows similar steps as the proof of the multivariate covering lemma. The only difference is that now the set of index tuples is not the usual Cartesian product. By symmetry, it suffices to investigate the case $(m_1, m_2) = (1,1)$. For notational convenience, hereafter we denote $\mathcal{B}_j(1) = \mathcal{B}_j$, $j \in \{1,2\}$.

Let

$$\begin{aligned}\mathcal{A} = \{ & (k_0,k_1,k_2,\ell_0,\ell_1,\ell_2) : (U_0^n(k_0), U_1^n(k_1), U_2^n(k_2), \\ & V_0^n(\ell_0), V_1^n(\ell_1), V_2^n(\ell_2)) \in \mathcal{T}_\epsilon^{(n)}, \\ & (k_0,k_1,k_2) \in \mathcal{B}_1, (\ell_0,\ell_1,\ell_2) \in \mathcal{B}_2 \}. \end{aligned} \tag{147}$$

Then, we have

$$\begin{aligned}\mathbb{P}(\mathcal{E}(1,1)) &= \mathbb{P}(|\mathcal{A}| = 0) \\ &\leq \mathbb{P}\left((|\mathcal{A}| - \mathbb{E}[|\mathcal{A}|])^2 \geq \mathbb{E}[|\mathcal{A}|]^2 \right) \\ &\overset{(a)}{\leq} \frac{\mathrm{Var}(|\mathcal{A}|)}{\mathbb{E}[|\mathcal{A}|]^2} \end{aligned} \tag{148}$$

where (a) follows from Chebyshev's inequality. For convenience, denote

$$\begin{aligned}\phi(k_0,k_1,k_2,\ell_0,\ell_1,\ell_2) = & \\ \mathbb{1}\{ & (U_0^n(k_0), U_1^n(k_1), U_2^n(k_2), \\ & V_0^n(\ell_0), V_1^n(\ell_1), V_2^n(\ell_2)) \in \mathcal{T}_\epsilon^{(n)} \}. \end{aligned} \tag{149}$$

Then, the set size $|\mathcal{A}|$ conditioned on the random bin assignments $\mathcal{B}_1$ and $\mathcal{B}_2$ can be expressed as

$$\mathbb{E}[|\mathcal{A}||\mathcal{B}_1,\mathcal{B}_2] = \sum_{(k_1,k_2)\in\mathcal{B}_1} \sum_{(\ell_1,\ell_2)\in\mathcal{B}_2} \phi(k_0,k_1,k_2,\ell_0,\ell_1,\ell_2). \tag{150}$$

For $a_0,a_1,a_2,b_0,b_1,b_2 \in \{1,2\}$, let

$$\begin{aligned} & p(a_0,a_1,a_2,b_0,b_1,b_2) && (151)\\ & = \mathbb{E}[\phi(1,1,1,1,1,1)\phi(a_0,a_1,a_2,b_0,b_1,b_2)], && (152)\\ & Q(a_0,a_1,a_2,b_0,b_1,b_2) && (153)\\ & = \big|\big\{ (k_0,k_1,k_2,\ell_0,\ell_1,\ell_2,k_0',k_1',k_2',\ell_0',\ell_1',\ell_2') : \\ & \quad (k_0,k_1,k_2) \in \mathcal{B}_1, (\ell_0,\ell_1,\ell_2) \in \mathcal{B}_2, (k_0',k_1',k_2') \in \mathcal{B}_1, \\ & \quad (\ell_0',\ell_1',\ell_2') \in \mathcal{B}_2, \mathcal{F}_0^{(a_0)}, \mathcal{F}_1^{(a_1)}, \mathcal{F}_2^{(a_2)}, \mathcal{G}_0^{(b_0)}, \mathcal{G}_1^{(b_1)}, \mathcal{G}_2^{(b_2)} \big\}\big|, && (154) \end{aligned}$$

where $\mathcal{F}_j^{(1)} = \left(\mathcal{F}_j^{(2)}\right)^c = \{k_j = k_j'\}$ and $\mathcal{G}_j^{(1)} = \left(\mathcal{G}_j^{(2)}\right)^c = \{\ell_j = \ell_j'\}$, for $j \in \{0,1,2\}$. Then, we have

$$\begin{aligned} & \mathbb{E}[|\mathcal{A}||\mathcal{B}_1,\mathcal{B}_2] && (155)\\ & = \sum_{(k_0,k_1,k_2)\in\mathcal{B}_1} \sum_{(\ell_0,\ell_1,\ell_2)\in\mathcal{B}_2} \mathbb{E}[\phi(k_0,k_1,k_2,\ell_0,\ell_1,\ell_2)] && (156)\\ & = Q(1,1,1,1,1,1)p(1,1,1,1,1,1), && (157) \end{aligned}$$

and

$$\begin{aligned} & \mathbb{E}[|\mathcal{A}|^2|\mathcal{B}_1,\mathcal{B}_2] && (158)\\ & = \sum_{(k_0,k_1,k_2)\in\mathcal{B}_1} \sum_{(\ell_0,\ell_1,\ell_2)\in\mathcal{B}_2} \sum_{(k_0',k_1',k_2')\in\mathcal{B}_1} \sum_{(\ell_0',\ell_1',\ell_2')\in\mathcal{B}_2} \\ & \qquad \mathbb{E}[\phi(k_0,k_1,k_2,\ell_0,\ell_1,\ell_2)\phi(k_0',k_1',k_2',\ell_0',\ell_1',\ell_2')] && (159)\\ & = \sum_{a_0,a_1,a_2,b_0,b_1,b_2} Q(a_0,a_1,a_2,b_0,b_1,b_2)p(a_0,a_1,a_2,b_0,b_1,b_2). && (160) \end{aligned}$$

Hence Equality (161) on top of the next page holds. Denote $I = \Gamma(U_0,U_1,U_2,V_0,V_1,V_2)$. By the joint typicality lemma [15, p. 29], it holds that

$$p(1,1,1,1,1,1) \geq 2^{-n(I+\delta(\epsilon))}, \tag{162}$$

$$\begin{aligned} p(a_0,a_1,a_2,b_0,b_1,b_2) \leq\ & 2^{-n(I+\sum_{i\in\Omega_{\mathsf{u}}^c} H(U_i)+\sum_{j\in\Omega_{\mathsf{v}}^c} H(V_j))} \\ & \cdot 2^{n(H(U(\Omega_{\mathsf{u}}^c),V(\Omega_{\mathsf{v}}^c)|U(\Omega_{\mathsf{u}}),V(\Omega_{\mathsf{v}}))+\delta(\epsilon))}, \end{aligned} \tag{163}$$

where $\Omega_{\mathsf{u}} = \bigcup_{j=0}^2 \kappa_j(a_j)$, $\Omega_{\mathsf{v}} = \bigcup_{j=0}^2 \kappa_j(b_j)$, and

$$\kappa_j(x) = \begin{cases} \{j\} & \text{if } x = 1, \\ \emptyset & \text{otherwise.} \end{cases} \tag{164}$$

Also, for all $a_0,a_1,a_2,b_0,b_1,b_2 \in \{1,2\}$, we have

$$\begin{aligned} & \mathbb{E}[Q(a_0,a_1,a_2,b_0,b_1,b_2)] \\ & = 2^{n\left(\sum_{i=0}^2 a_i R_{\mathsf{u}i} + \sum_{j=0}^2 b_j R_{\mathsf{v}j}\right)} \\ & \quad \cdot 2^{-n\left(1+\mathbb{1}\left\{\bigcup_{i=0}^2 \{a_i=2\}\right\}\right)R_1} \cdot 2^{-n\left(1+\mathbb{1}\left\{\bigcup_{j=0}^2 \{b_j=2\}\right\}\right)R_2}. \end{aligned} \tag{165}$$

Finally, (161) and thus (148) can be further upper bounded using (162), (163), (165). It can be checked that the corresponding upper bound tends to zero as $n \to \infty$ if the condition (146) holds, which establishes the lemma. ■

Appendix C
Proof of Theorem 3

First, we state the cut-set bound for the capacity region of the memoryless Gaussian C-RAN model. The proof follows by applying the standard cut-set argument (see [24, Theorem 15.10.1]) to the considered model and then specializing it to the memoryless Gaussian case.

Proposition 1: If a rate tuple $(R_1,\cdots,R_L)$ is achievable for the downlink N-BS L-user C-RAN with BS cooperation, then it must satisfy the inequality

$$\begin{aligned}\sum_{\ell\in\mathcal{D}} R_\ell \leq & \sum_{k\in\mathcal{S}^c} C_k + \sum_{j\in\mathcal{S}} \sum_{k\in\mathcal{S}^c} C_{kj} \\ & + \frac{1}{2}\log\det\left(\mathsf{I} + \mathsf{G}(\mathcal{D},\mathcal{S})\mathsf{K}(\mathcal{S}|\mathcal{S}^c)\mathsf{G}^T(\mathcal{D},\mathcal{S})\right), \end{aligned} \tag{166}$$

for all $\mathcal{S} \subseteq [N]$ and all nonempty subsets $\mathcal{D} \subseteq [L]$ for some covariance matrix $\mathsf{K} \succeq 0$ with $\mathsf{K}_{jj} \leq P$. Here $\mathsf{K}(\mathcal{S}|\mathcal{S}^c)$ is the conditional covariance matrix of $X(\mathcal{S})$ given $X(\mathcal{S}^c)$ for $X^N \sim \mathcal{N}(0,\mathsf{K})$ and $\mathsf{G}(\mathcal{S},\mathcal{D})$ is defined such that

$$\begin{bmatrix} Y(\mathcal{D}) \\ Y(\mathcal{D}^c) \end{bmatrix} = \begin{bmatrix} \mathsf{G}(\mathcal{D},\mathcal{S}) & \mathsf{G}(\mathcal{D},\mathcal{S}^c) \\ \mathsf{G}(\mathcal{D}^c,\mathcal{S}) & \mathsf{G}(\mathcal{D}^c,\mathcal{S}^c) \end{bmatrix} \begin{bmatrix} X(\mathcal{S}) \\ X(\mathcal{S}^c) \end{bmatrix} + \begin{bmatrix} Z(\mathcal{D}) \\ Z(\mathcal{D}^c) \end{bmatrix}. \tag{167}$$

Now we are ready to prove Theorem 3. It can be shown that the rate constraint of the DDF scheme can also be expressed as

$$\begin{aligned}\sum_{\ell\in\mathcal{D}} R_\ell < & \sum_{k\in\mathcal{S}^c} C_k + \sum_{j\in\mathcal{S}} \sum_{k\in\mathcal{S}^c} C_{kj} + I(X(\mathcal{S}); U(\mathcal{D})|X(\mathcal{S}^c)) \\ & - \sum_{k\in\mathcal{S}^c} I(X_k; X(\mathcal{S}_k^c)) - \sum_{\ell\in\mathcal{D}} I(U_\ell; U(\mathcal{D}_\ell), X^N|Y_\ell), \end{aligned} \tag{168}$$

$$\frac{\mathrm{Var}(|\mathcal{A}|)}{\mathbb{E}[|\mathcal{A}|]^2} = \frac{\mathbb{E}[\mathbb{E}[|\mathcal{A}|^2|\mathcal{B}_1,\mathcal{B}_2]] - (\mathbb{E}[\mathbb{E}[|\mathcal{A}||\mathcal{B}_1,\mathcal{B}_2]])^2}{(\mathbb{E}[\mathbb{E}[|\mathcal{A}||\mathcal{B}_1,\mathcal{B}_2]])^2} = \frac{\sum_{(a_0,a_1,a_2,b_0,b_1,b_2)\neq(2,2,2,2,2,2)} \mathbb{E}[Q(a_0,a_1,a_2,b_0,b_1,b_2)]p(a_0,a_1,a_2,b_0,b_1,b_2)}{(\mathbb{E}[Q(1,1,1,1,1,1)]p(1,1,1,1,1,1))^2} \tag{161}$$

Then, we set X_k to be i.i.d. $\mathcal{N}(0,P)$ for all $k \in [N]$ and

$$U_\ell = \sum_{k=1}^{N} g_{\ell k} X_k + \hat{Z}_\ell, \tag{169}$$

where $\hat{Z}_\ell \sim \mathcal{N}(0,1)$ are mutually independent and independent of (X^N, Y^L). Then, we have

$$\begin{aligned}\sum_{\ell\in\mathcal{D}} R_\ell < &\sum_{k\in\mathcal{S}^c} C_k + \sum_{j\in\mathcal{S}}\sum_{k\in\mathcal{S}^c} C_{kj} \\ &+\frac{1}{2}\log\det\left(\mathsf{I} + P\mathsf{G}(\mathcal{D},\mathcal{S})\mathsf{G}^T(\mathcal{D},\mathcal{S})\right) \\ &-\sum_{\ell\in\mathcal{D}}\frac{1}{2}\log\left(1+\frac{\sum_{k\in\mathcal{S}} g_{\ell k}^2 P}{1+\sum_{k\in\mathcal{S}} g_{\ell k}^2 P}\right),\end{aligned} \tag{170}$$

which can be further relaxed as

$$\begin{aligned}\sum_{\ell\in\mathcal{D}} R_\ell < &\sum_{k\in\mathcal{S}^c} C_k + \sum_{j\in\mathcal{S}}\sum_{k\in\mathcal{S}^c} C_{kj} \\ &+\frac{1}{2}\log\det\left(\mathsf{I} + P\mathsf{G}(\mathcal{D},\mathcal{S})\mathsf{G}^T(\mathcal{D},\mathcal{S})\right) - \frac{|\mathcal{D}|}{2}.\end{aligned} \tag{171}$$

On the other hand, the cut-set bound for the Gaussian case is given by

$$\begin{aligned}\sum_{\ell\in\mathcal{D}} R_\ell \le &\sum_{k\in\mathcal{S}^c} C_k + \sum_{j\in\mathcal{S}}\sum_{k\in\mathcal{S}^c} C_{kj} \\ &+\frac{1}{2}\log\det\left(\mathsf{I} + \mathsf{G}(\mathcal{D},\mathcal{S})\mathsf{K}(\mathcal{S}|\mathcal{S}^c)\mathsf{G}^T(\mathcal{D},\mathcal{S})\right), \quad (172)\\ \overset{(a)}{=} &\sum_{k\in\mathcal{S}^c} C_k + \sum_{j\in\mathcal{S}}\sum_{k\in\mathcal{S}^c} C_{kj} \\ &+\frac{1}{2}\log\det\left(\mathsf{I} + \mathsf{G}^T(\mathcal{D},\mathcal{S})\mathsf{G}(\mathcal{D},\mathcal{S})\mathsf{K}(\mathcal{S}|\mathcal{S}^c)\right), \quad (173)\end{aligned}$$

where (a) follows from Sylvester's determinant identity. The term $\det\left(\mathsf{I} + \mathsf{G}^T(\mathcal{D},\mathcal{S})\mathsf{G}(\mathcal{D},\mathcal{S})\mathsf{K}(\mathcal{S}|\mathcal{S}^c)\right)$ can be upper bounded in two different ways. Note that the symmetric matrices $\mathsf{G}^T(\mathcal{D},\mathcal{S})\mathsf{G}(\mathcal{D},\mathcal{S})$ and $\mathsf{K}(\mathcal{S}|\mathcal{S}^c)$ are positive semidefinite. When $\mathcal{S}$ is an empty set, the inner bound matches the cut-set bound. In the following, we consider the case $|\mathcal{S}| \ge 1$.

First, we have

$$\begin{aligned}&\det\left(\mathsf{I} + \mathsf{G}^T(\mathcal{D},\mathcal{S})\mathsf{G}(\mathcal{D},\mathcal{S})\mathsf{K}(\mathcal{S}|\mathcal{S}^c)\right) \\ &\le \det\left(\mathsf{I} + P\mathsf{G}^T(\mathcal{D},\mathcal{S})\mathsf{G}(\mathcal{D},\mathcal{S})\right)\cdot\det\left(\mathsf{I}+\frac{1}{P}\mathsf{K}(\mathcal{S}|\mathcal{S}^c)\right) \\ &\overset{(a)}{\le} \det\left(\mathsf{I} + P\mathsf{G}(\mathcal{D},\mathcal{S})\mathsf{G}^T(\mathcal{D},\mathcal{S})\right)\cdot 2^{|\mathcal{S}|},\end{aligned} \tag{174}$$

where (a) follows from Sylvester's determinant identity and Hadamard's inequality.

Second, denote by $\lambda_j(A)$ the j-th largest eigenvalue of the symmetric matrix A. For notational convenience, we denote $\mathsf{G}' = \mathsf{G}^T(\mathcal{D},\mathcal{S})\mathsf{G}(\mathcal{D},\mathcal{S})$ and $\mathsf{K}' = \mathsf{K}(\mathcal{S}|\mathcal{S}^c)$. Note that the matrix G' has at most $|\mathcal{D}|$ nonzero eigenvalues and $\lambda_1(\mathsf{K}') \le \mathrm{tr}(\mathsf{K}') \le |\mathcal{S}|P$. Thus, we have

$$\begin{aligned}\det\left(\mathsf{I}+\mathsf{G}'\mathsf{K}'\right) &= \prod_{i=1}^{|\mathcal{S}|}(1+\lambda_i(\mathsf{G}'\mathsf{K}')) && (175)\\ &\overset{(a)}{\le} \prod_{i=1}^{|\mathcal{S}|}(1+\lambda_i(\mathsf{G}')\lambda_1(\mathsf{K}')) && (176)\\ &\le \prod_{i=1}^{|\mathcal{S}|}(1+\lambda_i(\mathsf{G}')|\mathcal{S}|P) && (177)\\ &= \det\left(\mathsf{I}+|\mathcal{S}|P\mathsf{G}(\mathcal{D},\mathcal{S})\mathsf{G}^T(\mathcal{D},\mathcal{S})\right) && (178)\\ &\le \det\left(\mathsf{I}+P\mathsf{G}(\mathcal{D},\mathcal{S})\mathsf{G}^T(\mathcal{D},\mathcal{S})\right)\cdot|\mathcal{S}|^{|\mathcal{D}|}, && (179)\end{aligned}$$

where (a) follows from [25, 7.3.P16].

To summarize, the cut-set bound can be relaxed as

$$\begin{aligned}\sum_{\ell\in\mathcal{D}} R_\ell \le &\sum_{k\in\mathcal{S}^c} C_k + \sum_{j\in\mathcal{S}}\sum_{k\in\mathcal{S}^c} C_{kj} \\ &+\det\left(\mathsf{I} + P\mathsf{G}(\mathcal{D},\mathcal{S})\mathsf{G}^T(\mathcal{D},\mathcal{S})\right) \\ &+\frac{1}{2}\min\{|\mathcal{S}|, |\mathcal{D}|\log|\mathcal{S}|\}.\end{aligned} \tag{180}$$

Comparing the relaxed inner bound (171) and outer bound (180), we conclude that the DDF scheme achieves within $\min\left\{\frac{L+N}{2}, \frac{L+L\log N}{2}\right\}$ bits per dimension from the cut-set bound and thus from the capacity region.

References

[1] O. Simeone, A. Maeder, M. Peng, O. Sahin, and W. Yu, "Cloud radio access network: Virtualizing wireless access for dense heterogeneous systems," *Journal of Communications and Networks*, vol. 18, pp. 135–149, Apr. 2016.

[2] M. Peng, Y. Sun, X. Li, Z. Mao, and C. Wang, "Recent advances in cloud radio access networks: System architectures, key techniques, and open issues," *IEEE Communications Surveys & Tutorials*, vol. 18, pp. 2282–2308, thirdquarter 2016.

[3] R. Zakhour and D. Gesbert, "Optimized data sharing in multicell MIMO with finite backhaul capacity," *IEEE Trans. Signal Process.*, vol. 59, pp. 6102–6111, Dec. 2011.

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

[5] 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, pp. 5646–5658, Nov. 2013.

[6] P. Patil and W. Yu, "Generalized compression strategy for the downlink cloud radio access network," in *arXiv:1801.00394*, Jan. 2018.

[7] 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.

[8] S. N. Hong and G. Caire, "Compute-and-forward strategies for cooperative distributed antenna systems," *IEEE Trans. Inf. Theory*, vol. 59, pp. 5227–5243, Sep. 2013.

[9] N. Liu and W. Kang, “A new achievability scheme for downlink multicell processing with finite backhaul capacity,” in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Honolulu, HI, Jun. 2014.

[10] W. Kang, N. Liu, and W. Chong, “The Gaussian multiple access diamond channel,” *IEEE Trans. Inf. Theory*, vol. 61, pp. 6049–6059, Nov. 2015.

[11] K. Johansson, A. Furuskär, and J. Zander, “Modelling the cost of heterogeneous wireless access networks,” *Int. J. Mobile Network Design and Innovation*, vol. 2, no. 1, pp. 58–66, May 2007.

[12] B. Timus and J. Zander, “Incremental deployment with self-backhauling base stations in urban environment,” in *2009 IEEE International Conference on Communications Workshops*, June 2009, pp. 1–5.

[13] H. Lundquist, “Death by starvation?:backhaul and 5g,” *ComSoc Technology News*, Sep. 2015. [Online]. Available: https://www.comsoc.org/ctn/death-starvation-backhaul-and-5g

[14] R. Gray and A. Wyner, “Source coding for a simple network,” *Bell Systems Technical Journal*, vol. 53, no. 9, pp. 1681–1721, 1974.

[15] A. El Gamal and Y.-H. Kim, *Network Information Theory*. New York: Cambridge University Press, 2011.

[16] A. Orlitsky and J. R. Roche, “Coding for computing,” *IEEE Trans. Inf. Theory*, vol. 47, pp. 903–917, Mar. 2001.

[17] I. B. Gattegno, Z. Goldfeld, and H. H. Permuter, “Fourier–Motzkin elimination software for information theoretic inequalities,” in *arXiv:1610.03990 [cs.IT]*, Oct. 2016.

[18] K. Marton, “A coding theorem for the discrete memoryless broadcast channel,” *IEEE Trans. Inf. Theory*, vol. 25, pp. 306–311, May 1979.

[19] S. S. Bidokhti and G. Kramer, “Capacity bounds for diamond networks with an orthogonal broadcast channel,” *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7103–7122, Dec 2016.

[20] W. Zhao, D. Y. Ding, and A. Khisti, “Capacity bounds for a class of diamond networks with conferencing relays,” *IEEE Commun. Lett.*, vol. 19, pp. 1881–1884, Nov. 2015.

[21] T. Cover and A. El Gamal, “Capacity theorems for the relay channel,” *IEEE Trans. Inf. Theory*, vol. 25, pp. 572–584, Sep. 1979.

[22] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless systems*, 2nd ed. Cambridge University Press, 2011.

[23] D. Traskov and G. Kramer, “Reliable communication in networks with multi-access interference,” in *Proc. IEEE Information Theory Workshop (ITW)*, Lake Tahoe, CA, Sep. 2007.

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[25] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge University Press, 2013.