On the Optimal Fronthaul Compression and Decoding Strategies for Uplink Cloud Radio Access Networks

Yuhan Zhou, *Member, IEEE*, Yinfei Xu, *Student Member, IEEE*, Wei Yu, *Fellow, IEEE*, and Jun Chen, *Senior Member, IEEE*

***Abstract*— This paper investigates the compress-and-forward scheme for an uplink cloud radio access network (C-RAN) model, where multi-antenna base stations (BSs) are connected to a cloud-computing-based central processor (CP) via capacity-limited fronthaul links. The BSs compress the received signals with Wyner-Ziv coding and send the representation bits to the CP; the CP performs the decoding of all the users' messages. Under this setup, this paper makes progress toward the optimal structure of the fronthaul compression and CP decoding strategies for the compress-and-forward scheme in the C-RAN. On the CP decoding strategy design, this paper shows that under a sum fronthaul capacity constraint, a generalized successive decoding strategy of the quantization and user message codewords that allows arbitrary interleaved order at the CP achieves the same rate region as the optimal joint decoding. Furthermore, it is shown that a practical strategy of successively decoding the quantization codewords first, then the user messages, achieves the same maximum sum rate as joint decoding under individual fronthaul constraints. On the joint optimization of user transmission and BS quantization strategies, this paper shows that if the input distributions are assumed to be Gaussian, then under joint decoding, the optimal quantization scheme for maximizing the achievable rate region is Gaussian. Moreover, Gaussian input and Gaussian quantization with joint decoding achieve to within a constant gap of the capacity region of the Gaussian multiple-input multiple-output (MIMO) uplink C-RAN model. Finally, this paper addresses the computational aspect of optimizing uplink MIMO C-RAN by showing that under fixed Gaussian input, the sum rate maximization problem over the Gaussian quantization noise covariance matrices can be formulated as convex optimization problems, thereby facilitating its efficient solution.**

Manuscript received January 19, 2016; revised August 19, 2016; accepted October 2, 2016. Date of publication October 13, 2016; date of current version November 18, 2016. This work was supported by the Natural Sciences and Engineering Research Council of Canada. Y. Xu was supported in part by a Grant from the University Grants Committee of the Hong Kong Special Administrative Region, China, under Project AoE/E- 02/08 and in part by the National Natural Science Foundation of China under Grant 61401086. This paper was presented in part at the 2015 IEEE International Symposium on Information Theory.

Y. Zhou was with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada. He is now with Qualcomm Technologies Inc., San Diego, CA 92121 USA (e-mail: yzhou@ece.utoronto.ca).

Y. Xu was with the School of Information Science and Engineering, Southeast University, Nanjing 210096, China. He is now with the Department of Information Engineering, Institute of Network Coding, The Chinese University of Hong Kong, Hong Kong (e-mail: yinfeixu@inc.cuhk.edu.hk).

W. Yu is with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: weiyu@ece.utoronto.ca).

J. Chen is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4L8, Canada (e-mail: junchen@ece.mcmaster.ca).

Communicated by O. Simeone, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2016.2617862

***Index Terms*— Cloud radio access network, multiple-access relay channel, compress-and-forward, fronthaul compression, joint decoding, generalized successive decoding.**

I. Introduction

Cloud Radio Access Network (C-RAN) is an emerging mobile network architecture in which base-stations (BSs) in multiple cells are connected to a cloud-computing based central processor (CP) through wired/wireless fronthaul links. In the deployment of a C-RAN system, the BSs degenerate into remote antennas heads implementing only radio functionalities, such as frequency up/down conversion, sampling, filtering, and power amplification. The baseband operations at the BSs are migrated to the CP. The C-RAN model effectively virtualizes radio-access operations such as the encoding and decoding of user information and the optimization of radio resources [1]. Advanced joint multicell processing techniques, such as the coordinated multi-point (CoMP) and network multiple-input multiple-output (MIMO), can be efficiently supported by the C-RAN architecture, potentially enabling significantly higher data rates than conventional cellular networks [2].

This paper considers the uplink of a MIMO C-RAN system under finite-capacity fronthaul constraints, as shown in Fig. 1, which consists of multiple remote users sending independent messages to the CP through multiple BSs serving as relay nodes. Both the user terminals and the BSs are equipped with multiple antennas. The BSs and the CP are connected via noiseless fronthaul links with finite capacity. This channel model can be thought of as a two-hop relay network, with an interference channel between the users and the BSs, followed by a noiseless multiple-access channel between the BSs and the CP. This paper assumes that a compress-and-forward relaying strategy is employed, in which the relaying BSs perform distributed lossy source coding to compress the received signals and forward the representation bits to the CP through digital fronthaul links, and all the user messages are eventually decoded at the CP. The lossy source coding implemented at BSs involves Wyner-Ziv coding typically consisting of quantization followed by binning in order to achieve high compression efficiency by leveraging the correlation between the received signals across different BSs, which is different from the point-to-point fronthaul compression implemented in today's conventional C-RAN systems.

A key question in the design of compress-and-forward strategy in uplink C-RAN is the optimal input coding strategy

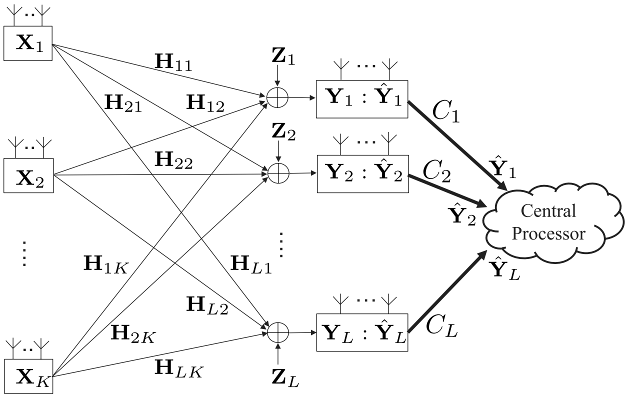

Fig. 1. The uplink C-RAN model under finite-capacity fronthaul constraints.

at the user terminals, the optimal relaying strategy at the BSs, and the optimal decoding strategy at the CP. Toward this end, this paper restricts attention to the strategy of compressing the received signals at the BSs, then either *joint decoding* of the quantization and message codewords simultaneously, or *generalized successive decoding* of the quantization and message codewords in some arbitrary order at the CP. Under this assumption, this paper makes the following contributions toward revealing the structure of the optimal compress-and-forward strategy.

First, motivated by the fact that successive decoding is much easier to implement than joint decoding, we seek to understand whether successive decoding at the CP can perform as well as joint decoding. Toward this end, this paper shows that generalized successive decoding indeed achieves the same rate region as joint decoding for an uplink C-RAN model under a sum fronthaul constraint. Further, although not necessarily so for the general rate region, if one focuses on maximizing the sum rate, the particular strategy of successively decoding the quantization codewords first, then the user messages, achieves the optimal sum rate.

Second, we seek to understand the optimal input distribution and quantization schemes in uplink C-RAN. Although it is well known that joint Gaussian strategies are not necessarily optimal, this paper shows that if we fix the input distribution to be Gaussian, then the optimal quantization scheme is Gaussian under joint decoding, and vice versa. Moreover, joint Gaussian signaling can be shown to achieve the capacity region of the Gaussian multiple-input multiple-output (MIMO) uplink C-RAN model to within a constant gap. Finally, this paper makes progress on the computational front by showing that under the joint Gaussian assumption, the optimization of the quantization covariance matrices for maximizing the sum rate can be formulated as a convex optimization problem. These results suggest that joint Gaussian input signaling and Gaussian quantization is a suitable strategy for the uplink C-RAN.

A. Related Work

The achievable rate region of compress-and-forward with joint decoding for the uplink C-RAN model was first characterized in [3] for a single-transmitter model then in [4] for the multi-transmitter case. However, the number of rate constraints in the joint decoding rate region grows exponentially with the size of the network [3, Proposition IV.1], which makes the evaluation of the achievable rate computationally prohibitive. The achievable rate region of the compress-and-forward strategy with practical successive decoding, in which the quantization codewords are decoded first, then the user messages are decoded based on the recovered quantization codewords, has also been studied for the uplink C-RAN model [5, Theorem 1]. One of the objectives of this paper is to illustrate the relationship between joint decoding and successive decoding. In the existing literature, the equivalence between these two decoding schemes is first demonstrated for single-source, single-destination, and single-relay networks [6, Appendix 16C], then shown for single-source, single-destination, and multiple-relay networks [7], under either block-by-block forward decoding or block-by-block backward decoding. This paper further demonstrates that in the case of uplink C-RAN, which is a multiple-source, single-destination, multiple-relay network, the optimality of successive decoding still holds under suitable conditions.

In general, it is challenging to find the optimal joint input and quantization noise distributions that maximize the achievable rate of the compress-and-forward scheme for uplink C-RAN. Gaussian signaling is not necessarily optimal—in particular, in a simple example of uplink C-RAN with one user and two BSs shown in [5], binary input is shown to outperform Gaussian input for a broad range of signal-to-noise ratios (SNRs). However, Gaussian input and Gaussian quantization can be shown to be approximately optimal. In fact, the uplink C-RAN model is an example of a general Gaussian relay network with multiple sources and a single destination for which a generalization of compress-and-forward with joint decoding (referred to as noisy network coding scheme [8]–[11]) and with Gaussian input and Gaussian quantization can be shown to achieve to within a constant gap to the information theoretical capacity of the overall network. Instead of using noisy network coding, our previous work [12] shows that successive decoding can achieve the sum capacity of uplink C-RAN to within constant gap, if the fronthaul links are subjected to a sum capacity constraint. In this work, we further demonstrate that the compress-and-forward scheme with joint decoding can achieve to within a constant gap to the entire capacity region of the uplink C-RAN model with individual fronthaul constraints; same is true for successive decoding under suitable condition.

An important theoretical result obtained in this paper is that if the input distributions of the uplink C-RAN model are fixed to be Gaussian, then Gaussian quantizer is in fact optimal under joint decoding. Finding the optimal quantization for the C-RAN model is related to the mutual information constraint problem [13], for which entropy power inequality is used to show that Gaussian quantization is optimal for a three-node relay network with Gaussian input. However, it is challenging to extend this approach to the uplink C-RAN model, which has multiple sources. This paper provides a novel proof of the optimality of Gaussian quantization based on the de Bruijn identity and the Fisher information inequality. The idea of the proof is inspired by the connection between the C-RAN model and the CEO problem in source coding [14], where a source is described to a central unit by remote agents with noisy observations. The solution to the CEO problem is known for the scalar Gaussian case [15], [16]; significant recent progress has been made in the vector case, e.g., [17]. The similarity between the uplink C-RAN model and the CEO problem has been noted in [5], based on which a capacity upper bound for the uplink C-RAN model is established. In this paper, we use techniques for establishing the outer bound for the Gaussian vector CEO problem [18] to prove the optimality of Gaussian quantization. We also remark the connection between this quantization optimization problem and the information bottleneck method [19], for which joint Gaussian distribution is shown to be Pareto optimal. The technique used in this paper is a significantly simpler alternative to the enhancement technique given in [20].

This paper also makes progress in observing that the optimization of Gaussian quantization noise covariance matrices for maximizing the (weighted) sum rate of uplink C-RAN can be reformulated as a convex optimization problem. The quantization noise covariance optimization problem for uplink C-RAN has been considered extensively in the literature. Various optimization algorithms have been developed to maximize the achievable rates of the compress-and-forward scheme for the case of either successive decoding of the quantization codewords followed by the user messages [21], [22] or joint decoding of the quantization codewords and user messages simultaneously [23]. In particular, a zero-duality gap result has been shown for the weighted sum rate maximization problem under a sum fronthaul capacity constraint in [21] based on a time-sharing argument to facilitate the algorithm design for searching optimal quantization noise covariance matrices. However, the optimization problems formulated in these works (i.e., [21]–[23]) are inherently nonconvex, hence only locally convergent algorithms are obtained. Instead, this paper provides a convex formulation of the problem that allows globally optimal Gaussian quantization noise covariance matrices to be found. Note that here the optimization of the quantization noise covariance matrix is performed under the fixed Gaussian input. The joint optimization of the input signal and quantization noise covariance matrices remains a computationally challenging difficult problem [24].

B. Main Contributions

This paper establishes several information theoretic results on the compress-and-forward scheme for the uplink MIMO C-RAN model with finite-capacity fronthaul links. A summary of our main contributions is as follows:

- This paper demonstrates that generalized successive decoding for compress-and-forward, which allows the decoding of the quantization and user message codewords in an arbitrary order, can achieve the same rate region as joint decoding for compress-and-forward under a sum fronthaul capacity constraint. Further, successive decoding of the quantization codewords first, then the user message codewords, can achieve the same maximum sum rate as joint decoding under individual fronthaul constraints.
- This paper shows that under Gaussian input and Gaussian quantization, compress-and-forward with joint decoding achieves to within a constant gap of the capacity region of the uplink MIMO C-RAN model. Combining with the result above, the same constant-gap result also holds for generalized successive decoding under a sum fronthaul constraint and for successive decoding for sum rate maximization.
- This paper shows that under fixed Gaussian input, Gaussian quantization maximizes the achievable rate region under joint decoding. Combining with the optimality result for successive decoding, this also implies that under fixed Gaussian input, Gaussian quantization is optimal for generalized successive decoding under a sum fronthaul constraint, and for successive decoding for sum rate maximization.
- Under joint Gaussian signaling and Gaussian quantization, the optimization of quantization noise covariance matrices for maximizing weighted sum rate under joint decoding and for maximizing sum rate under successive decoding can be formulated as convex optimization problems, which facilitate their efficient solution.

C. Paper Organization and Notation

The rest of the paper is organized as follows. Section II introduces the channel model for the uplink MIMO C-RAN and characterizes the achievable rate regions for compress-and-forward schemes with joint decoding and generalized successive decoding respectively. Section III demonstrates the rate-region optimality of generalized successive decoding under a sum fronthaul constraint and the sum-rate optimality of successive decoding. Section IV focuses on establishing the optimality of Gaussian quantizers with joint decoding under Gaussian input. In addition, Section IV also establishes the approximate capacity of the uplink MIMO C-RAN model to within constant gap, and shows the convex formulation of the (weighted) sum rate maximization problems over the quantization noise covariance matrices. Section V concludes the paper.

Notation: Boldface letters denote vectors or matrices, where context should make the distinction clear. Superscripts $(\cdot)^{\mathsf{T}}$, $(\cdot)^{\dagger}$ and $(\cdot)^{-1}$ denote transpose operation, Hermitian transpose and matrix inverse operators; $\mathbb{E}[\cdot]$ and $\mathrm{Tr}(\cdot)$ denote expectation and matrix trace operators; $\mathrm{co}(\cdot)$ denotes the convex closure operation; $p(\cdot)$ denotes the probability distribution function in this paper. We use $\mathbf{X}_i^j = \left(\mathbf{X}_i, \mathbf{X}_{i+1}, \ldots, \mathbf{X}_j\right)$ to denote a matrix with $(j - i + 1)$ columns for $1 \leq i \leq j$. For a vector/matrix $\mathbf{X}$, $\mathbf{X}_{\mathcal{S}}$ denotes a vector/matrix with elements whose indices are elements of $\mathcal{S}$. Given matrices $\{\mathbf{X}_1, \ldots, \mathbf{X}_L\}$, $\mathrm{diag}\left(\{\mathbf{X}_\ell\}_{\ell=1}^L\right)$ denotes the block diagonal matrix formed with $\mathbf{X}_\ell$ on the diagonal. For random vectors $\mathbf{X}$ and $\mathbf{Y}$, $\mathbf{J}(\mathbf{X}|\mathbf{Y})$ denotes the Fisher information matrix of $\mathbf{X}$ conditional on $\mathbf{Y}$; $\mathrm{cov}(\mathbf{X}|\mathbf{Y})$ denotes the covariance matrix of $\mathbf{X}$ conditional on $\mathbf{Y}$.

II. Achievable Rate Regions for Uplink C-RAN

A. Channel Model

This paper considers an uplink C-RAN model, where K mobile users communicate with a CP through L BSs, as shown in Fig. 1. The noiseless digital fronthaul link connecting the BS ℓ to the CP has the capacity of C_ℓ bits per complex dimension. The fronthaul capacity C_ℓ is the maximum long-term average throughput of the ℓth fronthaul link, i.e., $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} C_\ell(i) \le C_\ell$, where $C_\ell(i)$ represents the instantaneous transmission rate of the ℓth fronthaul link at the ith time slot. Each user terminal is equipped with M antennas; each BS is equipped with N antennas. Perfect channel state information (CSI) is assumed to be available to all the BSs and to the CP. For simple notation, we denote $\mathcal{K} = \{1, \cdots, K\}$ and $\mathcal{L} = \{1, \cdots, L\}$ in this paper.

Let $\mathbf{X}_k \in \mathbb{C}^M$ be the signal transmitted by the kth user, which is subject to per-user transmit power constraint of P_k, i.e. $\mathbb{E}\left[\mathbf{X}_k\mathbf{X}_k^\dagger\right] \le P_k$. The signal received at the ℓth BS can be expressed as

$$\mathbf{Y}_\ell = \sum_{k=1}^{K} \mathbf{H}_{\ell,k}\mathbf{X}_k + \mathbf{Z}_\ell, \quad \ell = 1, 2, \ldots, L, \tag{1}$$

where $\mathbf{Z}_\ell \sim \mathcal{CN}(\mathbf{0}, \Sigma_\ell)$ represents the additive Gaussian noise for BS ℓ and is independent across different BSs, and $\mathbf{H}_{\ell,k}$ denotes the complex channel matrix from user k to BS ℓ.

We consider the compress-and-forward scheme [25], [26] applied to the uplink C-RAN system, in which the BSs compress the received signals $\mathbf{Y}_\ell$, and forward the quantization bits to the CP for decoding. At the CP, the user messages are decoded using either joint decoding or some form of successive decoding. In joint decoding, the quantization codewords and the message codewords are decoded *simultaneously*, whereas, in successive decoding, the quantization codewords and messages are decoded *successively* in some prescribed order. Different orderings can potentially result in different achievable rates.

B. Achievable Rate-Fronthaul Regions for Joint Decoding, Successive Decoding, and Generalized Successive Decoding

In the following, we present the achievable rate-fronthaul regions of compress-and-forward with joint decoding and different forms of successive decoding.

Proposition 1 ([3, Proposition IV.1]): For the uplink C-RAN model shown in Fig. 1, the achievable rate-fronthaul region of compress-and-forward with joint decoding, $\mathcal{P}_{JD}^*$, is the closure of the convex hull of all $(R_1, \cdots, R_K, C_1, \ldots, C_L) \in \mathbb{R}_+^{K+L}$ satisfying

$$\sum_{k\in\mathcal{T}} R_k < \sum_{\ell\in\mathcal{S}} \left[C_\ell - I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_\mathcal{K}\right)\right] + I\left(\mathbf{X}_\mathcal{T}; \hat{\mathbf{Y}}_{\mathcal{S}^c}|\mathbf{X}_{\mathcal{T}^c}\right) \tag{2}$$

for all $\mathcal{T} \subseteq \mathcal{K}$ and $\mathcal{S} \subseteq \mathcal{L}$, for some product distribution $\prod_{k=1}^{K} p(\mathbf{x}_k)\prod_{\ell=1}^{L} p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell)$ such that $\mathbb{E}\left[\mathbf{X}_k\mathbf{X}_k^\dagger\right] \le P_k$ for $k = 1, \ldots, K$.

Note that for the uplink C-RAN model, the rate region (2) given by compress-and-forward with joint decoding is identical to the rate region of the noisy network coding scheme [9], which is an extension of the compress-and-forward scheme to the general multiple access relay network by using joint decoding at the receiver and block Markov coding at the transmitters.

As a more practical decoding strategy, successive decoding of quantization codewords first, and then the user messages at the CP can also be used in uplink C-RAN. The following proposition states the rate-fronthaul region achieved by successive decoding.

Proposition 2: ([5, Theorem 1]): For the uplink C-RAN model shown in Fig. 1, the achievable rate-fronthaul region of compress-and-forward with successive decoding, $\mathcal{P}_{SD}^*$, is the closure of the convex hull of all $(R_1, \cdots, R_K, C_1, \ldots, C_L) \in \mathbb{R}_+^{K+L}$ satisfying

$$\sum_{k\in\mathcal{T}} R_k < I\left(\mathbf{X}_\mathcal{T}; \hat{\mathbf{Y}}_\mathcal{L}|\mathbf{X}_{\mathcal{T}^c}\right), \quad \forall\, \mathcal{T} \subseteq \mathcal{K}, \tag{3}$$

and

$$I\left(\mathbf{Y}_\mathcal{S}; \hat{\mathbf{Y}}_\mathcal{S}|\hat{\mathbf{Y}}_{\mathcal{S}^c}\right) < \sum_{\ell\in\mathcal{S}} C_\ell, \quad \forall\, \mathcal{S} \subseteq \mathcal{L}, \tag{4}$$

for some product distribution $\prod_{k=1}^{K} p(\mathbf{x}_k)\prod_{\ell=1}^{L} p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell)$ such that $\mathbb{E}\left[\mathbf{X}_k\mathbf{X}_k^\dagger\right] \le P_k$ for $k = 1, \ldots, K$.

Note that (3) is the multiple-access rate region, (4) represents the Berger-Tung rate region for distributed lossy compression [6, Theorem 12.1], while (2) incorporates the joint decoding of the quantization codewords and the user messages. Because of its lower decoding complexity, successive decoding is usually preferred for practical implementation of the uplink C-RAN systems [21], [22]. Note that in the above strategy, successive decoding applies only to the vector $\mathbf{X}_k$ (user message codewords) and the vector $\mathbf{Y}_\ell$ (quantization codewords); the elements within vectors $\mathbf{X}_k$ and $\mathbf{Y}_\ell$ are still decoded jointly.

It is possible to improve upon the successive decoding scheme by allowing arbitrary interleaved decoding orders between quantization codewords and user message codewords. We call this the generalized successive decoding scheme in this paper. The generalized successive decoding scheme is first suggested in [27] under the name of joint base-station successive interference cancelation scheme. In such a successive decoding strategy, the set of potential decoding orders includes all the permutations of quantization and user message codewords.

Denote π as a permutation on the set of quantization and user message codewords $\left(\hat{\mathbf{Y}}_1, \hat{\mathbf{Y}}_2, \ldots, \hat{\mathbf{Y}}_L, \mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_K\right)$. For a given permutation π, the decoding order is given by the index of the elements in π, i.e., $\pi(1) \to \pi(2) \to \cdots \to \pi(L + K)$. For example, consider an uplink C-RAN model as shown in Fig. 1 with 2 BSs and 2 users. If $\pi = \left(\hat{\mathbf{Y}}_1, \mathbf{X}_1, \hat{\mathbf{Y}}_2, \mathbf{X}_2\right)$, then the decoding of $\hat{\mathbf{Y}}_2$ and $\mathbf{X}_2$ can use both previously decoded user messages and quantization codewords as side information. The resulting rate region is characterized as

$$\begin{cases} R_1 < I\left(\mathbf{X}_1; \hat{\mathbf{Y}}_1\right), \\ R_2 < I\left(\mathbf{X}_2; \hat{\mathbf{Y}}_1, \hat{\mathbf{Y}}_2 | \mathbf{X}_1\right), \end{cases} \tag{5}$$

for some product distribution $p(\mathbf{x}_1)p(\mathbf{x}_2)p(\hat{\mathbf{y}}_1|\mathbf{y}_1)p(\hat{\mathbf{y}}_2|\mathbf{y}_2)$ that satisfies

$$\begin{cases} C_1 > I\left(\mathbf{Y}_1; \hat{\mathbf{Y}}_1\right), \\ C_2 > I\left(\mathbf{Y}_2; \hat{\mathbf{Y}}_2 | \hat{\mathbf{Y}}_1, \mathbf{X}_1\right). \end{cases} \tag{6}$$

Let $\mathcal{I}_{\mathbf{X}_k}$, $\mathcal{I}_{\mathbf{Y}_\ell}$ denote the indices of user messages that are decoded before $\mathbf{X}_k$ and $\mathbf{Y}_\ell$ under the permutation π, respectively. Likewise, let $\mathcal{J}_{\mathbf{X}_k}$, $\mathcal{J}_{\mathbf{Y}_\ell}$ denote the indices of quantization codewords that are decoded before $\mathbf{X}_k$ and $\mathbf{Y}_\ell$ under the permutation π, respectively. The rate-fronthaul region of generalized successive decoding for uplink C-RAN is stated in the following proposition.

Proposition 3: For the uplink C-RAN model shown in Fig. 1, the achievable rate-fronthaul region of generalized successive decoding with decoding order π, $\mathcal{P}_{GSD}(\pi)$, is the closure of the convex hull of all $(R_1, \cdots, R_K, C_1, \ldots, C_L) \in \mathbb{R}_+^{K+L}$ satisfying

$$R_k < I\left(\mathbf{X}_k; \hat{\mathbf{Y}}_{\mathcal{J}_{\mathbf{X}_k}} | \mathbf{X}_{\mathcal{I}_{\mathbf{X}_k}}\right), \quad \forall\, k \in \mathcal{K}, \tag{7}$$

and

$$C_\ell > I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell | \hat{\mathbf{Y}}_{\mathcal{J}_{\mathbf{Y}_\ell}}, \mathbf{X}_{\mathcal{I}_{\mathbf{Y}_\ell}}\right), \quad \forall\, \ell \in \mathcal{L}, \tag{8}$$

for some product distribution $\prod_{k=1}^{K} p\left(\mathbf{x}_k\right) \prod_{\ell=1}^{L} p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell)$ such that $\mathbb{E}\left[\mathbf{X}_k\mathbf{X}_k^\dagger\right] \leq P_k$ for $k = 1, \ldots, K$. The generalized successive decoding region $\mathcal{P}_{GSD}^*$ is defined to be the closure of the convex hull of the union of regions $\mathcal{P}_{GSD}(\pi)$ over all possible permutation π's, i.e.,

$$\mathcal{P}_{GSD}^* = \mathrm{co}\left(\bigcup_{\pi} \mathcal{P}_{GSD}(\pi)\right). \tag{9}$$

III. Optimality of Successive Decoding

In general, we have $\mathcal{P}_{SD}^* \subseteq \mathcal{P}_{GSD}^* \subseteq \mathcal{P}_{JD}^*$. However, successive decoding is more desirable than joint decoding, not only because of its lower complexity, but also due to the fact that its rate region can be more easily evaluated. Thus, there is a tradeoff between complexity and performance in designing decoding strategies for uplink C-RAN. To further understand this tradeoff, this section establishes that: 1) By allowing arbitrary decoding orders of quantization and message codewords, the generalized successive decoding actually achieves the same rate region as joint decoding under a sum fronthaul constraint; 2) The practical successive decoding strategy in which the BSs decode the quantization codewords first, then the user messages, actually achieves the same maximum sum rate as joint decoding under individual fronthaul constraints.

A. Optimality of Generalized Successive Decoding Under a Sum Fronthaul Constraint

This section shows that in the special case where the fronthaul links are subject to a sum capacity constraint, generalized successive decoding achieves the rate region as joint decoding. In this model, the fronthaul capacities are constrained by $\sum_{\ell=1}^{L} C_\ell \leq C$ and $C_\ell \geq 0$, justifiable in situations where the fronthaul are implemented in shared medium (e.g. wireless fronthaul links), as has been considered in [21] and [12]. Under the sum fronthaul capacity constraint C, the rate regions achieved by with joint decoding $\mathcal{R}_{JD,s}^*$ is defined as

$$\mathcal{R}_{JD,s}^* = \left\{ (R_1, \ldots, R_K) \,\middle|\, \begin{array}{c} (R_1, \cdots, R_K, C_1, \ldots, C_L) \in \mathcal{P}_{JD}^*, \\ \sum\limits_{\ell=1}^{L} C_\ell \leq C, \;\; C_\ell \geq 0 \end{array} \right\}. \tag{10}$$

Likewise, the rate region achieved with generalized successive decoding $\mathcal{R}_{GSD,s}^*$ is given by

$$\mathcal{R}_{GSD,s}^* = \left\{ (R_1, \ldots, R_K) \,\middle|\, \begin{array}{c} (R_1, \cdots, R_K, C_1, \ldots, C_L) \in \mathcal{P}_{GSD}^*, \\ \sum_{\ell=1}^{L} C_\ell \leq C, \;\; C_\ell \geq 0 \end{array} \right\}. \tag{11}$$

The following theorem states the main result of this section.

Theorem 1: For the uplink C-RAN model with the sum fronthaul capacity constraint $\sum_{\ell=1}^{L} C_\ell \leq C$ and $C_\ell \geq 0$, the rate region achieved by generalized successive decoding and joint coding are identical, i.e., $\mathcal{R}_{GSD,s}^* = \mathcal{R}_{JD,s}^*$.

Proof: See Appendix A. ■

The roadmap for the proof of Theorem 1 shares the same idea as the characterization of the rate distortion region for the CEO problem under logarithmic loss [28] and the capacity region for the multiple-access channel [29], which uses the properties of submodular polyhedron (see Appendix B). Specifically, in order to show $\mathcal{R}_{GSD,s}^* = \mathcal{R}_{JD,s}^*$, we show that under fixed product distribution $\prod_{k=1}^{K} p(\mathbf{x}_k) \prod_{\ell=1}^{L} p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell)$, every extreme point of the polyhedron $(\mathcal{R}_{JD,s}^*, C)$ is dominated by the points in the polyhedron defined by $(\mathcal{R}_{GSD,s}^*, C)$. We conjecture that Theorem 1 holds also for the case of individual fronthaul capacity constraints. However, in that case, finding the dominant faces of polyhedron $\mathcal{P}_{JD}^*$ becomes much more difficult, it appears non-trivial to extend the current proof to the case of individual fronthaul constraints.

B. Optimality of Successive Decoding for Maximizing Sum Rate

As a special instance of generalized successive decoding, successive decoding reconstructs quantization codewords first, then user message codewords in a sequential order. In what follows, we show that the optimal sum rate achieved by this special successive decoding is the same as that achieved by joint decoding.

Under fixed input distribution and fixed fronthaul capacities C_ℓ, for $\ell = 1, \dots, L$, the maximum sum rate achieved by joint decoding $R^*_{JD,SUM}$ is defined as

$$R^*_{JD,SUM} = \begin{cases} \max \displaystyle\sum_{k=1}^{K} R_k \\ \text{s.t. } (R_1, \cdots, R_K, C_1, \dots, C_L) \in \mathcal{P}^*_{JD}. \end{cases} \tag{12}$$

Likewise, the maximum sum rate for successive decoding $R_{SD,SUM}$ is given by

$$R^*_{SD,SUM} = \begin{cases} \max \displaystyle\sum_{k=1}^{K} R_k \\ \text{s.t. } (R_1, \cdots, R_K, C_1, \dots, C_L) \in \mathcal{P}^*_{SD}. \end{cases} \tag{13}$$

The following theorem demonstrates the optimality of successive decoding for maximizing uplink C-RAN under individual fronthaul constraints.

Theorem 2: For the uplink C-RAN model with fronthaul capacities C_ℓ shown in Fig. 1, the maximum sum rates achieved by successive decoding and joint decoding are the same, i.e., $R^*_{SD,SUM} = R^*_{JD,SUM}$.

Proof: See Appendix C. ■

We remark that Theorem 2 can be thought as a generalization of a result in [7] that shows under block-by-block forward decoding, the compress-and-forward scheme with compression-message successive decoding achieves the same maximum rate as that with compression-message joint decoding for a single-source, single-destination relay network. The uplink C-RAN is a multiple-source, single-destination relay network. If all the user terminals are regarded as one super transmitter, then it follows from [7] that successive decoding and joint decoding achieve the same maximum sum rate. However, the proof in [7] is quite complicated. In this paper, we provide an alternative proof technique for showing the optimality of successive decoding for sum rate maximization in uplink C-RAN. The new proof utilizes the properties of submodular optimization, which is simpler than the proof provided in [7]. The proofs of Theorem 2 and Theorem 1 illustrate the usefulness of submodular optimization in establishing this type of results.

It is remarked that successive decoding and joint decoding achieve the same sum rate, but do not achieve the same rate region. The achievable rate region of generalized successive decoding is in general larger than that of successive decoding. For example, consider the compress-and-forward scheme for maximizing the rate of user 1, R_1, only. The optimal decoding order should be $\mathbf{X}_{\mathcal{K}\setminus\{1\}} \rightarrow \hat{\mathbf{Y}}_{\mathcal{L}} \rightarrow \mathbf{X}_1$. With this decoding order, user 1 can achieve larger rate than using the decoding order of $\hat{\mathbf{Y}}_{\mathcal{L}} \rightarrow \mathbf{X}_{\mathcal{K}}$, because the decoded user messages $\mathbf{X}_2, \mathbf{X}_3, \dots, \mathbf{X}_K$ can serve as side information for the decoding of $\hat{\mathbf{Y}}_{\mathcal{L}}$. In general, to maximize a weighted sum rate, one needs to maximize over $(L+K)!$ orderings for generalized successive decoding. The main result of this section shows however that for maximizing the sum rate in uplink C-RAN, successive decoding of the quantization codewords first, and then the user messages is optimal; this reduces the search space considerably to $L!K!$ decoding orders.

IV. Uplink C-RAN With Gaussian Input and Gaussian Quantization

In this section, we specialize to the compress-and-forward scheme for uplink C-RAN with Gaussian input signal at the users and Gaussian quantization at the BSs. Although it is known that joint Gaussian distribution is suboptimal for uplink C-RAN [5], Gaussian input is desirable, because it leads to achievable rate regions that can be easily evaluated. In the following section, it is shown that with Gaussian input and Gaussian quantization, compress-and-forward with joint decoding can achieve the capacity region of uplink C-RAN to within a constant gap. The gap depends on the network size but is independent of the channel gain matrix and the SNR. We further establish the optimality of Gaussian compression at the relaying BSs for joint decoding, if the input is Gaussian. These results can be further extended to generalized successive decoding under a sum fronthaul constraint and successive decoding for the maximum sum rate. Additionally, under Gaussian signaling, the optimization of quantization noise covariance matrices for weighted sum-rate maximization under joint decoding and for sum rate maximization under practical successive decoding can be cast as convex optimization problems, thereby facilitating their efficient numerical solution. Throughout this section, we focus on the achievable rates under the fixed Gaussian input, and the fixed fronthaul capacity constraints C_ℓ for $\ell = 1, \dots, L$.

A. Achievable Rate Regions Under Gaussian Input and Gaussian Quantization

We let the input distribution be Gaussian, i.e., $\mathbf{X}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_k)$, then evaluate the rate regions for the compress-and-forward scheme with joint decoding and successive decoding under Gaussian quantization, denoted as $\mathcal{R}^G_{JD,GIn}$ and $\mathcal{R}^G_{SD,GIn}$, respectively. Set $\prod_{\ell=1}^{L} p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell) \sim \mathcal{CN}(\mathbf{y}_\ell, \mathbf{Q}_\ell)$, where $\mathbf{Q}_\ell$ is the Gaussian quantization noise covariance matrix at the ℓth BS.

With Gaussian input and Gaussian quantization, we have

$$I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}}) = \log \frac{|\Sigma_\ell + \mathbf{Q}_\ell|}{|\mathbf{Q}_\ell|} \tag{14}$$

and

$$I\left(\mathbf{X}_{\mathcal{T}}; \hat{\mathbf{Y}}_{\mathcal{S}^c}|\mathbf{X}_{\mathcal{T}^c}\right) = \log \frac{\left|\mathbf{H}_{\mathcal{S}^c,\mathcal{T}}\mathbf{K}_{\mathcal{T}}\mathbf{H}^\dagger_{\mathcal{S}^c,\mathcal{T}} + \text{diag}\left(\{\Sigma_\ell + \mathbf{Q}_\ell\}_{\ell\in\mathcal{S}^c}\right)\right|}{\left|\text{diag}\left(\{\Sigma_\ell + \mathbf{Q}_\ell\}_{\ell\in\mathcal{S}^c}\right)\right|}. \tag{15}$$

The achievable rate region (2) for joint decoding can be evaluated as

$$\sum_{k\in\mathcal{T}} R_k < \sum_{\ell\in\mathcal{S}} \left[C_\ell - \log \frac{|\Sigma_\ell + \mathbf{Q}_\ell|}{|\mathbf{Q}_\ell|}\right] + \log \frac{\left|\mathbf{H}_{\mathcal{S}^c,\mathcal{T}}\mathbf{K}_{\mathcal{T}}\mathbf{H}^\dagger_{\mathcal{S}^c,\mathcal{T}} + \text{diag}\left(\{\Sigma_\ell + \mathbf{Q}_\ell\}_{\ell\in\mathcal{S}^c}\right)\right|}{\left|\text{diag}\left(\{\Sigma_\ell + \mathbf{Q}_\ell\}_{\ell\in\mathcal{S}^c}\right)\right|}, \tag{16}$$

for all $\mathcal{T} \subseteq \mathcal{K}$ and $\mathcal{S} \subseteq \mathcal{L}$.

Likewise the achievable rate expression (3) for successive decoding becomes

$$\sum_{k\in\mathcal{T}} R_k < \log \frac{\left|\mathbf{H}_{\mathcal{S}^c,\mathcal{K}}\mathbf{K}_{\mathcal{K}}\mathbf{H}^{\dagger}_{\mathcal{L},\mathcal{K}} + \mathrm{diag}\left(\{\Sigma_\ell+\mathbf{Q}_\ell\}_{\ell\in\mathcal{L}}\right)\right|}{\left|\mathrm{diag}\left(\{\Sigma_\ell+\mathbf{Q}_\ell\}_{\ell\in\mathcal{L}}\right)\right|}, \tag{17}$$

for all $\mathcal{T}\subseteq\mathcal{K}$.

In deriving the fronthaul constraint (4), we start with evaluating the mutual information

$$\begin{aligned}
I&\left(\mathbf{Y}_{\mathcal{S}};\hat{\mathbf{Y}}_{\mathcal{S}}|\hat{\mathbf{Y}}_{\mathcal{S}^c}\right)\\
&= I\left(\mathbf{X}_{\mathcal{K}},\mathbf{Y}_{\mathcal{S}};\hat{\mathbf{Y}}_{\mathcal{S}}|\hat{\mathbf{Y}}_{\mathcal{S}^c}\right) - I\left(\mathbf{X}_{\mathcal{K}};\hat{\mathbf{Y}}_{\mathcal{S}}|\mathbf{Y}_{\mathcal{S}},\hat{\mathbf{Y}}_{\mathcal{S}^c}\right)\\
&= I\left(\mathbf{X}_{\mathcal{K}};\hat{\mathbf{Y}}_{\mathcal{S}}|\hat{\mathbf{Y}}_{\mathcal{S}^c}\right) + I\left(\mathbf{Y}_{\mathcal{S}};\hat{\mathbf{Y}}_{\mathcal{S}}|\mathbf{X}_{\mathcal{K}},\hat{\mathbf{Y}}_{\mathcal{S}^c}\right)\\
&\quad - I\left(\mathbf{X}_{\mathcal{K}};\hat{\mathbf{Y}}_{\mathcal{S}}|\mathbf{Y}_{\mathcal{S}},\hat{\mathbf{Y}}_{\mathcal{S}^c}\right)\\
&\overset{(a)}{=} I\left(\mathbf{X}_{\mathcal{K}};\hat{\mathbf{Y}}_{\mathcal{S}}|\hat{\mathbf{Y}}_{\mathcal{S}^c}\right) + I\left(\mathbf{Y}_{\mathcal{S}};\hat{\mathbf{Y}}_{\mathcal{S}}|\mathbf{X}_{\mathcal{K}}\right)\\
&\overset{(b)}{=} I\left(\mathbf{X}_{\mathcal{K}};\hat{\mathbf{Y}}_{\mathcal{L}}\right) - I\left(\mathbf{X}_{\mathcal{K}};\hat{\mathbf{Y}}_{\mathcal{S}^c}\right) + \sum_{\ell\in\mathcal{S}} I(\mathbf{Y}_\ell;\hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}})
\end{aligned} \tag{18}$$

for all $\mathcal{S}\subseteq\mathcal{L}$, where the equality (a) follows from the fact that

$$I\left(\mathbf{Y}_{\mathcal{S}};\hat{\mathbf{Y}}_{\mathcal{S}}|\mathbf{X}_{\mathcal{K}},\hat{\mathbf{Y}}_{\mathcal{S}^c}\right) = I\left(\mathbf{Y}_{\mathcal{S}};\hat{\mathbf{Y}}_{\mathcal{S}}|\mathbf{X}_{\mathcal{K}}\right) \tag{19}$$

and

$$I\left(\mathbf{X}_{\mathcal{K}};\hat{\mathbf{Y}}_{\mathcal{S}}|\mathbf{Y}_{\mathcal{S}},\hat{\mathbf{Y}}_{\mathcal{S}^c}\right) = 0, \tag{20}$$

and equality (b) follows from the fact that

$$I\left(\mathbf{Y}_{\mathcal{S}};\hat{\mathbf{Y}}_{\mathcal{S}}|\mathbf{X}_{\mathcal{K}}\right) = \sum_{\ell\in\mathcal{S}} I(\mathbf{Y}_\ell;\hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}}). \tag{21}$$

The above equations (19)-(21) follow from the Markov chain

$$\hat{\mathbf{Y}}_i \leftrightarrow \mathbf{Y}_i \leftrightarrow \mathbf{X}_{\mathcal{K}} \leftrightarrow \mathbf{Y}_j \leftrightarrow \hat{\mathbf{Y}}_j, \quad \forall\, i\neq j.$$

We further evaluate the mutual information expression (18) with Gaussian input and Gaussian quantization, which yields that

$$\begin{aligned}
I&\left(\mathbf{Y}_{\mathcal{S}};\hat{\mathbf{Y}}_{\mathcal{S}}|\hat{\mathbf{Y}}_{\mathcal{S}^c}\right)\\
&= \log\frac{\left|\mathbf{H}_{\mathcal{L},\mathcal{K}}\mathbf{K}_{\mathcal{K}}\mathbf{H}^{\dagger}_{\mathcal{L},\mathcal{K}} + \mathrm{diag}\left(\{\Sigma_\ell+\mathbf{Q}_\ell\}_{\ell\in\mathcal{L}}\right)\right|}{\left|\mathrm{diag}\left(\{\Sigma_\ell+\mathbf{Q}_\ell\}_{\ell\in\mathcal{L}}\right)\right|}\\
&\quad - \log\frac{\left|\mathbf{H}_{\mathcal{S}^c,\mathcal{K}}\mathbf{K}_{\mathcal{K}}\mathbf{H}^{\dagger}_{\mathcal{S}^c,\mathcal{K}} + \mathrm{diag}\left(\{\Sigma_\ell+\mathbf{Q}_\ell\}_{\ell\in\mathcal{S}^c}\right)\right|}{\left|\mathrm{diag}\left(\{\Sigma_\ell+\mathbf{Q}_\ell\}_{\ell\in\mathcal{S}^c}\right)\right|}\\
&\quad + \sum_{\ell\in\mathcal{S}}\log\frac{|\Sigma_\ell+\mathbf{Q}_\ell|}{|\mathbf{Q}_\ell|}\\
&= \log\frac{\left|\mathbf{H}_{\mathcal{L},\mathcal{K}}\mathbf{K}_{\mathcal{K}}\mathbf{H}^{\dagger}_{\mathcal{L},\mathcal{K}} + \mathrm{diag}\left(\{\Sigma_\ell+\mathbf{Q}_\ell\}_{\ell\in\mathcal{L}}\right)\right|}{\left|\mathbf{H}_{\mathcal{S}^c,\mathcal{K}}\mathbf{K}_{\mathcal{K}}\mathbf{H}^{\dagger}_{\mathcal{S}^c,\mathcal{K}} + \mathrm{diag}\left(\{\Sigma_\ell+\mathbf{Q}_\ell\}_{\ell\in\mathcal{S}^c}\right)\right|}\\
&\quad - \sum_{\ell\in\mathcal{S}}\log|\mathbf{Q}_\ell|\\
&\le \sum_{\ell\in\mathcal{S}} C_\ell.
\end{aligned}$$

Instead of parameterizing the rate expressions over $\mathbf{Q}_\ell$ as in above, in this section, we introduce the following reparameterization, which is crucial for proving our main results. Define

$$\mathbf{B}_\ell = \left(\Sigma_\ell + \mathbf{Q}_\ell\right)^{-1}. \tag{22}$$

We represent the rate regions of joint decoding and successive decoding in terms of $\mathbf{B}_\ell$ in the following.

Proposition 4: For the uplink C-RAN model shown in Fig. 1 and under fixed Gaussian input $\mathbf{X}_{\mathcal{K}}\sim\mathcal{CN}(\mathbf{0},\mathbf{K}_{\mathcal{K}})$ with $\mathbf{K}_{\mathcal{K}} = \mathrm{diag}\left(\{\mathbf{K}_k\}_{k\in\mathcal{K}}\right)$. The rate-fronthaul region for joint decoding under Gaussian quantization, $\mathcal{P}^G_{JD,GIn}$, is the closure of the convex hull of all $(R_1,\cdots,R_K,C_1,\ldots,C_L)$ satisfying

$$\begin{aligned}
\sum_{k\in\mathcal{T}} R_k &< \sum_{\ell\in\mathcal{S}}\left[C_\ell - \log\frac{\left|\Sigma_\ell^{-1}\right|}{\left|\Sigma_\ell^{-1}-\mathbf{B}_\ell\right|}\right]\\
&\quad + \log\frac{\left|\sum_{\ell\in\mathcal{S}^c}\mathbf{H}^{\dagger}_{\ell,\mathcal{T}}\mathbf{B}_\ell\mathbf{H}_{\ell,\mathcal{T}} + \mathbf{K}_{\mathcal{T}}^{-1}\right|}{\left|\mathbf{K}_{\mathcal{T}}^{-1}\right|}
\end{aligned} \tag{23}$$

for all $\mathcal{T}\subseteq\mathcal{K}$ and $\mathcal{S}\subseteq\mathcal{L}$, for some $0\preceq\mathbf{B}_\ell\preceq\Sigma_\ell^{-1}$, where $\mathbf{K}_{\mathcal{T}} = \mathbb{E}\left[\mathbf{X}_{\mathcal{T}}\mathbf{X}_{\mathcal{T}}^{\dagger}\right]$ is the covariance matrix of $\mathbf{X}_{\mathcal{T}}$, and $\mathbf{H}_{\ell,\mathcal{T}}$ denotes the channel matrix from $\mathbf{X}_{\mathcal{T}}$ to $\mathbf{Y}_\ell$. Furthermore, under the fixed fronthaul capacity constraints C_ℓ for $\ell = 1,\ldots,L$, the rate region achieved by joint decoding $\mathcal{R}^G_{JD,GIn}$ is defined as

$$\begin{aligned}
&\mathcal{R}^G_{JD,GIn}\\
&= \left\{(R_1,\ldots,R_K) : (R_1,\cdots,R_K,C_1,\ldots,C_L)\in\mathcal{P}^G_{JD,GIn}\right\}.
\end{aligned} \tag{24}$$

Proposition 5: For the uplink C-RAN model shown in Fig. 1 and under fixed Gaussian input $\mathbf{X}_{\mathcal{K}}\sim\mathcal{CN}(\mathbf{0},\mathbf{K}_{\mathcal{K}})$ with $\mathbf{K}_{\mathcal{K}} = \mathrm{diag}\left(\{\mathbf{K}_k\}_{k\in\mathcal{K}}\right)$. The rate-fronthaul region for successive decoding, $\mathcal{P}^G_{SD,GIn}$, is the closure of the convex hull of all $(R_1,\cdots,R_K,C_1,\ldots,C_L)$ satisfying

$$\sum_{k\in\mathcal{T}} R_k < \log\frac{\left|\sum_{\ell=1}^{L}\mathbf{H}^{\dagger}_{\ell,\mathcal{T}}\mathbf{B}_\ell\mathbf{H}_{\ell,\mathcal{T}} + \mathbf{K}_{\mathcal{T}}^{-1}\right|}{\left|\mathbf{K}_{\mathcal{T}}^{-1}\right|},\ \forall\,\mathcal{T}\subseteq\mathcal{K}, \tag{25}$$

and

$$\begin{aligned}
\log\frac{\left|\sum_{\ell=1}^{L}\mathbf{H}^{\dagger}_{\ell,\mathcal{K}}\mathbf{B}_\ell\mathbf{H}_{\ell,\mathcal{K}} + \mathbf{K}_{\mathcal{K}}^{-1}\right|}{\left|\sum_{\ell\in\mathcal{S}^c}\mathbf{H}^{\dagger}_{\ell,\mathcal{K}}\mathbf{B}_\ell\mathbf{H}_{\ell,\mathcal{K}} + \mathbf{K}_{\mathcal{K}}^{-1}\right|} + \sum_{\ell\in\mathcal{S}}\log\frac{\left|\Sigma_\ell^{-1}\right|}{\left|\Sigma_\ell^{-1}-\mathbf{B}_\ell\right|}&\\
< \sum_{\ell\in\mathcal{S}} C_\ell,\ \forall\,\mathcal{S}\subseteq\mathcal{L},&
\end{aligned} \tag{26}$$

for some $0\preceq\mathbf{B}_\ell\preceq\Sigma_\ell^{-1}$, where $\mathbf{K}_{\mathcal{T}} = \mathbb{E}\left[\mathbf{X}_{\mathcal{T}}\mathbf{X}_{\mathcal{T}}^{\dagger}\right]$ is the covariance matrix of $\mathbf{X}_{\mathcal{T}}$, and $\mathbf{H}_{\ell,\mathcal{T}}$ denotes the channel matrix from $\mathbf{X}_{\mathcal{T}}$ to $\mathbf{Y}_\ell$. Moreover, under the fixed fronthaul capacity constraints C_ℓ for $\ell = 1,\ldots,L$, the rate region achieved by successive decoding $\mathcal{R}^G_{SD,GIn}$ is defined as

$$\mathcal{R}^G_{SD,GIn} = \left\{(R_1,\ldots,R_K) : (R_1,\cdots,R_K,C_1,\ldots,C_L) \in \mathcal{P}^G_{SD,GIn}\right\}. \tag{27}$$

B. Gaussian Input and Gaussian Quantization Achieve Capacity to within Constant Gap

With Gaussian input and Gaussian quantization, the rate region of joint decoding (23) can be shown to be within a constant gap to the capacity region of uplink C-RAN. This constant-gap result is stated in the following theorem.

Theorem 3: For any rate tuple $(R_1, R_2, \ldots, R_K)$ within the cut-set bound for uplink C-RAN with fixed fronthaul capacities of C_ℓ shown in Fig. 1, the rate tuple $(R_1 - \eta, R_2 - \eta, \ldots, R_K - \eta)$, with $\eta = NL + M$ is achievable for compress-and-forward with Gaussian input, Gaussian quantization, and joint decoding, where L is the number of BSs in the network, M is the number of transmit antennas at user, and N is the number of receive antennas at BS, i.e., $(R_1 - \eta, R_2 - \eta, \ldots, R_K - \eta) \in \mathcal{R}^G_{JD,GIn}$.

Proof: See Appendix D. ■

Although the uplink C-RAN model is an example of a relay network for which noisy network coding approach applies and it is known that compress-and-forward with joint decoding achieves the same rate region as noisy network coding for uplink C-RAN, we remark that Theorem 3 does not immediately follow from the constant-gap optimality result of noisy network coding [9]. The constant-gap optimality of noisy network coding is proven for Gaussian relay networks, whereas the uplink C-RAN model contains fronthaul links which are digital connections and not Gaussian channels.

Combining with our earlier results on the optimality of successive decoding, constant-gap optimality results can also be obtained for compress-and-forward with generalized successive decoding and successive decoding. These results are summarized in the following corollary.

Corollary 1: For the uplink C-RAN model as shown in Fig. 1, compress-and-forward with generalized successive decoding, under Gaussian input and Gaussian quantization achieves the capacity region to within $NL + M$ bits per complex dimension if the fronthaul links are subjected to a sum capacity constraint $\sum_{\ell=1}^{L} C_\ell \leq C$. Furthermore, compress-and-forward with successive decoding, under Gaussian input and Gaussian quantization, achieves the sum capacity of an uplink C-RAN model with individual fronthaul constraints to within $NL + MK$ bits per complex dimension.

C. Optimality of Gaussian Quantization Under Joint Decoding

For the Gaussian uplink MIMO C-RAN model, it is known that Gaussian input and Gaussian quantization are not jointly optimal [5]. However, if the quantization noise is fixed as Gaussian, then the optimal input distribution must be Gaussian. This is because the channel reduces to a conventional Gaussian multiple-access channel in this case. The main result of this section is that the converse is also true, i.e., under fixed Gaussian input, Gaussian quantization actually maximizes the achievable rate region of the uplink C-RAN model under joint decoding.

Under fixed fronthaul capacity constraints C_ℓ for $\ell = 1, \ldots, L$, we let $\mathcal{R}^*_{JD,GIn}$ denote the rate region of joint decoding under Gaussian input and optimal quantization. In the following, we first define Fisher information and state the two main tools for proving this result: the Bruijn identity and the Fisher information inequality. We then present the main theorem on the optimality of Gaussian quantization for joint decoding, i.e., $\mathcal{R}^G_{JD,GIn} = \mathcal{R}^*_{JD,GIn}$.

Definition 1: Let $(\mathbf{X}, \mathbf{Y})$ be a pair of random vectors with joint probability distribution function $p(\mathbf{x}, \mathbf{y})$. The Fisher information matrix of $\mathbf{X}$ is defined as

$$\mathbf{J}(\mathbf{X}) = \mathbb{E}\left[\nabla \log p(\mathbf{X}) \nabla \log p(\mathbf{X})^{\mathsf{T}}\right]. \tag{28}$$

Likewise, the Fisher information matrix of $\mathbf{X}$ conditional on $\mathbf{Y}$ is defined as

$$\mathbf{J}(\mathbf{X}|\mathbf{Y}) = \mathbb{E}\left[\nabla \log p(\mathbf{X}|\mathbf{Y}) \nabla \log p(\mathbf{X}|\mathbf{Y})^{\mathsf{T}}\right]. \tag{29}$$

Lemma 1: (Fisher Information Inequality, [30] [18, Lemma 2]): Let $(\mathbf{U}, \mathbf{X})$ be an arbitrary complex random vector, where the conditional Fisher information of $\mathbf{X}$ conditioned on $\mathbf{U}$ exists. We have

$$\log\left|(\pi e)\mathbf{J}^{-1}(\mathbf{X}|\mathbf{U})\right| \leq h(\mathbf{X}|\mathbf{U}). \tag{30}$$

Lemma 2 (Bruijn Identity, [31] [18, Lemma 3]): Let $(\mathbf{V}_1, \mathbf{V}_2)$ be an arbitrary random vector with finite second moments, and $\mathbf{N}$ be a zero-mean Gaussian random vector with covariance Λ_N. Assume $(\mathbf{V}_1, \mathbf{V}_2)$ and $\mathbf{N}$ are independent. We have

$$\operatorname{cov}(\mathbf{V}_2|\mathbf{V}_1, \mathbf{V}_2 + \mathbf{N}) = \Lambda_N - \Lambda_N \mathbf{J}(\mathbf{V}_2 + \mathbf{N}|\mathbf{V}_1)\Lambda_N. \tag{31}$$

Theorem 4: For the uplink C-RAN under fixed Gaussian input distribution and assuming joint decoding, Gaussian quantization is optimal, i.e., $\mathcal{R}^G_{JD,GIn} = \mathcal{R}^*_{JD,GIn}$.

Proof: Recall that the achievable rate region of the compress-and-forward scheme under joint decoding is given by the set of $(R_1, \ldots, R_K)$ derived from (2) under the joint distribution

$$p\left(\mathbf{x}_1, \ldots, \mathbf{x}_K, \mathbf{y}_1, \ldots, \mathbf{y}_L, \hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_L\right) = \prod_{k=1}^{K} p(\mathbf{x}_k) \prod_{\ell=1}^{L} p(\mathbf{y}_\ell|\mathbf{x}_1, \ldots, \mathbf{x}_K) \prod_{\ell=1}^{L} p\left(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell\right). \tag{32}$$

For fixed Gaussian input $\mathbf{X}_{\mathcal{K}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_{\mathcal{K}})$ and fixed $\prod_{\ell=1}^{L} p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell)$, choose $\mathbf{B}_\ell$ with $\mathbf{0} \preceq \mathbf{B}_\ell \preceq \Sigma_\ell^{-1}$ such that

$$\operatorname{cov}\left(\mathbf{Y}_\ell|\mathbf{X}_{\mathcal{K}}, \hat{\mathbf{Y}}_\ell\right) = \Sigma_\ell - \Sigma_\ell \mathbf{B}_\ell \Sigma_\ell, \quad \ell = 1, \cdots, L.$$

We proceed to show that the achievable rate region as given by (23) with a Gaussian $\prod_{\ell=1}^{L} p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell) \sim \mathcal{CN}(\mathbf{Y}_\ell, \mathbf{Q}_\ell)$, where $\mathbf{Q}_\ell = \mathbf{B}_\ell^{-1} - \Sigma_\ell$, is as large as that of (2) under Gaussian input.

First, note that

$$\begin{aligned} & I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell | \mathbf{X}_\mathcal{K}\right) \\ &\quad = \log |(\pi e)\Sigma_\ell| - h\left(\mathbf{Y}_\ell | \mathbf{X}_\mathcal{K}, \hat{\mathbf{Y}}_\ell\right) \\ &\quad \geq \log |(\pi e)\Sigma_\ell| - \log \left|(\pi e) \operatorname{cov}\left(\mathbf{Y}_\ell | \mathbf{X}_\mathcal{K}, \hat{\mathbf{Y}}_\ell\right)\right| \\ &\quad = \log \frac{\left|\Sigma_\ell^{-1}\right|}{\left|\Sigma_\ell^{-1} - \mathbf{B}_\ell\right|}, \quad \ell = 1, \cdots, L, \end{aligned} \tag{33}$$

where we use the fact that Gaussian distribution maximizes differential entropy.

Moreover, we have

$$\begin{aligned} I\left(\mathbf{X}_\mathcal{T}; \hat{\mathbf{Y}}_{\mathcal{S}^c} | \mathbf{X}_{\mathcal{T}^c}\right) &= h\left(\mathbf{X}_\mathcal{T}\right) - h\left(\mathbf{X}_\mathcal{T} | \mathbf{X}_{\mathcal{T}^c}, \hat{\mathbf{Y}}_{\mathcal{S}^c}\right) \\ &\leq \log |\mathbf{K}_\mathcal{T}| - \log \left|\mathbf{J}^{-1}\left(\mathbf{X}_\mathcal{T} | \mathbf{X}_{\mathcal{T}^c}, \hat{\mathbf{Y}}_{\mathcal{S}^c}\right)\right|, \end{aligned}$$

where the inequality is due to Lemma 1. Since

$$\mathbf{Y}_{\mathcal{S}^c} = \mathbf{H}_{\mathcal{S}^c,\mathcal{T}}\mathbf{X}_\mathcal{T} + \mathbf{H}_{\mathcal{S}^c,\mathcal{T}^c}\mathbf{X}_{\mathcal{T}^c} + \mathbf{Z}_{\mathcal{S}^c},$$

it follows from the MMSE estimation of Gaussian random vectors that

$$\begin{aligned} \mathbf{X}_\mathcal{T} &= \mathbb{E}\left[\mathbf{X}_\mathcal{T} | \mathbf{X}_{\mathcal{T}^c}, \mathbf{Y}_{\mathcal{S}^c}\right] + \mathbf{N}_{\mathcal{T},\mathcal{S}^c} \\ &= \sum_{\ell \in \mathcal{S}^c} \mathbf{G}_{\mathcal{T},\ell}\left(\mathbf{Y}_\ell - \mathbf{H}_{\ell,\mathcal{T}^c}\mathbf{X}_{\mathcal{T}^c}\right) + \mathbf{N}_{\mathcal{T},\mathcal{S}^c}, \end{aligned}$$

where

$$\mathbf{G}_{\mathcal{T},\ell} = \left(\mathbf{K}_\mathcal{T}^{-1} + \sum_{j \in \mathcal{S}^c} \mathbf{H}_{j,\mathcal{T}}^\dagger \Sigma_j^{-1} \mathbf{H}_{j,\mathcal{T}}\right)^{-1} \mathbf{H}_{\ell,\mathcal{T}}^\dagger \Sigma_\ell^{-1},$$

and $\mathbf{N}_{\mathcal{T},\mathcal{S}^c} \sim \mathcal{CN}\left(\mathbf{0}, \Lambda_\mathbf{N}\right)$ with covariance matrix

$$\Lambda_\mathbf{N} = \left(\mathbf{K}_\mathcal{T}^{-1} + \sum_{\ell \in \mathcal{S}^c} \mathbf{H}_{\ell,\mathcal{T}}^\dagger \Sigma_\ell^{-1} \mathbf{H}_{\ell,\mathcal{T}}\right)^{-1}. \tag{34}$$

Here $\mathbb{E}\left[\mathbf{X}_\mathcal{T} | \mathbf{X}_{\mathcal{T}^c}, \mathbf{Y}_{\mathcal{S}^c}\right]$ is the MMSE estimator of $\mathbf{X}_\mathcal{T}$ from $\mathbf{X}_{\mathcal{T}^c}, \mathbf{Y}_{\mathcal{S}^c}$. The error in estimation is $\mathbf{N}_{\mathcal{T},\mathcal{S}^c}$, and the MMSE matrix is $\Lambda_\mathbf{N}$.

By the matrix complementary identity between Fisher information matrix and MMSE in Lemma 2, we have

$$\begin{aligned} & \mathbf{J}\left(\mathbf{X}_\mathcal{T} | \mathbf{X}_{\mathcal{T}^c}, \hat{\mathbf{Y}}_{\mathcal{S}^c}\right) \\ &= \Lambda_\mathbf{N}^{-1} \\ &\quad - \Lambda_\mathbf{N}^{-1} \operatorname{cov}\left(\sum_{\ell \in \mathcal{S}^c} \mathbf{G}_{\mathcal{T},\ell}(\mathbf{Y}_\ell - \mathbf{H}_{\ell,\mathcal{T}^c}\mathbf{X}_{\mathcal{T}^c}) | \mathbf{X}_\mathcal{K}, \hat{\mathbf{Y}}_{\mathcal{S}^c}\right) \Lambda_\mathbf{N}^{-1} \\ &= \Lambda_\mathbf{N}^{-1} - \Lambda_\mathbf{N}^{-1} \operatorname{cov}\left(\sum_{\ell \in \mathcal{S}^c} \mathbf{G}_{\mathcal{T},\ell}\mathbf{Y}_\ell | \mathbf{X}_\mathcal{K}, \hat{\mathbf{Y}}_{\mathcal{S}^c}\right) \Lambda_\mathbf{N}^{-1} \\ &= \Lambda_\mathbf{N}^{-1} - \Lambda_\mathbf{N}^{-1} \left[\sum_{\ell \in \mathcal{S}^c} \mathbf{G}_{\mathcal{T},\ell} \operatorname{cov}\left(\mathbf{Y}_\ell | \mathbf{X}_\mathcal{K}, \hat{\mathbf{Y}}_\ell\right) \mathbf{G}_{\mathcal{T},\ell}^\dagger\right] \Lambda_\mathbf{N}^{-1} \\ &= \Lambda_\mathbf{N}^{-1} - \sum_{\ell \in \mathcal{S}^c} \mathbf{H}_{\ell,\mathcal{T}}^\dagger \left(\Sigma_\ell^{-1} - \mathbf{B}_\ell\right) \mathbf{H}_{\ell,\mathcal{T}} \\ &= \mathbf{K}_\mathcal{T}^{-1} + \sum_{\ell \in \mathcal{S}^c} \mathbf{H}_{\ell,\mathcal{T}}^\dagger \mathbf{B}_\ell \mathbf{H}_{\ell,\mathcal{T}}. \end{aligned}$$

Therefore,

$$\begin{aligned} I\left(\mathbf{X}_\mathcal{T}; \hat{\mathbf{Y}}_{\mathcal{S}^c} | \mathbf{X}_{\mathcal{T}^c}\right) &\leq \log \frac{\left|\mathbf{J}(\mathbf{X}_\mathcal{T} | \mathbf{X}_{\mathcal{T}^c}, \hat{\mathbf{Y}}_{\mathcal{S}^c})\right|}{\left|\mathbf{K}_\mathcal{T}^{-1}\right|} \\ &= \log \frac{\left|\mathbf{K}_\mathcal{T}^{-1} + \sum_{\ell \in \mathcal{S}^c} \mathbf{H}_{\ell,\mathcal{T}}^\dagger \mathbf{B}_\ell \mathbf{H}_{\ell,\mathcal{T}}\right|}{\left|\mathbf{K}_\mathcal{T}^{-1}\right|} \end{aligned} \tag{35}$$

for all $\mathcal{T} \subseteq \mathcal{K}$ and $\mathcal{S} \subseteq \mathcal{L}$. Combining (33) and (35), we conclude that $\mathcal{R}_{JD,GIn}^G$ as derived from (23) is as large as $\mathcal{R}_{JD,GIn}^*$. Therefore, $\mathcal{R}_{JD,GIn}^G = \mathcal{R}_{JD,GIn}^*$. ■

D. Optimization of Gaussian Input and Gaussian Quantization Noise Covariance Matrices

This section addresses the numerical optimization of the Gaussian input and quantization noise covariance matrices for uplink MIMO C-RAN under given fronthaul capacity constraints. First, we note that even when restricting to Gaussian input and Gaussian quantization, the joint optimization of input and quantization noise covariance matrices is still a challenging problem for the uplink MIMO C-RAN. However, if we fix the quantization noise covariance, then the input optimization reduces to that of optimizing a conventional Gaussian multiple-access channel. In particular, the problem of maximizing the weighted sum rate can be formulated as a convex optimization, which can be readily solved [32].

Conversely, if we fix the transmit covariance matrix, the optimization of quantization noise covariance can in some cases be formulated as convex optimization. The key enabling fact is the reparameterization in term of $\mathbf{B}_\ell$ (22), instead of direct optimization over $\mathbf{Q}_\ell$. Consider first the case of joint decoding. Using (23) under the fixed C_ℓ for $\ell = 1, \ldots, L$, the weighted sum rate maximization problem can be formulated over $\{R_k, \mathbf{B}_\ell\}$ as follows:

$$\begin{aligned} \max_{R_k, \mathbf{B}_\ell} \quad & \sum_{k=1}^{K} \mu_k R_k \\ \text{s.t.} \quad & \sum_{k \in \mathcal{T}} R_k \leq \log \frac{\left|\sum_{\ell \in \mathcal{S}^c} \mathbf{H}_{\ell,\mathcal{T}}^\dagger \mathbf{B}_\ell \mathbf{H}_{\ell,\mathcal{T}} + \mathbf{K}_\mathcal{T}^{-1}\right|}{\left|\mathbf{K}_\mathcal{T}^{-1}\right|} \\ & \quad + \sum_{\ell \in \mathcal{S}} \left[C_\ell - \log \frac{\left|\Sigma_\ell^{-1}\right|}{\left|\Sigma_\ell^{-1} - \mathbf{B}_\ell\right|}\right], \ \forall\, \mathcal{T} \subseteq \mathcal{K},\ \forall \mathcal{S} \subseteq \mathcal{L}, \\ & \mathbf{0} \preceq \mathbf{B}_\ell \preceq \Sigma_\ell^{-1}, \quad \forall\, \ell \in \mathcal{L}, \end{aligned} \tag{36}$$

where μ_k represents the weight associated with user k, which is typically determined from upper layer protocols. The key observation is that the above problem is convex in $\{R_k, \mathbf{B}_\ell\}$. However, we also note that because of joint decoding, the number of constraints is exponential in the size of the network. Consequently, the above optimization problem can only be solved for small networks in practice.

Note that the above formulation considers the optimization of instantaneous achievable rates R_k under instantaneous fronthaul capacity constraints C_ℓ in a fixed time slot. The solution obtained, however, also applies to the more general case of optimizing the weighted sum rates under weighted sum fronthaul constraint (e.g., $\sum_{\ell=1}^{L} \nu_\ell C_\ell \leq C$). This is because if we consider a slightly more general formulation of optimizing an objective of

$$\max_{R_k, \mathbf{B}_\ell, C_\ell} \sum_{k=1}^{K} \mu_k R_k - \gamma \sum_{\ell=1}^{L} \nu_\ell C_\ell \tag{37}$$

under the same constraints as in (36) and $\sum_{\ell=1}^{L} \nu_\ell C_\ell \leq C$. Such an optimization problem is convex, so time-sharing is not needed. For this reason, the rest of this section considers the formulation with instantaneous rates only.

We now consider the weighted sum-rate maximization problem for the case of successive decoding of the quantization codewords followed by the user messages. However, the direct characterization of successive decoding rate (25)-(26) does not give rise to a convex formulation. Nevertheless, for the special case of maximizing the sum rate (i.e., with $\mu_1 = \cdots = \mu_K = 1$), using Theorem 2, which shows that successive decoding achieves the same maximum sum rate as joint decoding, the sum-rate maximization problem with successive decoding can be equivalently formulated as follows:

Theorem 5: For the uplink C-RAN model with individual fronthaul capacity constraint C_ℓ as shown in Fig. 1, the sum rate maximization problem under successive decoding can be formulated as the following convex problem:

$$\begin{aligned} \max_{R, \mathbf{B}_\ell} \ & R \\ \text{s.t.} \ & R \leq \sum_{\ell \in \mathcal{S}} \left[C_\ell - \log \frac{\left|\Sigma_\ell^{-1}\right|}{\left|\Sigma_\ell^{-1} - \mathbf{B}_\ell\right|} \right] \\ & \qquad + \log \frac{\left|\sum_{\ell \in \mathcal{S}^c} \mathbf{H}_{\ell,\mathcal{T}}^{\dagger} \mathbf{B}_\ell \mathbf{H}_{\ell,\mathcal{T}} + \mathbf{K}_{\mathcal{K}}^{-1}\right|}{\left|\mathbf{K}_{\mathcal{K}}^{-1}\right|}, \ \forall \mathcal{S} \subseteq \mathcal{L}, \\ & \mathbf{0} \preceq \mathbf{B}_\ell \preceq \Sigma_\ell^{-1}, \ \forall \ \ell \in \mathcal{L}. \end{aligned} \tag{38}$$

Further, if the fronthaul links are subject to a sum capacity constraint of C, the sum rate maximization problem can be formulated as the following convex problem:

$$\begin{aligned} \max_{R, \mathbf{B}_\ell} \ & R \\ \text{s.t.} \ & R \leq \log \frac{\left|\sum_{\ell=1}^{L} \mathbf{H}_{\ell,\mathcal{K}}^{\dagger} \mathbf{B}_\ell \mathbf{H}_{\ell,\mathcal{K}} + \mathbf{K}_{\mathcal{K}}^{-1}\right|}{\left|\mathbf{K}_{\mathcal{K}}^{-1}\right|}, \\ & R + \sum_{\ell=1}^{L} \log \frac{\left|\Sigma_\ell^{-1}\right|}{\left|\Sigma_\ell^{-1} - \mathbf{B}_\ell\right|} \leq C, \\ & \mathbf{0} \preceq \mathbf{B}_\ell \preceq \Sigma_\ell^{-1}, \ \forall \ \ell \in \mathcal{L}. \end{aligned} \tag{39}$$

We remark that the formulation for uplink C-RAN with individual fronthaul capacities (38) has exponential number of constraints, because the CP in effect needs to search over $L!$ different decoding orders of quantization codewords at the BSs. In practical implementation, a heuristic method can be used to determine the decoding orders of quantization codewords for avoiding the exponential search [24], [33]. Alternatively, if the C-RAN has a sum fronthaul constraint, then the number of constraints is linear in network size, because we only need to consider the case of $\mathcal{S} = \mathcal{L}$ and $\mathcal{S} = \emptyset$ in (38). Consequently, the resulting quantization noise covariance optimization problem (39) can be solved in polynomial time. Note that convexity is a key advantage of the above problem formulations as compared to previous approaches in the literature (e.g. [21], [22]) that parameterize the optimization problem over the quantization noise covariance $\mathbf{Q}_\ell$, which leads to a nonconvex formulation.

We emphasize the importance of Gaussian input for the convex formulation in Theorem 5. Suppose that both input signal $\mathbf{X}_{\mathcal{K}}$ and compressed signal $\hat{\mathbf{Y}}_\ell$ are discrete random vectors with finite alphabet. For fixed input distribution, the sum-rate maximization problem under the sum fronthaul constraint can be written as

$$\begin{aligned} \max_{p(\hat{\mathbf{y}}_\ell | \mathbf{y}_\ell)} \ & I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{L}}\right), \\ \text{s.t.} \ & I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}\right) \leq C, \\ & p\left(\hat{\mathbf{y}}_\ell | \mathbf{y}_\ell\right) \geq 0, \ \sum_{\hat{\mathbf{y}}_\ell} p\left(\hat{\mathbf{y}}_\ell | \mathbf{y}_\ell\right) = 1, \ \forall \ \ell \in \mathcal{L}. \end{aligned} \tag{40}$$

The above problem can be thought as a variant of the information bottleneck method [19], which can be solved by a generalized Blahut-Arimoto (BA) algorithm [34], [35]. However, due to the non-convex nature of problem (40), the generalized BA algorithm can only converge to a local optimum.

V. Conclusion

This paper provides a number of information theoretical results on the optimal compress-and-forward scheme for the uplink MIMO C-RAN model, where the BSs are connected to a CP through noiseless fronthaul links of limited capacities. It is shown that the generalized successive decoding scheme, which allows arbitrary decoding orders between quantization and user message codewords, can achieve the same rate region as joint decoding under a sum fronthaul constraint. Moreover, the practical successive decoding of the quantization codewords followed by the user messages is shown to achieve the same maximum sum rate as joint decoding under individual fronthaul constraints. In addition, if the input distribution is assumed to be Gaussian, it is shown that Gaussian quantization maximizes the achievable rate region of joint decoding. With Gaussian input signaling, the optimization of Gaussian quantization for maximizing the weighted sum rate under joint decoding and the sum rate under successive decoding can be cast as convex optimization problems, which facilitates efficient numerical solution. Finally, Gaussian input and Gaussian quantization achieve the capacity region of the uplink C-RAN model to within constant gap. Collectively, these results provide justifications for the practical choice of using Gaussian-like input signals at the user terminals, Gaussian-like quantization at the relaying BSs, and successive decoding of quantization codewords followed by user messages at the CP for implementing uplink MIMO C-RAN.

VI. Appendix A
Optimality of Generalized Successive Decoding

In this appendix, we prove Theorem 1, which states the equivalence between generalized successive decoding and joint decoding under a sum-capacity fronthaul constraint. We begin by introducing an outer bound for the achievable rate region of joint decoding under a sum fronthaul constraint. Under the sum fronthaul capacity constraint, define the rate-fronthaul region for joint decoding $\mathcal{P}^o_{JD,s}$ as the closure of the convex hull of all $(R_1, R_2, \ldots, R_K, C)$ satisfying

$$\begin{cases} \sum_{k\in\mathcal{T}} R_k < \min\left\{C - \sum_{\ell\in\mathcal{L}} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_\mathcal{K}\right),\right. \\ \qquad\qquad \left. I\left(\mathbf{X}_\mathcal{T}; \hat{\mathbf{Y}}_\mathcal{L}|\mathbf{X}_{\mathcal{T}^c}\right)\right\}, \quad \forall\, \mathcal{T} \subseteq \mathcal{K}, \\ C > \sum_{\ell\in\mathcal{L}} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_\mathcal{K}\right), \end{cases} \tag{41}$$

for some product distribution $\prod_{k=1}^K p\left(\mathbf{x}_k\right) \prod_{\ell=1}^L p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell)$. Under fixed sum fronthaul constraint C, define the region $\mathcal{R}^o_{JD,s}$ as follows

$$\mathcal{R}^o_{JD,s} = \left\{(R_1, \ldots, R_K) : (R_1, \cdots, R_K, C) \in \mathcal{P}^o_{JD,s}\right\}. \tag{42}$$

Note that the rate region $\mathcal{R}^o_{JD,s}$ is an outer bound for joint decoding rate region (10) because only the constraints corresponding to $\mathcal{S} = \emptyset$ and $\mathcal{S} = \mathcal{L}$ are included. These constraints turn out to be the only active ones under the sum fronthaul constraint $\sum_{\ell=1}^L C_\ell \le C$ and $C_\ell \ge 0$.

Under the sum fronthaul constraint, the generalized successive decoding region $\mathcal{P}_{GSD,s}(\pi)$ for decoding order π can be derived from (2) by letting $\sum_{\ell=1}^L C_\ell = C$. More specifically, $\mathcal{P}_{GSD,s}(\pi)$ is the closure of the convex hull of all $(R_1, R_2, \ldots, R_K, C)$ satisfying

$$\begin{cases} R_k < I\left(\mathbf{X}_k; \hat{\mathbf{Y}}_{\mathcal{J}_{\mathbf{X}_k}}|\mathbf{X}_{\mathcal{I}_{\mathbf{X}_k}}\right), \quad \forall\, k \in \mathcal{K}, \\ C > \sum_{\ell=1}^L I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\hat{\mathbf{Y}}_{\mathcal{J}_{\mathbf{Y}_\ell}}, \mathbf{X}_{\mathcal{I}_{\mathbf{Y}_\ell}}\right), \end{cases} \tag{43}$$

for some product distribution $\prod_{k=1}^K p\left(\mathbf{x}_k\right) \prod_{\ell=1}^L p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell)$, where $\mathcal{I}_{\mathbf{X}_k}$, $\mathcal{I}_{\mathbf{Y}_\ell}$ are the indices of user messages that are decoded before $\mathbf{X}_k$ and $\mathbf{Y}_\ell$ under the permutation π, and $\mathcal{J}_{\mathbf{X}_k}$, $\mathcal{J}_{\mathbf{Y}_\ell}$ are the indices of the quantization codewords that are decoded before $\mathbf{X}_k$ and $\mathbf{Y}_\ell$ under decoding order π. Define $\mathcal{P}^*_{GSD,s}$ to be the closure of the convex hull of all $\mathcal{P}_{GSD,s}(\pi)$'s over decoding order π's, i.e., $\mathcal{P}^*_{GSD,s} = \text{co}\left(\bigcup_\pi \mathcal{P}_{GSD,s}(\pi)\right)$.

We say a point $(R_1, \ldots, R_K, C)$ is *dominated* by a point in $\mathcal{P}^*_{GSD,S}$ if there exists some $(R'_1, \ldots, R'_K, C')$ in $\mathcal{P}^*_{GSD,s}$ for which $R_k \le R'_k$ for $k = 1, 2, \ldots, K$, and $C \ge C'$.

Given the definitions of $\mathcal{R}^*_{GSD,s}$, $\mathcal{R}^*_{JD,s}$ and $\mathcal{R}^o_{JD,s}$, it is easy to see that $\mathcal{R}^*_{GSD,s} \subseteq \mathcal{R}^*_{JD,s} \subseteq \mathcal{R}^o_{JD,s}$. To show $\mathcal{R}^*_{GSD,s} = \mathcal{R}^*_{JD,s}$, it suffices to show $\mathcal{R}^o_{JD,s} \subseteq \mathcal{R}^*_{GSD,s}$, which is equivalent to showing that if a point $(R_1, R_2, \ldots, R_K, C) \in \mathcal{P}^o_{JD,s}$, then the same point $(R_1, R_2, \ldots, R_K, C) \in \mathcal{P}^*_{GSD,s}$ also. To show this, it suffices to show that for any fixed product distribution $\prod_{k=1}^K p\left(\mathbf{x}_k\right) \prod_{\ell=1}^L p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell)$ and fixed C, each extreme point $(R_1, \ldots, R_K, C)$ as defined by (41) is dominated by a point in $\mathcal{P}^*_{GSD,s}$ with the average sum fronthaul capacity requirement at most C.

To this end, define a set function $f : 2^\mathcal{K} \to \mathbb{R}$ as follows:

$$f\left(\mathcal{T}\right) := \min\left\{C - \sum_{\ell\in\mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_\mathcal{K}),\ I\left(\mathbf{X}_\mathcal{T}; \hat{\mathbf{Y}}_\mathcal{L}|\mathbf{X}_{\mathcal{T}^c}\right)\right\},$$

for each $\mathcal{T} \subseteq \mathcal{K}$. It can be verified that the function f is a submodular function (Appendix B, Lemma 3). By construction, $(R_1, R_2, \ldots, R_K)$ as defined by (42) satisfies

$$\sum_{k\in\mathcal{T}} R_k \le f\left(\mathcal{T}\right),$$

which is a submodular polyhedron associated with f.

It follows by basic results in submodular optimization (Appendix VI, Proposition 6) that, for a linear ordering $i_1 \prec i_2 \prec \cdots \prec i_K$ on the set $\mathcal{K}$, an extreme point of $\mathcal{R}^*_{JD,s}$ can be computed as follows

$$\tilde{R}_{i_j} = f\left(\{i_1, \ldots, i_j\}\right) - f\left(\{i_1, \ldots, i_{j-1}\}\right).$$

Furthermore, the extreme points of $\mathcal{R}^o_{JD,s}$ can be enumerated over all the orderings of $\mathcal{K}$. Each ordering of $\mathcal{K}$ is analyzed in the same manner, hence for notational simplicity we only consider the natural ordering $i_j = j$ in the following proof.

By construction,

$$\begin{aligned} \tilde{R}_j = \min&\left\{C - \sum_{\ell\in\mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_\mathcal{K}),\ I\left(\mathbf{X}_1^j; \hat{\mathbf{Y}}_\mathcal{L}|\mathbf{X}_{j+1}^K\right)\right\} \\ &- \min\left\{C - \sum_{\ell\in\mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_\mathcal{K}),\ I\left(\mathbf{X}_1^{j-1}; \hat{\mathbf{Y}}_\mathcal{L}|\mathbf{X}_j^K\right)\right\}. \end{aligned} \tag{44}$$

Due to the fact that $I\left(\mathbf{X}_1^j; \hat{\mathbf{Y}}_\mathcal{L}|\mathbf{X}_{j+1}^K\right) \ge I\left(\mathbf{X}_1^{j-1}; \hat{\mathbf{Y}}_\mathcal{L}|\mathbf{X}_j^K\right)$, for some product distribution $\prod_{k=1}^K p\left(\mathbf{x}_k\right) \prod_{\ell=1}^L p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell)$, equation (44) can yield two different results. Case 1: the first term $C - \sum_{\ell\in\mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_\mathcal{K})$ in the minima in equation (44) is not active for any j; Case 2: the term $C - \sum_{\ell\in\mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_\mathcal{K})$ is active starting with some index j.

- Case 1 holds if $C \ge I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}_\mathcal{L}\right) + \sum_{\ell\in\mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_\mathcal{K})$. In this case the resulting extreme point $\mathbf{r}^1_{JD} = (\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_K, C)$ satisfies

$$\begin{cases} \tilde{R}_j = I\left(\mathbf{X}_j; \hat{\mathbf{Y}}_\mathcal{L}|\mathbf{X}_{j+1}^K\right), \text{ for } j = 1, 2, \ldots, K-1, \\ \tilde{R}_K = I\left(\mathbf{X}_K; \hat{\mathbf{Y}}_\mathcal{L}\right), \\ C = I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}_\mathcal{L}\right) + \sum_{\ell\in\mathcal{L}} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_\mathcal{K}\right). \end{cases}$$

  Consider successive decoding with the decoding order $\hat{\mathbf{Y}}_\mathcal{L} \to \mathbf{X}_K \to \cdots \to \mathbf{X}_1$. The extreme point $(R_1^*, \ldots, R_K^*, C^*) \in \mathcal{P}^*_{GSD,s}$ corresponding to this decoding order is

$$\begin{cases} \tilde{R}^*_j = I\left(\mathbf{X}_j; \hat{\mathbf{Y}}_\mathcal{L}|\mathbf{X}_{j+1}^K\right), \text{ for } j = 1, 2, \ldots, K-1, \\ \tilde{R}^*_K = I\left(\mathbf{X}_K; \hat{\mathbf{Y}}_\mathcal{L}\right), \\ C^* = I(\mathbf{Y}_\mathcal{L}; \hat{\mathbf{Y}}_\mathcal{L}). \end{cases}$$

Following the Markov chain

$$\hat{\mathbf{Y}}_i \leftrightarrow \mathbf{Y}_i \leftrightarrow \mathbf{X}_{\mathcal{K}} \leftrightarrow \mathbf{Y}_j \leftrightarrow \hat{\mathbf{Y}}_j, \quad \forall\, i \neq j,$$

it can be shown that

$$\sum_{\ell\in\mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}}) + I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{L}}\right) = I(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}).$$

Clearly, $\mathbf{r}^1_{JD}$ can be achieved by the decoding order of $\hat{\mathbf{Y}}_{\mathcal{L}} \to \mathbf{X}_K \to \cdots \to \mathbf{X}_1$. Thus, $\mathbf{r}^1_{JD}$ is dominated by a point in $\mathcal{P}^*_{GSD,s}$.

- Case 2 holds if $C \leq I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{L}}\right) + \sum_{\ell\in\mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}})$. We let $\mathbf{X}^i_j = \emptyset$ for $i < j$, and assume that

$$I\left(\mathbf{X}_1^{j-1}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_j^K\right) \leq C - \sum_{\ell\in\mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}})$$

and

$$C - \sum_{\ell\in\mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}}) \leq I\left(\mathbf{X}_1^j; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right)$$

for some $1 \leq j \leq K$. The resulting extreme point $\mathbf{r}^2_{JD} = (\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_K, C)$ satisfies

$$\begin{cases} \tilde{R}_i = I\left(\mathbf{X}_i; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{i+1}^K\right), & \text{for } i < j, \\ \tilde{R}_i = \left[C - \sum_{\ell\in\mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}}) - I\left(\mathbf{X}_1^{j-1}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_i^K\right)\right]^+, & \\ & \text{for } i = j, \\ \tilde{R}_i = 0, & \text{for } i > j, \\ C = I\left(\mathbf{X}_1^j; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right) + \sum_{\ell\in\mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}}), & \end{cases}$$

where $[\cdot]^+$ means $\max\{\cdot, 0\}$. Note that users with index $i > j$ are inactive, and are essentially removed from the network. In this case, the rate-fronthaul tuple does not correspond to a specific corner point obtained with a specific generalized successive decoding order, but that it lies on the convex-hull of two corner points of two different generalized successive decoding orders. To obtain a visualization on Case 2, the rate-fronthaul region for a two-user C-RAN model under a fixed joint distribution $p\left(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1, \mathbf{y}_2, \hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2\right)$ is illustrated in Fig. 2. In the case of $K = j = 2$, it is shown that the rate-fronthaul tuple $\mathbf{r}^2_{JD}$ lies on the convex-hull of two corner points $\mathbf{r}^{(1)}$ and $\mathbf{r}^{(2)}$.

To prove the statement mathematically, we consider generalized successive decoding with the following two different decoding orders: (i) Decoding order 1 satisfies

$$\mathbf{X}_K \to \ldots \to \mathbf{X}_{j+1} \to \hat{\mathbf{Y}}_{\mathcal{L}} \to \mathbf{X}_j \to \ldots \to \mathbf{X}_1.$$

The extreme point $\mathbf{r}^{(1)}_{GSD} = (R_1^{(1)}, \ldots, R_K^{(1)}, C^{(1)})$ of $\mathcal{P}^*_{GSD,s}$ corresponding to Decoding order 1 satisfies

$$\begin{cases} R_i^{(1)} = I\left(\mathbf{X}_i; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{i+1}^K\right), & \text{for } i \leq j, \\ R_i^{(1)} = 0, & \text{for } i > j, \\ C^{(1)} = I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right), & \end{cases}$$

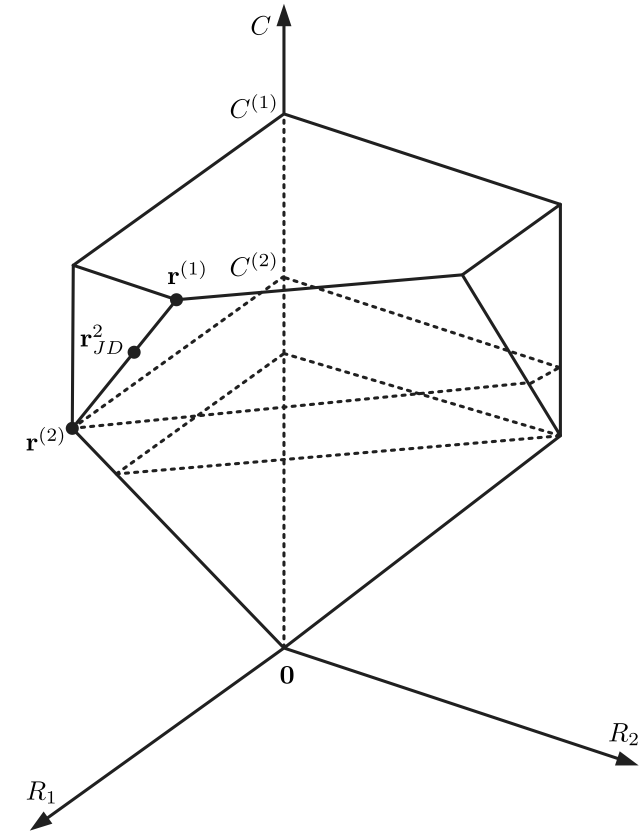

Fig. 2. An illustration of the rate-fronthaul tuple in Case 2 in Appendix A with a two-user C-RAN model under a fixed joint distribution $p\left(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1, \mathbf{y}_2, \hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2\right)$.

where $C^{(1)}$ represents the required fronthaul capacity in order to achieve the above rate tuple $(R_1^{(1)}, \ldots, R_K^{(1)})$ with decoding order 1.

(ii) Decoding order 2 is

$$\mathbf{X}_K \to \ldots \to \mathbf{X}_j \to \hat{\mathbf{Y}}_{\mathcal{L}} \to \mathbf{X}_{j-1} \to \ldots \to \mathbf{X}_1.$$

The extreme point $\mathbf{r}^{(2)}_{GSD} = (R_1^{(2)}, \ldots, R_K^{(2)}, C^{(2)})$ of $\mathcal{P}^*_{GSD,s}$ corresponding to Decoding order 2 satisfies

$$\begin{cases} R_i^{(2)} = I\left(\mathbf{X}_i; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{i+1}^K\right), & \text{for } i < j, \\ R_i^{(2)} = 0, & \text{for } i \geq j, \\ C^{(2)} = I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_j^K\right), & \end{cases}$$

where $C^{(2)}$ represents the required fronthaul capacity in order to achieve the above rate tuple $(R_1^{(2)}, \ldots, R_K^{(2)})$ with decoding order 2. Observe that the rate tuples $(R_1^{(1)}, \ldots, R_K^{(1)})$ and $(R_1^{(2)}, \ldots, R_K^{(2)})$ given by above two decoding orders different at only the jth component, where $R_j^{(1)} = I\left(\mathbf{X}_j; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right)$ and $R_j^{(2)} = 0$ and $R_i^{(1)} = R_i^{(2)} = \tilde{R}_i$ for all $i < j$. Now choose a parameter θ such that

$$\theta = \frac{C - \sum_{\ell\in\mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}}) - I\left(\mathbf{X}_1^{j-1}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_j^K\right)}{I\left(\mathbf{X}_j; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right)}. \tag{45}$$

Following the Markov chain $\mathbf{X}_{\mathcal{K}} \leftrightarrow \mathbf{Y}_{\mathcal{L}} \leftrightarrow \hat{\mathbf{Y}}_{\mathcal{L}}$, we have the following identity,

$$1-\theta = \frac{I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right) - C}{I\left(\mathbf{X}_j; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right)}.$$

Consider the following point: $\mathbf{r}_{GSD}^{\theta} = \theta \mathbf{r}_{GSD}^{(1)} + (1-\theta)\mathbf{r}_{GSD}^{(2)}$, which is in $\mathcal{P}_{GSD,s}^*$. The corresponding sum fronthaul requirement is given by

$$\begin{aligned}
&\theta C^{(1)} + (1-\theta)C^{(2)} \\
&= \theta I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right) + (1-\theta) I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j}^K\right) \\
&= C \times \frac{I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right) - I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j}^K\right)}{I\left(\mathbf{X}_j; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right)} \\
&\quad + \frac{I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right)}{I\left(\mathbf{X}_j; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right)} \times \Big[I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j}^K\right) \\
&\qquad - I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{1}^K\right) - I\left(\mathbf{X}_1^{j-1}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j}^K\right)\Big] \\
&\overset{(c)}{=} C \times \frac{I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right) - I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j}^K\right)}{I\left(\mathbf{X}_j; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right)} \\
&\quad + \frac{I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right)}{I\left(\mathbf{X}_j; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right)} \times \Big[I\left(\mathbf{X}_1^{j-1}, \mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j}^K\right) \\
&\qquad - I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{1}^K\right) - I\left(\mathbf{X}_1^{j-1}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j}^K\right)\Big] \\
&\overset{(d)}{\leq} C \times \frac{I\left(\mathbf{X}_j, \mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right) - I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j}^K\right)}{I\left(\mathbf{X}_j; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right)} \\
&= C,
\end{aligned} \tag{46}$$

where the equality (c) follows from the fact that $I\left(\mathbf{X}_1^{j-1}, \mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j}^K\right) = I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j}^K\right)$ due to Markov chain $\mathbf{X}_{\mathcal{K}} \leftrightarrow \mathbf{Y}_{\mathcal{L}} \leftrightarrow \hat{\mathbf{Y}}_{\mathcal{L}}$, and inequality (d) follows from the fact that $I\left(\mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right) \leq I\left(\mathbf{X}_j, \mathbf{Y}_{\mathcal{L}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{j+1}^K\right)$. Thus, we have that $\mathbf{r}_{JD}^2$ is dominated by some point lying on line segment between $\mathbf{r}_{GSD}^{(1)}$ and $\mathbf{r}_{GSD}^{(2)}$, which lies in $\mathcal{P}_{GSD,s}^*$.

Therefore, for every extreme point $(\tilde{R}_1, \ldots, \tilde{R}_K)$ of $\mathcal{R}_{JD}^o$, the point $(\tilde{R}_1, \ldots, \tilde{R}_K, C)$ lies in $\mathcal{P}_{GSD,s}^*$. This completes the proof.

Appendix B
Submodular Functions

In this appendix, we review some basic results in submodular optimization used proving Theorem 1 and Theorem 2. We tailor our statements toward submodularity and supermodularity, which are used in the proofs.

We begin with the definition of submodular function.

Definition 2: Let $\mathcal{D} = \{1, \ldots, n\}$ be a finite set. A set function $f : 2^{\mathcal{D}} \to \mathbb{R}$ is submodular if for all $\mathcal{S}, \mathcal{T} \subseteq \mathcal{D}$,

$$f(\mathcal{S}) + f(\mathcal{T}) \geq f(\mathcal{S} \cup \mathcal{T}) + f(\mathcal{S} \cap \mathcal{T}). \tag{47}$$

Algorithm 1 Greedy Algorithm for Submodular Polyhedron

```
comment: Returns extreme point $(v_1, \ldots, v_n)$ of $\mathcal{P}(f)$ with the ordering $\prec$.
for $j = 1, \ldots, n$ do
    Set $v_j = f\left(\{i_1, i_2, \ldots, i_j\}\right) - f\left(\{i_1, i_2, \ldots, i_{j-1}\}\right)$.
end for
return $(v_1, \ldots, v_n)$
```

Definition 3: Let $\mathcal{E} = \{1, \ldots, m\}$ be a finite set. A set function $g : 2^{\mathcal{E}} \to \mathbb{R}$ is supermodular if for all $\mathcal{S}, \mathcal{T} \subseteq \mathcal{E}$,

$$g(\mathcal{S}) + g(\mathcal{T}) \leq g(\mathcal{S} \cup \mathcal{T}) + g(\mathcal{S} \cap \mathcal{T}). \tag{48}$$

If the function f is submodular, we call a polyhedron defined by

$$\mathcal{P}(f) = \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n : \sum_{i \in \mathcal{S}} x_i \leq f(\mathcal{S}),\ \forall\, \mathcal{S} \subseteq \mathcal{D} \right\} \tag{49}$$

the submodular polyhedron associated with the submodular function f. Similarly, we define the supermodular polyhedron $\mathcal{P}(g)$ to be the set of $(x_1, \ldots, x_n) \in \mathbb{R}^n$ satisfying

$$\sum_{i \in \mathcal{T}} x_i \geq g(\mathcal{T}),\ \forall\, \mathcal{T} \subseteq \mathcal{E}. \tag{50}$$

We say a point in $\mathcal{P}(f)$ is an extreme point if it cannot be expressed as a convex combination of the other two points in $\mathcal{P}(f)$.

One important property of submodular polyhedron is that all the extreme points can be enumerated through solving a linear optimization. The following proposition provides an algorithm that enumerates the extreme points of $\mathcal{P}(f)$.

Proposition 6 ([36], [37]): For a linear ordering $i_1 \prec i_2 \prec \cdots \prec i_n$ of the elements in $\mathcal{D}$, Algorithm 1 returns an extreme point $(v_1, \ldots, v_n)$ of $\mathcal{P}(f)$. Moreover, all extreme points of $\mathcal{P}(f)$ can be enumerated by considering all linear orderings of the elements of $\mathcal{D}$.

Proposition 6 is the key tool we employ to prove Theorem 1 and Theorem 2. In order to apply this proposition, we require the following lemmas,

Lemma 3: For any joint distribution $\prod_{k=1}^K p\left(\mathbf{x}_k\right) \prod_{\ell=1}^L p\left(\mathbf{y}_\ell|\mathbf{x}_1^K\right) \prod_{\ell=1}^L p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell)$ and fixed $C \in \mathbb{R}$, the set function $f : 2^{\mathcal{K}} \to \mathbb{R}$ defined as follows

$$f\left(\mathcal{T}\right) := \min \left\{ C - \sum_{\ell \in \mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}}),\ I\left(\mathbf{X}_{\mathcal{T}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{\mathcal{T}^c}\right) \right\}$$

is submodular.

Proof: Define a set function $f'\left(\mathcal{T}\right) = I\left(\mathbf{X}_{\mathcal{T}}; \hat{\mathbf{Y}}_{\mathcal{L}}|\mathbf{X}_{\mathcal{T}^c}\right)$. By definition, it can be verified that function f' is submodular [38]. Under fixed sum fronthaul capacity C and conditional distribution $\prod_{\ell=1}^L p_{\hat{\mathbf{Y}}_\ell|\mathbf{Y}_\ell}$, the expression $C - \sum_{\ell \in \mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}})$ is a constant. Let $C' = C - \sum_{\ell \in \mathcal{L}} I(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}})$. Now the problem reduces to show that $f\left(\mathcal{T}\right) = \min\left\{C', f'\left(\mathcal{T}\right)\right)$ is submodular.

Next, observe that f' is monotonically increasing, i.e., if $\mathcal{S} \subset \mathcal{T}$, then $f'(\mathcal{S}) \le f'(\mathcal{T})$. Thus, fixing $\mathcal{S}, \mathcal{T} \subseteq \mathcal{K}$, we can assume without loss of generality that

$$f'(\mathcal{S} \cap \mathcal{T}) \le f'(\mathcal{S}) \le f'(\mathcal{T}) \le f'(\mathcal{S} \cup \mathcal{T})$$

If $C' \le f'(\mathcal{S} \cap \mathcal{T})$, then $f(\mathcal{S}) = f(\mathcal{T}) = f(\mathcal{S} \cap \mathcal{T}) = f(\mathcal{T}) \le f'(\mathcal{S} \cup \mathcal{T}) = C'$. Clearly, f is then submodular. On the other hand, if $C' \ge f'(\mathcal{S} \cup \mathcal{T})$, then $f(\mathcal{S}) = f'(\mathcal{S})$, $f(\mathcal{T}) = f'(\mathcal{T})$, $f(\mathcal{S} \cap \mathcal{T}) = f'(\mathcal{S} \cap \mathcal{T})$, and $f(\mathcal{S} \cup \mathcal{T}) = f'(\mathcal{S} \cup \mathcal{T})$, f is also submodular. Thus, it suffices to check the following three cases:

- Case 1: $f'(\mathcal{S} \cap \mathcal{T}) \le C' \le f'(\mathcal{S}) \le f'(\mathcal{T}) \le f'(\mathcal{S} \cup \mathcal{T})$. By definition of function f, we have
$$f(\mathcal{S}) + f(\mathcal{T}) \ge C' + f'(\mathcal{S} \cap \mathcal{T}) = f(\mathcal{S} \cup \mathcal{T}) + f(\mathcal{S} \cap \mathcal{T}).$$
- Case 2: $f'(\mathcal{S} \cap \mathcal{T}) \le f'(\mathcal{S}) \le C' \le f'(\mathcal{T}) \le f'(\mathcal{S} \cup \mathcal{T})$. Since f' is monotonically increasing, we have
$$\begin{aligned} f(\mathcal{S}) + f(\mathcal{T}) = f'(\mathcal{S}) + C' &\ge f'(\mathcal{S} \cap \mathcal{T}) + f(\mathcal{S} \cup \mathcal{T}) \\ &= f(\mathcal{S} \cap \mathcal{T}) + f(\mathcal{S} \cup \mathcal{T}). \end{aligned}$$
- Case 3: $f'(\mathcal{S} \cap \mathcal{T}) \le f'(\mathcal{S}) \le f'(\mathcal{T}) \le C' \le f'(\mathcal{S} \cup \mathcal{T})$. In this case, the submodularity of f' and the fact of $f' \le f$ imply that
$$\begin{aligned} f(\mathcal{S}) + f(\mathcal{T}) &= f'(\mathcal{S}) + f'(\mathcal{T}) \\ &\ge f'(\mathcal{S} \cap \mathcal{T}) + f'(\mathcal{S} \cup \mathcal{T}) \\ &\ge f(\mathcal{S} \cap \mathcal{T}) + f(\mathcal{S} \cup \mathcal{T}). \end{aligned}$$

Hence, $f = \min\{C', f'\}$ is submodular, which completes the proof of Lemma 3. ∎

Lemma 4: For any joint distribution $\prod_{k=1}^{K} p(\mathbf{x}_k) \prod_{\ell=1}^{L} p\left(\mathbf{y}_\ell|\mathbf{x}_1^K\right) \prod_{\ell=1}^{L} p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell)$ and fixed $R \in \mathbb{R}$, define the set function $g : 2^{\mathcal{L}} \to \mathbb{R}$ as:

$$g(\mathcal{S}) := R + \sum_{\ell \in \mathcal{S}} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}}\right) - I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{S}^c}\right),$$

and the corresponding non-negative set function $g^+ : 2^{\mathcal{L}} \to \mathbb{R}_+$ as $g^+ = \max\{g, 0\}$. The functions g and g^+ are supermodular.

Proof: We first prove that the set function $g'(\mathcal{T}) = I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{T}}\right)$ is submodular. To this end, we evaluate

$$\begin{aligned} & g'(\mathcal{T} \cap \mathcal{S}) + g'(\mathcal{T} \cup \mathcal{S}) \\ &\quad = I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{T} \cup \mathcal{S}}\right) + I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{T} \cap \mathcal{S}}\right) \\ &\quad = I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{S}}, \hat{\mathbf{Y}}_{\mathcal{S}^c \cap \mathcal{T}}\right) + I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{T} \cap \mathcal{S}}\right) \\ &\quad = g'(\mathcal{S}) + g'(\mathcal{T}) + I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{S}^c \cap \mathcal{T}}|\hat{\mathbf{Y}}_{\mathcal{S}}\right) \\ &\qquad - I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{S}^c \cap \mathcal{T}}|\hat{\mathbf{Y}}_{\mathcal{T} \cap \mathcal{S}}\right). \end{aligned}$$

Furthermore,

$$\begin{aligned} & I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{S}^c \cap \mathcal{T}}|\hat{\mathbf{Y}}_{\mathcal{S}}\right) - I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{S}^c \cap \mathcal{T}}|\hat{\mathbf{Y}}_{\mathcal{T} \cap \mathcal{S}}\right) \\ &\quad = h\left(\hat{\mathbf{Y}}_{\mathcal{S}^c \cap \mathcal{T}}|\hat{\mathbf{Y}}_{\mathcal{S}}\right) - h\left(\hat{\mathbf{Y}}_{\mathcal{S}^c \cap \mathcal{T}}|\hat{\mathbf{Y}}_{\mathcal{S}}, \mathbf{X}_{\mathcal{K}}\right) \\ &\qquad - h\left(\hat{\mathbf{Y}}_{\mathcal{S}^c \cap \mathcal{T}}|\hat{\mathbf{Y}}_{\mathcal{T} \cap \mathcal{S}}\right) + h\left(\hat{\mathbf{Y}}_{\mathcal{S}^c \cap \mathcal{T}}|\hat{\mathbf{Y}}_{\mathcal{T} \cap \mathcal{S}}, \mathbf{X}_{\mathcal{K}}\right) \\ &\quad = h\left(\hat{\mathbf{Y}}_{\mathcal{S}^c \cap \mathcal{T}}|\hat{\mathbf{Y}}_{\mathcal{S}}\right) - h\left(\hat{\mathbf{Y}}_{\mathcal{S}^c \cap \mathcal{T}}|\hat{\mathbf{Y}}_{\mathcal{S} \cap \mathrm{T}}\right) \\ &\quad \le 0. \end{aligned}$$

Therefore, $g'(\mathcal{T} \cap \mathcal{S}) + g'(\mathcal{T} \cup \mathcal{S}) \le g'(\mathcal{S}) + g'(\mathcal{T})$, which proves that g' is submodular.

In the following, we prove that g is supermodular. Evaluate $g(\mathcal{S}) + g(\mathcal{T})$ as

$$\begin{aligned} & g(\mathcal{S}) + g(\mathcal{T}) \\ &\quad = 2R + \sum_{\ell \in \mathcal{S}} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}}\right) + \sum_{\ell \in \mathcal{T}} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}}\right) \\ &\qquad - I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{S}^c}\right) - I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{T}^c}\right) \\ &\quad \overset{(e)}{\le} 2R + \sum_{\ell \in \mathcal{S} \cup \mathcal{T}} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}}\right) + \sum_{\ell \in \mathcal{S} \cap \mathcal{T}} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}}\right) \\ &\qquad - I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{(\mathcal{S} \cap \mathcal{T})^c}\right) - I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{(\mathcal{S} \cup \mathcal{T})^c}\right) \\ &\quad = g(\mathcal{S} \cap \mathcal{T}) + g(\mathcal{S} \cup \mathcal{T}), \end{aligned}$$

where inequality (e) follows from the fact that $g'(\mathcal{T}) = I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{T}}\right)$ is a submodular function.

Therefore, we show that g is supermodular. Following the result of [28, Lemma 6], it can be shown that $g^+ = \max\{g, 0\}$ is also supermodular. ∎

Appendix C
Optimality of Successive Decoding for Maximizing Sum Rate

Similar to the proof of Theorem 1, Theorem 2 can also be proven using submodular optimization. In the following, we consider the region $(R, C_1, \ldots, C_L)$, and prove that joint decoding and successive decoding achieve the same maximum rate using the properties of submodular optimization.

Definition 4: Define $\mathcal{P}_{JD}^s$ to be the closure of the convex hull of all $(R, C_1, \ldots, C_L)$ satisfying

$$R < \sum_{\ell \in \mathcal{S}} \left[C_\ell - I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell|\mathbf{X}_{\mathcal{K}}\right)\right] + I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{S}^c}\right), \ \forall\, \mathcal{S} \subseteq \mathcal{L}, \tag{51}$$

for some product distribution $\prod_{k=1}^{K} p(\mathbf{x}_k) \prod_{\ell=1}^{L} p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell)$.

Definition 5: Define $\mathcal{P}_{SD}^s$ to be the closure of the convex hull all $(R, C_1, \ldots, C_L)$ satisfying

$$\begin{cases} R < I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_{\mathcal{L}}\right), \\ \displaystyle\sum_{\ell \in \mathcal{S}} C_\ell > I\left(\mathbf{Y}_{\mathcal{S}}; \hat{\mathbf{Y}}_{\mathcal{S}}|\hat{\mathbf{Y}}_{\mathcal{S}^c}\right), \quad \forall\, \mathcal{S} \subseteq \mathcal{L}, \end{cases} \tag{52}$$

for some product distribution $\prod_{k=1}^{K} p(\mathbf{x}_k) \prod_{\ell=1}^{L} p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell)$.

Note that $\mathcal{P}_{JD}^s$ represents the sum-rate and fronthaul-capacity region of joint decoding. All the partial sums over $\mathcal{S}$ in (51) can be strictly attained with equality depending on the values of the fronthaul capacities C_ℓ for $\ell = 1, \ldots, L$ and the sum rate R. Similarly, $\mathcal{P}_{SD}^s$ corresponds to the region of successive decoding. For fixed product distribution $\prod_{k=1}^{K} p(\mathbf{x}_k) \prod_{\ell=1}^{L} p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell)$, we say a point $(R, C_1, \ldots, C_L)$ is dominated by a point $(R', C_1', \ldots, C_L')$ in $\mathcal{P}_{SD}^s$ if $C_\ell' \le C_\ell$ for $\ell = 1, \ldots, L$ and $R' \ge R$.

Clearly, the maximum sum rate achieved by joint decoding is always larger or equal to that achieved by successive decoding, i.e., $R_{JD,SUM}^* \ge R_{SD,SUM}^*$. To show $R_{JD,SUM}^* =$

$R^*_{SD,SUM}$, it remains to show that $R^*_{JD,SUM} \leq R^*_{SD,SUM}$. For any given product distribution $\prod_{k=1}^{K} p(\mathbf{x}_k) \prod_{\ell=1}^{L} p(\hat{\mathbf{y}}_\ell|\mathbf{y}_\ell)$ and joint decoding sum rate R_{JD}, define $\mathcal{P}_C \subset \mathbb{R}_+^L$ to be the set of $(C_1, \ldots, C_L)$ such that

$$\sum_{\ell \in \mathcal{S}} C_\ell \geq \left[R_{JD} + \sum_{\ell \in \mathcal{S}} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell | \mathbf{X}_\mathcal{K}\right) - I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}_{\mathcal{S}^c}\right) \right]^+, \tag{53}$$

for all $\mathcal{S} \subseteq \mathcal{L}$. Now, to show $R^*_{JD,SUM} \leq R^*_{SD,SUM}$, it suffices to show that each extreme point of $(R_{JD}, \mathcal{P}_C)$ is dominated by a point in $\mathcal{P}^s_{SD}$ that achieves a sum rate greater or equal to the joint decoding sum rate R_{JD}.

To this end, define a set function $g : 2^\mathcal{L} \to \mathbb{R}$ as follows:

$$g(\mathcal{S}) := R_{JD} + \sum_{\ell \in \mathcal{S}} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell | \mathbf{X}_\mathcal{K}\right) - I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}_{\mathcal{S}^c}\right),$$

for each $\mathcal{S} \subseteq \mathcal{L}$. It can be verified that the function $g^+(\mathcal{S}) = \max\{g(\mathcal{S}), 0\}$ is a supermodular function (see Appendix B, Lemma 4). By construction, $\mathcal{P}_C$ is equal to the set of $(C_1, R_2, \ldots, C_L)$ satisfying

$$\sum_{\ell \in \mathcal{S}} C_\ell \geq g^+(\mathcal{S}), \quad \forall\, \mathcal{S} \subseteq \mathcal{L}.$$

Following the results in submodular optimization (Appendix B, Proposition 6), we have that for a linear ordering $i_1 \prec i_2 \prec \cdots \prec i_K$ on the set $\mathcal{K}$, an extreme point of $\mathcal{P}_C$ can be computed as follows

$$\tilde{C}_{i_j} = g^+\left(\{i_1, \ldots, i_j\}\right) - g^+\left(\{i_1, \ldots, i_{j-1}\}\right).$$

All the $L!$ extreme points of $\mathcal{P}_C$ can be analyzed in the same manner. For notational simplicity we only consider the natural ordering $i_j = j$ in the following proof.

By construction,

$$\tilde{C}_j = \left[R_{JD} + \sum_{\ell=1}^{j} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell | \mathbf{X}_\mathcal{K}\right) - I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j+1}\right) \right]^+ - \left[R_{JD} + \sum_{\ell=1}^{j-1} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell | \mathbf{X}_\mathcal{K}\right) - I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j}\right) \right]^+.$$

Let j be the first index for which $g(\{1, \ldots, j\}) > 0$. Then, by construction,

$$\begin{aligned} \tilde{C}_k &= I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}_k | \hat{\mathbf{Y}}^L_{k+1}\right) + I\left(\mathbf{Y}_k; \hat{\mathbf{Y}}_k | \mathbf{X}_\mathcal{K}\right) \\ &= I\left(\mathbf{Y}_k; \hat{\mathbf{Y}}_k | \hat{\mathbf{Y}}^L_{k+1}\right) \end{aligned}$$

for all $k > j$, where the Markov chain $\hat{\mathbf{Y}}_i \leftrightarrow \mathbf{Y}_i \leftrightarrow \mathbf{X}_\mathcal{K} \leftrightarrow \mathbf{Y}_j \leftrightarrow \hat{\mathbf{Y}}_j$, for $i \neq j$, is utilized in deriving the second equality. Clearly, all the $\tilde{C}_k$'s are in the successive decoding region $\mathcal{P}^s_{SD}$.

Moreover, we have $g\left(\{1, \ldots, j'\}\right) \leq 0$ for all $j' < j$. Thus, $\tilde{C}_j$ can be expressed as

$$\begin{aligned} \tilde{C}_j &= R_{JD} + \sum_{\ell=1}^{j} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell | \mathbf{X}_\mathcal{K}\right) - I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j+1}\right) \\ &= \alpha I\left(\mathbf{Y}_{j+1}; \hat{\mathbf{Y}}_{j+1} | \hat{\mathbf{Y}}^L_{j+1}\right) \end{aligned}$$

where $\alpha \in [0, 1]$ is defined as

$$\alpha = \frac{R_{JD} + \sum_{\ell=1}^{j} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell | \mathbf{X}_\mathcal{K}\right) - I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j+1}\right)}{I\left(\mathbf{Y}_{j+1}; \hat{\mathbf{Y}}_{j+1} | \hat{\mathbf{Y}}^L_{j+1}\right)}.$$

Consider the two following successive decoding schemes:

- Scheme 1: The CP decodes quantization codewords $\hat{\mathbf{Y}}_{j+1}, \ldots, \hat{\mathbf{Y}}_L$ first, then decodes the user message codewords $\mathbf{X}_\mathcal{K}$ sequentially. Note that the BSs with index $i \leq j$ are inactive, and are essentially removed from the network. The resulting extreme point $\mathbf{c}^{(1)} = (R^{(1)}_{SD}, C^{(1)}_1, \ldots, C^{(1)}_L)$ of $\mathcal{P}^s_{SD}$ satisfies
$$\begin{cases} C^{(1)}_i = 0, & \text{for } i \leq j, \\ C^{(1)}_i = I\left(\mathbf{Y}_i; \hat{\mathbf{Y}}_i | \hat{\mathbf{Y}}^L_{i+1}\right), & \text{for } i > j, \\ R^{(1)}_{SD} = I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j+1}\right). \end{cases}$$
- Scheme 2: The CP decodes quantization codewords $\hat{\mathbf{Y}}_j, \ldots, \hat{\mathbf{Y}}_L$ first, then decodes the user message codewords $\mathbf{X}_\mathcal{K}$ sequentially. Note that in this scheme, the BSs with index $i < j$ are inactive, and are essentially removed from the network. The resulting extreme point $\mathbf{c}^{(2)} = (R^{(2)}_{SD}, C^{(2)}_1, \ldots, C^{(2)}_L)$ of $\mathcal{P}^s_{SD}$ satisfies
$$\begin{cases} C^{(2)}_i = 0, & \text{for } i < j, \\ C^{(2)}_i = I\left(\mathbf{Y}_i; \hat{\mathbf{Y}}_i | \hat{\mathbf{Y}}^L_{i+1}\right), & \text{for } i \geq j, \\ R^{(2)}_{SD} = I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j}\right). \end{cases}$$

Since C_ℓ is defined to be the maximum long-term average throughput of fronthaul link ℓ, the following point: $\mathbf{c}^\alpha = (1-\alpha)\mathbf{c}^{(1)} + \alpha \mathbf{c}^{(2)}$ lies in $\mathcal{P}^s_{SD}$. The corresponding sum rate R_{SD} in $\mathbf{c}^\alpha$ is given by

$$\begin{aligned} &(1-\alpha) R^{(1)}_{SD} + \alpha R^{(2)}_{SD} \\ &= (1-\alpha) I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j+1}\right) + \alpha I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j}\right) \\ &\overset{(f)}{=} \frac{I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j}\right) - R_{JD} - \sum_{\ell=1}^{j-1} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell | \mathbf{X}_\mathcal{K}\right)}{I\left(\mathbf{Y}_{j+1}; \hat{\mathbf{Y}}_{j+1} | \hat{\mathbf{Y}}^L_{j+1}\right)} \\ &\qquad \times I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j+1}\right) \\ &\qquad + \frac{R_{JD} + \sum_{\ell=1}^{j} I\left(\mathbf{Y}_\ell; \hat{\mathbf{Y}}_\ell | \mathbf{X}_\mathcal{K}\right) - I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j+1}\right)}{I\left(\mathbf{Y}_{j+1}; \hat{\mathbf{Y}}_{j+1} | \hat{\mathbf{Y}}^L_{j+1}\right)} \\ &\qquad \times I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j}\right) \\ &\geq \frac{R_{JD} \times \left[I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j}\right) - I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j+1}\right) \right]}{I\left(\mathbf{Y}_{j+1}; \hat{\mathbf{Y}}_{j+1} | \hat{\mathbf{Y}}^L_{j+1}\right)} \\ &\qquad + \frac{I\left(\mathbf{Y}_j; \hat{\mathbf{Y}}_j | \mathbf{X}_\mathcal{K}\right) \times I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j}\right)}{I\left(\mathbf{Y}_{j+1}; \hat{\mathbf{Y}}_{j+1} | \hat{\mathbf{Y}}^L_{j+1}\right)} \\ &\overset{(g)}{\geq} R_{JD} \times \frac{I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j}\right) - I\left(\mathbf{X}_\mathcal{K}; \hat{\mathbf{Y}}^L_{j+1}\right) + I\left(\mathbf{Y}_j; \hat{\mathbf{Y}}_j | \mathbf{X}_\mathcal{K}\right)}{I\left(\mathbf{Y}_{j+1}; \hat{\mathbf{Y}}_{j+1} | \hat{\mathbf{Y}}^L_{j+1}\right)} \\ &= R_{JD}, \end{aligned} \tag{54}$$

where the equality (f) follows from the fact that $I\left(\mathbf{X}_{\mathcal{K}}, \mathbf{Y}_{j+1}; \hat{\mathbf{Y}}_{j+1}|\hat{\mathbf{Y}}_{j+1}^L\right) = I\left(\mathbf{Y}_{j+1}; \hat{\mathbf{Y}}_{j+1}|\hat{\mathbf{Y}}_{j+1}^L\right)$, and inequality (g) follows from the fact that $R_{JD} \le I\left(\mathbf{X}_{\mathcal{K}}; \hat{\mathbf{Y}}_j^L\right)$.

Therefore, for every extreme point $(\tilde{C}_1, \ldots, \tilde{C}_L)$ of $\mathcal{P}_C$, the point $(R_{JD}, \tilde{C}_1, \ldots, \tilde{C}_L)$ is dominated by a point in $\mathcal{P}_{SD}^s$. This proves Theorem 2.

VII. Appendix D
Constant-Gap Result for Compress-and-Forward With Joint Decoding

The idea of the proof is to compare the achievable rate of compress-and-forward with joint decoding with the following cut-set upper bound [6]

$$\sum_{k\in\mathcal{T}} R_k \le \min_{\mathcal{S}\subseteq\mathcal{L}} \left\{ \sum_{\ell\in\mathcal{S}} C_\ell + \log \frac{\left|\sum_{\ell\in\mathcal{S}^c} \mathbf{H}_{\ell,\mathcal{T}}^\dagger \Sigma_\ell^{-1} \mathbf{H}_{\ell,\mathcal{T}} + \mathbf{K}_{\mathcal{T}}^{-1}\right|}{\left|\mathbf{K}_{\mathcal{T}}^{-1}\right|} \right\} \tag{55}$$

for all $\emptyset \subset \mathcal{T} \subseteq \mathcal{K}$. In the expression of cut-set bound, the first term represents the cut across the fronthaul links in set $\mathcal{S}$, and the second term represents the cut from the users to the BSs in set $\mathcal{S}^c$.

Recall that the rate region for joint decoding (23) under Gaussian quantization is the set of $(R_1, \cdots, R_K)$ such that

$$\sum_{k\in\mathcal{T}} R_k < \sum_{\ell\in\mathcal{S}} \left[C_\ell - \log \frac{\left|\Sigma_\ell^{-1}\right|}{\left|\Sigma_\ell^{-1} - \mathbf{B}_\ell\right|} \right] + \log \frac{\left|\sum_{\ell\in\mathcal{S}^c} \mathbf{H}_{\ell,\mathcal{T}}^\dagger \mathbf{B}_\ell \mathbf{H}_{\ell,\mathcal{T}} + \mathbf{K}_{\mathcal{T}}^{-1}\right|}{\left|\mathbf{K}_{\mathcal{T}}^{-1}\right|}$$

for all $\emptyset \subset \mathcal{T} \subseteq \mathcal{K}$ and $\mathcal{S} \subseteq \mathcal{L}$, for some $0 \preceq \mathbf{B}_\ell \preceq \Sigma_\ell^{-1}$. We now show that if a rate tuple $(R_1, \cdots, R_K)$ is within the cut-set bound, then $(R_1 - \eta, \cdots, R_K - \eta)$ is in the achievable rate region of joint decoding, where

$$|\mathcal{T}|\eta \le \sum_{\ell\in\mathcal{S}} \log \frac{\left|\Sigma_\ell^{-1}\right|}{\left|\Sigma_\ell^{-1} - \mathbf{B}_\ell\right|} + \log \frac{\left|\sum_{\ell\in\mathcal{S}^c} \mathbf{H}_{\ell,\mathcal{T}}^\dagger \Sigma_\ell^{-1} \mathbf{H}_{\ell,\mathcal{T}} + \mathbf{K}_{\mathcal{T}}^{-1}\right|}{\left|\sum_{\ell\in\mathcal{S}^c} \mathbf{H}_{\ell,\mathcal{T}}^\dagger \mathbf{B}_\ell \mathbf{H}_{\ell,\mathcal{T}} + \mathbf{K}_{\mathcal{T}}^{-1}\right|} \tag{56}$$

is the gap between the cut-set bound and achievable rate of joint decoding.

Choose quantization noise level to be at the background noise level, i.e., $\mathbf{Q}_\ell = \Sigma_\ell$. Then we have

$$\mathbf{B}_\ell = (\Sigma_\ell + \mathbf{Q}_\ell)^{-1} = \frac{1}{2}\Sigma_\ell^{-1}.$$

Evaluating gap η with the above choice of $\mathbf{B}_\ell$ gives

$$\eta \le \frac{|\mathcal{S}|}{|\mathcal{T}|} \cdot N + M \le NL + M,$$

which completes the proof of Proposition 3.

References

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

[2] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

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

[4] A. Sanderovich, S. Shamai, and Y. Steinberg, "Distributed MIMO receiver—Achievable rates and upper bounds," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4419–4438, Oct. 2009.

[5] A. Sanderovich, S. Shamai, Y. Steinberg, and G. Kramer, "Communication via decentralized processing," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3008–3023, Jul. 2008.

[6] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[7] X. Wu and L.-L. Xie, "On the optimal compressions in the compress-and-forward relay schemes," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2613–2628, May 2013.

[8] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.

[9] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[10] M. H. Yassaee and M. R. Aref, "Slepian–Wolf coding over cooperative relay networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3462–3482, Jun. 2011.

[11] J. Hou and G. Kramer, "Short message noisy network coding with a decode–forward option," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 89–107, Jan. 2016.

[12] Y. Zhou and W. Yu, "Optimized backhaul compression for uplink cloud radio access network," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1295–1307, Jun. 2014.

[13] C. Tian and J. Chen, "Remote vector Gaussian source coding with decoder side information under mutual information and distortion constraints," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4676–4680, Oct. 2009.

[14] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem [multiterminal source coding]," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[15] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2577–2593, Jul. 2005.

[16] V. Prabhakaran, D. Tse, and K. Ramachandran, "Rate region of the quadratic Gaussian CEO problem," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun./Jul. 2004, p. 119.

[17] J. Wang and J. Chen, "Vector Gaussian multiterminal source coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5533–5552, Sep. 2014.

[18] E. Ekrem and S. Ulukus, "An outer bound for the vector Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6870–6887, Nov. 2014.

[19] N. Tishby, F. Pereira, and W. Bialek, "The information bottleneck method," in *Proc. 37th Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 1999, pp. 368–377.

[20] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information-theoretic problems," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1839–1851, May 2007.

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

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

[23] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "Joint decompression and decoding for cloud radio access networks," *IEEE Signal Process. Lett.*, vol. 20, no. 5, pp. 503–506, May 2013.

[24] Y. Zhou and W. Yu, "Fronthaul compression and transmit beamforming optimization for multi-antenna uplink C-RAN," *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4138–4151, Aug. 2016.

[25] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[26] A. El Gamal, M. Mohseni, and S. Zahedi, "Bounds on capacity and minimum energy-per-bit for AWGN relay channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1545–1561, Apr. 2006.

[27] L. Zhou and W. Yu, "Uplink multicell processing with limited backhaul via per-base-station successive interference cancellation," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 1981–1993, Oct. 2013.

[28] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 740–761, Jan. 2014.

[29] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels. I. Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.

[30] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1501–1518, Nov. 1991.

[31] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.

[32] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[33] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai (Shitz), "Fronthaul compression for cloud radio access networks: Signal processing advances inspired by network information theory," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 69–79, Nov. 2014.

[34] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.

[35] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 19, no. 3, pp. 357–359, May 1973.

[36] S. Fujishige, *Submodular Functions Optimization*, 2nd ed. New York, NY, USA: Elsevier, 2005.

[37] S. Iwata, "Submodular function minimization," *Math. Program.*, vol. 112, no. 1, pp. 45–64, 2008.

[38] X. Zhang, J. Chen, S. B. Wicker, and T. Berger, "Successive coding in multiuser information theory," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2246–2254, Jun. 2007.

Yuhan Zhou (S'08–M'16) received the B.E. degree in Electronic and Information Engineering from Jilin University, Jilin, China, in 2005, the M.A.Sc. degree from the University of Waterloo, ON, Canada, in 2009, and the Ph.D. degree from the University of Toronto, ON, Canada, in 2016, both in Electrical and Computer Engineering. Since 2016, he has been with Qualcomm Technologies Inc., San Diego, CA, USA. His research interests include wireless communications, network information theory, and convex optimization.

Yinfei Xu (S'10) was born in July 1986 in Nanjing, China. He received the B.E. and Ph.D. degrees in 2008 and 2016, respectively, both in Information Engineering, from Southeast University, Nanjing, China. Since March 2016, he has been in the Institute of Network Coding at The Chinese University of Hong Kong, Hong Kong, where he is currently a Postdoctoral Fellow. He was a visiting student in the Department of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada, from July 2014 to January 2015. His research interests include information theory, signal processing and wireless communications.

Wei Yu (S'97–M'02–SM'08–F'14) received the B.A.Sc. degree in Computer Engineering and Mathematics from the University of Waterloo, Waterloo, Ontario, Canada in 1997 and M.S. and Ph.D. degrees in Electrical Engineering from Stanford University, Stanford, CA, in 1998 and 2002, respectively. Since 2002, he has been with the Electrical and Computer Engineering Department at the University of Toronto, Toronto, Ontario, Canada, where he is now Professor and holds a Canada Research Chair (Tier 1) in Information Theory and Wireless Communications. His main research interests include information theory, optimization, wireless communications and broadband access networks.

Prof. Wei Yu currently serves on the IEEE Information Theory Society Board of Governors (2015-17). He is an IEEE Communications Society Distinguished Lecturer (2015-16). He served as an Associate Editor for IEEE TRANSACTIONS ON INFORMATION THEORY (2010-2013), as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS (2009-2011), as an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2004-2007), and as a Guest Editor for a number of special issues for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and the EURASIP JOURNAL ON APPLIED SIGNAL PROCESSING. He was a Technical Program co-chair of the IEEE Communication Theory Workshop in 2014, and a Technical Program Committee co-chair of the Communication Theory Symposium at the IEEE International Conference on Communications (ICC) in 2012. He was a member of the Signal Processing for Communications and Networking Technical Committee of the IEEE Signal Processing Society (2008-2013), then Vice Chair in 2016. Prof. Wei Yu received a Steacie Memorial Fellowship in 2015, an IEEE Communications Society Best Tutorial Paper Award in 2015, an IEEE ICC Best Paper Award in 2013, an IEEE Signal Processing Society Best Paper Award in 2008, the McCharles Prize for Early Career Research Distinction in 2008, the Early Career Teaching Award from the Faculty of Applied Science and Engineering, University of Toronto in 2007, and an Early Researcher Award from Ontario in 2006. He is a Highly Cited Researcher according to Thomson Reuters.

Jun Chen (S'03–M'06–SM'16) received the B.E. degree with honors in communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001 and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, in 2004 and 2006, respectively.

He was a Postdoctoral Research Associate in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana, IL, from September 2005 to July 2006, and a Postdoctoral Fellow at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, from July 2006 to August 2007. Since September 2007 he has been with the Department of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada, where he is currently an Associate Professor and a Joseph Ip Distinguished Engineering Fellow. His research interests include information theory, wireless communications, and signal processing. He received several awards for his research, including the Josef Raviv Memorial Postdoctoral Fellowship in 2006, the Early Researcher Award from the Province of Ontario in 2010, and the IBM Faculty Award in 2010. He is currently serving as an Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY.