On the Capacity of Cloud Radio Access Networks with Oblivious Relaying

Iñaki Estella Aguerri[†]   Abdellatif Zaidi[† ‡]   Giuseppe Caire[⋆]   Shlomo Shamai (Shitz)[∗]

[†]France Research Center, Huawei Technologies, Boulogne-Billancourt, 92100, France
[‡] Université Paris-Est, Champs-sur-Marne, 77454, France
[⋆] Technische Universitat Berlin, 10587 Berlin, Germany
[∗] Technion Institute of Technology, Technion City, Haifa 32000, Israel
{inaki.estella@huawei.com, abdellatif.zaidi@u-pem.fr, caire@tu-berlin.de,
sshlomo@ee.technion.ac.il}

Abstract—We study the transmission over a network in which users send information to a remote destination through relay nodes that are connected to the destination via finite-capacity error-free links, i.e., a cloud radio access network. The relays are constrained to operate without knowledge of the users' codebooks, i.e., they perform oblivious processing. The destination, or central processor, however, is informed about the users' codebooks. We establish a single-letter characterization of the capacity region of this model for a class of discrete memoryless channels in which the outputs at the relay nodes are independent given the users' inputs. We show that both relaying à-la Cover-El Gamal, i.e., compress-and-forward with joint decompression and decoding, and "noisy network coding", are optimal. The proof of the converse part establishes, and utilizes, connections with the Chief Executive Officer (CEO) source coding problem under logarithmic loss distortion measure. Extensions to general discrete memoryless channels are also investigated. In this case, we establish inner and outer bounds on the capacity region. For memoryless Gaussian channels within the studied class of channels, we characterize the capacity under Gaussian channel inputs.

I. INTRODUCTION

Cloud radio access networks (CRAN) provide a new architecture for next-generation wireless cellular systems in which base stations (BSs) are connected to a cloud-computing central processor (CP) via error-free finite-rate fronthaul links. This architecture is generally seen as an efficient means to increase spectral efficiency in cellular networks by enabling joint processing of the signals received by multiple BSs at the CP, thus alleviating the effect of interference. Other advantages include low cost deployment and flexible network utilization [1].

In a CRAN network, each BS acts essentially as a relay node; and so can in principle implement any relaying strategy, e.g., decode-and-forward [2, Theorem 1], compress-and-forward [2, Theorem 6] or combinations of them. Relaying strategies in CRANs can be divided roughly into two classes: i) strategies that require the relay nodes to know the users' codebooks (i.e., modulation, coding), such as decode-and-forward, compute-and-forward [3], [4] or variants of it, and ii) strategies in which the relay nodes operate without knowledge of the users' codebooks, often referred to as *oblivious relay processing* (or *nomadic transmitter*) [5]–[7]. This second class is composed essentially of strategies in which the relays imple-

ment forms of compress-and-forward [2], such as successive Wyner-Ziv compression [8]–[10] and noisy-network coding [11]. Schemes combining the two apporaches have been shown to possibly outperform the best of the two in [12], especially in scenarios in which there are more users than relay nodes.

In the spirit, however, a CRAN architecture is usually envisioned as one in which BSs operate as simple radio units (RUs) that are constrained to implement only the radio functionalities, such as analog-to-digital conversion and filtering, while the baseband functionalities are migrated to the CP. For this reason, while relaying schemes that involve partial or full decoding of the users' codewords can sometimes offer rate gains, they do not seem to be suitable in practice. In fact, such schemes assume that all or a subset of the relay nodes are fully aware (at all times) of the codebooks and encoding used by the users; and the signaling required to convey such information is generally prohibitive, particularly as networks become large. Instead, schemes in which relays perform oblivious processing are preferred. Oblivious processing was first introduced in [5]. The basic idea is that of using *randomized encoding* to model lack of information about codebooks. For related works, the reader may refer to [6], [13], [14]. In particular, [6] extends the original definition of oblivious processing of [5], which rules out time-sharing, to include settings in which encoders are allowed to switch among different codebooks and oblivious nodes are unaware of the codebooks but are given, or can acquire, time- or frequency-schedule information, which is generally less difficult to obtain. The framework is termed therein as "*oblivious processing with enabled time-sharing*".

In this work, we consider transmission over a CRAN in which the relay nodes are constrained to operate without knowledge of the users' codebooks, i.e., are oblivious, and only know time- or frequency-sharing information. Focusing on a class of discrete memoryless channels in which the relay outputs are independent conditionally on the users' inputs, we establish a single-letter characterization of the capacity region of this class of channels. We show that relaying à-la Cover-El Gamal, i.e., compress-and-forward with joint decompression and decoding, or noisy network coding, are optimal. For the proof of the converse part, we utilize useful connections with the Chief Executive Officer (CEO) source coding problem under logarithmic loss distortion measure [15]. For memoryless Gaussian channels within this class, we characterize the capacity under Gaussian channel inputs. Extensions to general discrete memoryless channels are also investigated. In this case, we establish inner and outer bounds on the capacity region.

Notation: Throughout, we use the following notation. Lower case letters denote scalars, e.g., x; upper case letters denote random variables, e.g., X, boldface lower case letters denote vectors, e.g., $\mathbf{x}$, and boldface upper case letters denote matrices, e.g., $\mathbf{X}$. Calligraphic letters denote sets, e.g., $\mathcal{X}$; and the cardinality of set $\mathcal{X}$ is denoted by $|\mathcal{X}|$. For a set of integers $\mathcal{K}$, the notation $X_{\mathcal{K}}$ denotes the set of random variables $\{X_k\}$ with indices k in the set $\mathcal{K}$, i.e., $X_{\mathcal{K}} = \{X_k\}_{k \in \mathcal{K}}$.

II. SYSTEM MODEL

Consider the discrete memoryless CRAN model shown in Figure 1. In this model, a set of users communicate with a central processor (CP) through a set of relay nodes that are connected to the CP via error-free finite-rate fronthaul links. Let $\mathcal{L} = \{1, \ldots, L\}$ denote the set of users, and $\mathcal{K} = \{1, \ldots, K\}$ denote the set of relays, and let C_k be the capacity of the link connecting relay node k to the CP, $k \in \mathcal{K}$. Similar to [6], the relays nodes are constrained to operate without knowledge of the users' codebook and only know time-sharing information, i.e., oblivious relay processing with enabled time sharing. The obliviousness of the relay nodes to the actual codebooks is modeled by the transmitters picking at random their selected codebooks and the relays not aware of the actual codebooks indices. Specifically, the codeword $X^n(F_l, M_l, Q^n)$ transmitted by encoder l, $l \in \mathcal{L}$, depends not only on the message $M_l \in [1, 2^{nR_l}]$, but also on the index F_l which runs over all possible codebooks of the given rate R_l, i.e., $F_l \in [1, |\mathcal{X}_l|^{n2^{nR_l}}]$ and the time sharing sequence Q^n. Formally, the model is defined as follows.

1) *Codebooks:* Transmitter l, $l \in \mathcal{L}$, sends message $m_l \in [1, 2^{nR_l}]$ to the CP using a codebook from a set $\{\mathcal{C}_l(F_l)\}$ that is indexed by $F_l \in [1, |\mathcal{X}_l|^{n2^{nR_l}}]$. The index F_l is picked at random and shared with the CP, but not to the relays.

2) *Time-sharing sequence:* All terminals, including the relay nodes, are aware of a time-sharing sequence Q^n, distributed as $p_{Q^n}(q^n) = \prod_{i=1}^n p_Q(q_i)$ for a pmf $p_Q(q)$.

3) *Encoding functions:* The encoding function at user l, $l \in \mathcal{L}$, is defined by a pair (p_{X_l}, ϕ_l) where p_{X_l} is a single-letter pmf and ϕ_l is a mapping $\phi_l : [1, |\mathcal{X}_l|^{n2^{nR_l}}] \times [1, 2^{nR_l}] \times \mathcal{Q}^n \to \mathcal{X}_l^n$, that maps the given codebook index F_l, message M_l and time-sharing variable Q^n to a channel input $X_l^n = \phi_l(F_l, M_l, Q^n)$. The probability of selecting codebook F_l, $F_l \in [1, |\mathcal{X}_l|^{n2^{nR_l}}]$, is given by

$$p_{F_l}(f_l, q^n) = \prod_{m_l \in [1, 2^{nR_l}]} p_{X_l^n|Q^n}(\phi_l(f_l, m_l)|q^n), \quad (1)$$

where $p_{X_l^n|Q^n}(x_l^n|q^n) = \prod_{i=1}^n p_{X_l|Q}(x_{l,i}|q_i)$ for some given pmf $p_{X_l|Q}(x_l|q)$.

4) *Relaying functions:* Relay node k, $k \in \mathcal{K}$, is unaware of the codebook indices $F_{\mathcal{L}} = (F_1, \ldots, F_L)$, and maps the received sequence $Y_k^n \in \mathcal{Y}_k^n$ into an index $J_k \in [1, 2^{nC_k}]$ as $J_k = \phi_k^r(Y_k^n, Q^n)$. The index J_k is then sent the to the CP over the error-free link of capacity C_k.

5) *Decoding function:* Upon receiving the indices $J_{\mathcal{K}} = (J_1 \ldots, J_K)$, the CP estimates the users' messages as

$$(\hat{M}_1, \ldots, \hat{M}_L) = g(F_1, \ldots, F_L, J_1, \ldots, J_K, Q^n), \quad (2)$$

where $g : [1, |\mathcal{X}_1|^{n2^{nR_1}}] \times \cdots \times [1, |\mathcal{X}_L|^{n2^{nR_L}}] \times [1, 2^{nC_1}] \times \cdots \times [1, 2^{nC_L}] \times \mathcal{Q}^n \to [1, 2^{nR_1}] \times \cdots \times [1, 2^{nR_L}]$ is the decoder mapping.
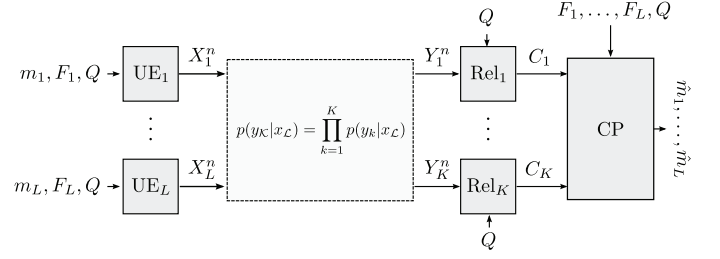

Fig. 1. CRAN model with oblivious relaying and time-sharing.

Definition 1. *A $(n, R_1, \ldots, R_L)$ code for the studied model with oblivious relay processing and enabled time-sharing consist of L encoding functions $\phi_l : [1, |\mathcal{X}_l|^{n2^{nR_l}}] \times [1, 2^{nR_l}] \times \mathcal{Q}^n \to \mathcal{X}_l^n$, K relaying functions $\phi_k^r : \mathcal{Y}_k^n \times \mathcal{Q}^n \to [1, 2^{nC_k}]$, and a decoding function $g : [1, |\mathcal{X}_1|^{n2^{nR_1}}] \times \cdots \times [1, |\mathcal{X}_L|^{n2^{nR_L}}] \times [1, 2^{nC_1}] \times \cdots \times [1, 2^{nC_L}] \times \mathcal{Q}^n \to [1, 2^{nR_1}] \times \cdots [1, 2^{nR_L}]$.*

Definition 2. *A rate tuple $(R_1, \ldots, R_L)$ is said to be achievable if, for any $\epsilon > 0$, there exist a sequence of $(n, R_1, \ldots, R_L)$ codes, such that, for sufficiently long block-length n, each user's message can be decoded by the CP at rate at least R_k with vanishing probability of error, i.e.,*

$$\Pr\{(M_1, \ldots, M_L) \neq (\hat{M}_1, \ldots, \hat{M}_L)\} \leq \epsilon. \quad (3)$$

For given $\mathcal{C}_{\mathcal{K}}$, the capacity region $\mathcal{R}(C_{\mathcal{K}})$ is the closure of all achievable rate tuples $(R_1, \ldots, R_L)$.

Due to space limitations, some of the results of this paper are only outlined or given without proofs. The detailed proofs can be found in [16].

A. A Class of Discrete Memoryless Channels

In this work, we establish the capacity region of the following class of discrete memoryless CRAN channels with oblivious relay processing and enabled time-sharing. In this class, the channel outputs at the relay nodes are independent conditionally on the users' inputs. That is,

$$p(y_{\mathcal{K}}^n | x_{\mathcal{L}}^n) = \prod_{i=1}^n \prod_{k=1}^K p(y_{k,i} | x_{\mathcal{L},i}), \quad (4)$$

or, equivalently, the following Markov chain holds,

$$Y_{k,i} \; \text{---} \; X_{\mathcal{L},i} \; \text{---} \; Y_{\mathcal{K}/k,i} \quad \text{for } k \in \mathcal{K} \text{ and } i \in [1, n]. \quad (5)$$

B. Oblivious Relaying with Enabled Time-Sharing

Similar to [6], the above constraint of oblivious relay processing with enabled time-sharing means that, in the absence of information regarding the indices $F_{\mathcal{L}}$ and the messages $M_{\mathcal{L}}$, a codeword $x_l^n(f_l, m_l|q^n)$ taken from a (n, R_l) codebook has independent but non-identically distributed entries.

Lemma 1. *Without the knowledge of the selected codebooks indices $(F_1, \ldots, F_L)$, the distribution of the transmitted codewords conditioned on the time-sharing sequence are given by $p_{X_l^n|Q^n}(x_l^n|q^n) = \prod_{i=1}^n p_{X_l|Q}(x_{l,i}|q_i)$. Consequently, the channel output at relay k, Y_k^n is distributed as*

$$p_{Y_k^n|Q^n}(y_k^n|q^n) = \prod_{i=1}^n \sum_{x_1, \ldots, x_L} p_{Y_k|X_{\mathcal{L}}}(y_{k,i}|x_{\mathcal{L},i}) \prod_{i=1}^L p_{X_l|Q}(x_{l,i}|q_i).$$

III. MAIN RESULTS

A. Capacity Region of the Studied Class of CRAN Channels

The main result of this paper is a single-letter characterization of the capacity region of the class of channels with oblivious relaying and enabled time-sharing that satisfy (5). The following theorem states the result.

Theorem 1. *For the class of discrete memoryless channels given by* (5) *with oblivious relay processing and enabled time-sharing, a rate tuple* $(R_1, \ldots, R_L)$ *is achievable if and only if for all* $\mathcal{T} \subseteq \mathcal{L}$ *and for all* $\mathcal{S} \subseteq \mathcal{K}$, *we have*

$$\sum_{t \in \mathcal{T}} R_t \leq \sum_{s \in \mathcal{S}} [C_s - I(Y_s; U_s | X_{\mathcal{L}}, Q)] \\ + I(X_{\mathcal{T}}; U_{\mathcal{S}^c} | X_{\mathcal{T}^c}, Q), \quad (6)$$

for some joint measure of the form $p(q) \prod_{l=1}^{L} p(x_l | q) \prod_{k=1}^{K} p(y_k | x_{\mathcal{L}}) \prod_{k=1}^{K} p(u_k | y_k, q).$

Proof: The proof of the converse part of Theorem 1 is relegated to Section V. The proof of the direct part can be obtained by applying the noisy network coding (NNC) scheme of [11, Theorem 1]. Alternatively, the rate region of Theorem 1 can also be achieved by a scheme that generalizes that of [7, Theorem 1], which is established in the case of a single transmit node, to the case of multiple users and accommodate time-sharing. By opposition to the NNC scheme, the latter scheme is based on compress-and-forward à la Cover-El Gamal with joint decoding and decompression at the CP (CoF-JD). □

Remark 1. *Key element for the proof of Theorem 1 is the connection with the chief executive officer (CEO) problem. For the case of* m *encoders,* $m \geq 3$, *while the characterization of the optimal rate-distortion region of this problem for general distortion measures has eluded the information theory, a characterization of the optimal region in the case of logarithmic loss distortion measure has been provided recently in [15].*

Remark 2. *In the special case in which* $K = L$ *and the memoryless channel* (4) *is such that* $Y_i = X_i$ *for all* $i = 1, \ldots, K$, *the source coding counter-part of the problem treated in this section reduces to a distributed source coding setting with independent sources (recall that the users input symbols are independent here) under logarithmic loss distortion measure. It is important to note that, for* $K > 2$ *and general (i.e., arbitrarily correlated) sources, the problem poses significant challenges, and is still to be solved, as the Berger-Tung achievability scheme is known to be suboptimal.*

Remark 3. *The sum-rate of Theorem 1 can also be achieved by a scheme in which the CP decodes explicitly the compression indices first, and then decodes the users' transmitted messages, i.e., decompression and decoding is not performed jointly. A similar observation is found in [10, Theorem 2].*

B. Memoryless Gaussian Model

In this section, we consider a memoryless Gaussian MIMO model of the studied CRAN with oblivious relay processing and enabled time sharing. The channel output at relay node k, equipped with M_k receive antennas, is given by

$$\mathbf{Y}_k = \mathbf{H}_{k,\mathcal{L}} \mathbf{X} + \mathbf{N}_k, \quad (7)$$

where $\mathbf{X} \triangleq [\mathbf{X}_1^T, \ldots, \mathbf{X}_L^T]^T$, $\mathbf{X}_l \in \mathbb{C}^{N_l}$ is the channel input vector of user l, $l \in \mathcal{L}$, $\mathbf{H}_{k,\mathcal{L}} \triangleq [\mathbf{H}_{k,1}, \ldots, \mathbf{H}_{k,L}]$, $\mathbf{H}_{k,l} \in \mathbb{C}^{M_k \times N_l}$ is the channel matrix connecting user l to relay node k, and $\mathbf{N}_k \in \mathbb{C}^{M_k}$ is the noise vector at relay node k, assumed to be Gaussian with $\mathbf{N}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_k)$. The transmission is subjected to power constraint $\mathrm{Tr}(\mathbf{K}_l) \leq P_k$, where $\mathbf{K}_l = \mathrm{E}[\mathbf{X}_l \mathbf{X}_l^H]$ is the covariance matrix of $\mathbf{X}_l$. The noises at the relay nodes are assumed to be independent; and so the studied Gaussian model satisfies the Markov chain (5).

Theorem 1 can be extended to continuous channels using standard techniques; and so it characterizes the capacity region of the model (7). The computation of this region, however, is not easy as it requires finding the optimal choices of the involved auxiliary random variables $U_1, \ldots, U_K$. The following theorem characterizes more explicitly the capacity region when the users are constrained to employ Gaussian signaling, i.e., for $Q = q$, $\mathbf{X}_{l,q} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_{l,q})$, for all $l \in \mathcal{L}$.

Theorem 2. *If the input vectors are Gaussian, the capacity region of the memoryless Gaussian MIMO model* (7) *is given by the set of all rate tuples* $(R_1, \ldots, R_L)$ *satisfying that for all* $\mathcal{T} \subseteq \mathcal{L}$ *and all* $\mathcal{S} \subseteq \mathcal{K}$

$$\sum_{t \in \mathcal{T}} R_t \leq \sum_{k \in \mathcal{S}} \left[C_k - \log \frac{|\boldsymbol{\Sigma}_k^{-1}|}{|\boldsymbol{\Sigma}_k^{-1} - \mathbf{B}_k|} \right] \quad (8)$$

$$+ \log \frac{|\sum_{k \in \mathcal{S}^c} \mathbf{H}_{k,\mathcal{T}}^H \mathbf{B}_k \mathbf{H}_{k,\mathcal{T}} + \mathbf{K}_{\mathcal{T}}^{-1}|}{|\mathbf{K}_{\mathcal{T}}^{-1}|}, \quad (9)$$

for some $\mathbf{0} \preceq \mathbf{B}_k \preceq \boldsymbol{\Sigma}_k^{-1}$, *where* $\mathbf{H}_{k,\mathcal{T}}$ *denotes the channel matrix connecting the input* $\mathbf{X}_{\mathcal{T}}$ *to the output* $\mathbf{Y}_k$, *formed by concatenating the matrices* $\mathbf{H}_{k,l}$, $l \in \mathcal{T}$, *horizontally.*

Remark 4. *Theorem 2 extends the result of [5, Theorem 5] to the case of* L *users and enabled time-sharing. In addition to showing that under the constraint of Gaussian input signaling, the quantization codewords can be chosen optimally to be Gaussian, the result of Theorem 2 also means that time-sharing is not needed in the memoryless Gaussian case with NNC and CoF-JD. Recall that, as shown through an example in [5], if the relays are aware of the users' codebooks restricting to Gaussian input signaling can be a severe constraint and is generally suboptimal.*

Remark 5. *In [10], the authors study the questions of optimal fronthaul compression and decoding strategies for uplink CRAN networks without oblivious processing constraints. It is shown that NNC with Gaussian input and Gaussian quantization achieve to within a constant gap of the capacity region of the Gaussian MIMO uplink CRAN. In this paper, we show that if only oblivious relay processing is allowed, NNC and CoF-JD are in fact optimal from a capacity viewpoint.*

IV. GENERAL DISCRETE MEMORYLESS MODEL

In this section, we focus on general discrete memoryless CRAN channels with oblivious relay processing and time sharing, i.e., the channel outputs at the relays are arbitrarily correlated and the Markov chain (5) does not necessarily hold. We establish bounds on the capacity region of the model. The results extend those of [5], which only consider a single transmitter and no time-sharing, to the case of multiple transmitters and allowed time-sharing.

The following theorem provides an inner bound on the capacity region of the general DM CRAN model with oblivious relay processing and time sharing.

Theorem 3. *For general DM CRAN channels with oblivious relay processing and enabled time-sharing, the set of rates $(R_1, \ldots, R_L)$ such that for all $\mathcal{T} \subseteq \mathcal{L}$ and all $\mathcal{S} \subseteq \mathcal{K}$,*

$$\sum_{t \in \mathcal{T}} R_t \leq \sum_{s \in \mathcal{S}} C_s - I(Y_\mathcal{S}; U_\mathcal{S}|X_\mathcal{L}, U_{\mathcal{S}^c}, Q) \tag{10}$$

$$+ I(X_\mathcal{T}; U_{\mathcal{S}^c}|X_{\mathcal{T}^c}, Q), \tag{11}$$

for some joint measure $p(q) \prod_{l=1}^{L} p(x_l|q) p(y_\mathcal{K}|x_\mathcal{L})$ $\cdot \prod_{k=1}^{K} p(u_k|y_k, q)$, is achievable.

We now provide an outer bound on the capacity region of the general DM CRAN model with oblivious relay processing and time-sharing. The following theorem states the result.

Theorem 4. *For general DM CRAN channels with oblivious relay processing and enabled time-sharing, if a rate-tuple $(R_1, \ldots, R_L)$ is achievable then for all $\mathcal{T} \subseteq \mathcal{L}$ and all $\mathcal{S} \subseteq \mathcal{K}$,*

$$\sum_{t \in \mathcal{T}} R_t \leq \sum_{s \in \mathcal{S}} C_s - I(Y_\mathcal{S}; U_\mathcal{S}|X_\mathcal{L}, U_{\mathcal{S}^c}, Q) \tag{12}$$

$$+ I(X_\mathcal{T}; U_{\mathcal{S}^c}|X_{\mathcal{T}^c}, Q), \tag{13}$$

for some $(Q, X_\mathcal{L}, Y_\mathcal{K}, U_\mathcal{K}, W)$ distributed according to $p(q) \prod_{l=1}^{L} p(x_l|q) \ p(y_\mathcal{K}|x_\mathcal{L}) \ p(w|q)$, $u_k = f_k(w, y_k, q)$ for $k = [1, K]$, for some random variable W and some deterministic functions $\{f_k\}$, $k \in \mathcal{K}$.

Remark 6. *The inner bound of Theorem 3 and the outer bound of Theorem 4 do not coincide in general. This is due to the fact that in Theorem 3, $U_1, \ldots, U_K$ satisfy the Markov chain $U_k \multimap (Y_k, Q) \multimap (X_\mathcal{L}, Y_{\mathcal{L}/k}, U_{\mathcal{K}/k})$, whereas in Theorem 4 each U_k is a function of Y_k but also of a "common" random variable W. In particular, the Markov chain $U_k \multimap (Y_k, Q) - \multimap U_{\mathcal{K}/k}$ does not necessarily hold for the auxiliary random variables of the outer bound.*

Remark 7. *As we already mentioned, the class (5) of DM CRAN channels connects with the CEO problem under logarithmic loss distortion measure. The rate-distortion region of this problem is characterized in the excellent contribution [15] for an arbitrary number of (source) encoders (see [15, Theorem 3] therein). For general DM CRAN channels, i.e., without the Markov chain (5) the model connects with the distributed source coding problem under logarithmic loss distortion measure. While a solution of the latter problem for the case of two encoders has been found in [15, Theorem 6], generalizing the result to the case of arbitrary number of encoders poses a significant challenge. In fact, as also mentioned in [15], the Berger-Tung inner bound is known to be generally suboptimal (e.g., see the Korner-Marton lossless modulo-sum problem [17]). Characterizing the capacity region of the general DM CRAN model under the constraint of oblivious relay processing and enabled time-sharing poses a similar challenge, except perhaps for the case of two relay nodes, results on which will be reported elsewhere.*

V. PROOF OF CONVERSE PART OF THEOREM 1

Assume the rate tuple $(R_1, \ldots, R_L)$ is achievable. Let $\mathcal{T}$ be a set of $\mathcal{L}$, $\mathcal{S}$ be a non-empty set of $\mathcal{K}$, and $J_k \triangleq \phi_k^r(Y_k^n, Q^n)$ be the message sent by relay $k \in \mathcal{K}$, and let $\tilde{Q} = Q^n$ be

the time-sharing variable. For simplicity we define $X_\mathcal{L}^n \triangleq (X_1^n, \ldots, X_L^n)$, $R_\mathcal{T} \triangleq \sum_{t \in \mathcal{T}} R_t$ and $C_\mathcal{S} \triangleq \sum_{k \in \mathcal{S}} C_k$. Define $U_{i,k} \triangleq (J_k, Y_k^{i-1})$ and $\bar{Q}_i \triangleq (X_\mathcal{L}^{i-1}, X_{\mathcal{L},i+1}^n, \tilde{Q})$.

From Fano's inequality, we have with $\epsilon_n \to 0$ for $n \to \infty$ (for vanishing probability of error), for all $\mathcal{T} \subseteq \mathcal{L}$,

$$H(m_\mathcal{T}|J_\mathcal{K}, F_\mathcal{L}, \tilde{Q}) \leq H(m_\mathcal{L}|J_\mathcal{K}, F_\mathcal{L}, \tilde{Q}) \leq n\epsilon_n. \tag{14}$$

We show the following inequality, used below in the proof.

$$H(X_\mathcal{T}^n|X_{\mathcal{T}^c}^n, J_\mathcal{K}, \tilde{Q}) \leq \sum_{i=1}^{n} H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, \bar{Q}_i) - nR_\mathcal{T} \tag{15}$$

$$\triangleq n\Gamma_\mathcal{T}. \tag{16}$$

Inequality (16) can be shown as follows.

$$nR_\mathcal{T} = H(m_\mathcal{T}) \tag{17}$$

$$= I(m_\mathcal{T}; J_\mathcal{K}, F_\mathcal{L}, \tilde{Q}) + H(m_\mathcal{T}|J_\mathcal{K}, F_\mathcal{L}, \tilde{Q}) \tag{18}$$

$$= I(m_\mathcal{T}; J_\mathcal{K}, F_\mathcal{T}|F_{\mathcal{T}^c}, \tilde{Q}) + H(m_\mathcal{T}|J_\mathcal{K}, F_\mathcal{L}, \tilde{Q}) \tag{19}$$

$$\leq I(m_\mathcal{T}; J_\mathcal{K}, F_\mathcal{T}|F_{\mathcal{T}^c}, \tilde{Q}) + n\epsilon_n \tag{20}$$

$$= H(J_\mathcal{K}, F_\mathcal{T}|F_{\mathcal{T}^c}, \tilde{Q}) - H(J_\mathcal{K}, F_\mathcal{T}|F_{\mathcal{T}^c}, m_\mathcal{T}, \tilde{Q}) + n\epsilon_n$$

$$= H(J_\mathcal{K}|F_{\mathcal{T}^c}, \tilde{Q}) + H(F_\mathcal{T}|F_{\mathcal{T}^c}, J_\mathcal{K}, \tilde{Q}) + n\epsilon_n \tag{21}$$

$$- H(F_\mathcal{T}|F_{\mathcal{T}^c}, m_\mathcal{T}, \tilde{Q}) - H(J_\mathcal{K}|F_{\mathcal{T}^c}, m_\mathcal{T}, F_\mathcal{T}, \tilde{Q}) \tag{22}$$

$$= I(m_\mathcal{T}, F_\mathcal{T}; J_\mathcal{K}|F_{\mathcal{T}^c}, \tilde{Q}) - I(F_\mathcal{T}; J_\mathcal{K}|F_{\mathcal{T}^c}, \tilde{Q}) + n\epsilon_n \tag{23}$$

$$\leq I(m_\mathcal{T}, F_\mathcal{T}; J_\mathcal{K}|F_{\mathcal{T}^c}, \tilde{Q}) + n\epsilon_n \tag{24}$$

$$\leq I(X_\mathcal{T}^n; J_\mathcal{K}|F_{\mathcal{T}^c}, \tilde{Q}) + n\epsilon_n \tag{25}$$

$$= H(X_\mathcal{T}^n|F_{\mathcal{T}^c}, \tilde{Q}) - H(X_\mathcal{T}^n|F_{\mathcal{T}^c}, J_\mathcal{K}, \tilde{Q}) + n\epsilon_n \tag{26}$$

$$\leq H(X_\mathcal{T}^n|X_{\mathcal{T}^c}^n, \tilde{Q}) - H(X_\mathcal{T}^n|X_{\mathcal{T}^c}^n, F_{\mathcal{T}^c}, J_\mathcal{K}, \tilde{Q}) + n\epsilon_n \tag{27}$$

$$= H(X_\mathcal{T}^n|X_{\mathcal{T}^c}^n, \tilde{Q}) - H(X_\mathcal{T}^n|X_{\mathcal{T}^c}^n, J_\mathcal{K}, \tilde{Q}) + n\epsilon_n, \tag{28}$$

where (17) follows since $m_\mathcal{T}$ are independent; (19) follows since $m_\mathcal{T}$ is independent of $\tilde{Q}$ and $F_{\mathcal{T}^c}$; (20) follows from (14); (23) follows since $m_\mathcal{T}$ is independent of $F_\mathcal{L}$; (25) follows from the data processing inequality;(27) follows since $X_{\mathcal{T}^c}^n, F_{\mathcal{T}^c}$ are independent from $X_\mathcal{T}^n$ and since conditioning reduces entropy and; (28) follows due to the Markov chain

$$X_\mathcal{T}^n \multimap (X_{\mathcal{T}^c}^n, J_\mathcal{K}, \tilde{Q}) \multimap F_{\mathcal{T}^c}. \tag{29}$$

Then, from (28) we have (16) as follows:

$$H(X_\mathcal{T}^n|X_{\mathcal{T}^c}^n, J_\mathcal{K}, \tilde{Q}) \tag{30}$$

$$\leq \sum_{i=1}^{n} H(X_{\mathcal{T},i}|X_{\mathcal{T}^c}^n, X_\mathcal{T}^{i-1}, \tilde{Q}) - nR_\mathcal{T} \tag{31}$$

$$= \sum_{i=1}^{n} H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, X_\mathcal{L}^{i-1}, X_{\mathcal{L},i+1}^n, \tilde{Q}) - nR_\mathcal{T} \tag{32}$$

$$= \sum_{i=1}^{n} H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, \bar{Q}_i) - nR_\mathcal{T} = n\Gamma_\mathcal{T}, \tag{33}$$

where (32) is due to Lemma 1.

Continuing from (28), we have

$$nR_\mathcal{T} \leq \sum_{i=1}^{n} H(X_{\mathcal{T},i}|X_{\mathcal{T}^c}^n, \tilde{Q}, X_\mathcal{T}^{i-1})$$

$$- H(X_{\mathcal{T},i}^n|X_{\mathcal{T}^c}^n, J_\mathcal{K}, X_\mathcal{T}^{i-1}, \tilde{Q}) + n\epsilon_n \tag{34}$$

$$= \sum_{i=1}^{n} H(X_{\mathcal{T},i}|X_{\mathcal{T}^c}^n, \tilde{Q}, X_\mathcal{T}^{i-1}, X_{\mathcal{T},i+1}^n)$$

$$- H(X_{\mathcal{T},i}|X_{\mathcal{T}^c}^n, J_{\mathcal{K}}, X_{\mathcal{T}}^{i-1}, \tilde{Q}) + n\epsilon_n \tag{35}$$

$$\leq \sum_{i=1}^n H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, \bar{Q}_i)$$
$$- H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, U_{\mathcal{K},i}, \bar{Q}_i) + n\epsilon_n \tag{36}$$

$$= \sum_{i=1}^n I(X_{\mathcal{T},i}; U_{\mathcal{K},i}|X_{\mathcal{T}^c,i}, \bar{Q}_i) + n\epsilon_n, \tag{37}$$

where (35) follows due to Lemma 1; and (36) follows since conditioning reduces entropy.

On the other hand, we have the following equality

$$I(Y_{\mathcal{S}}^n; J_{\mathcal{S}}|X_{\mathcal{L}}^n, J_{\mathcal{S}^c}, \tilde{Q}) = \sum_{k\in\mathcal{S}} I(Y_k^n; J_k|X_{\mathcal{L}}^n, \tilde{Q}) \tag{38}$$

$$= \sum_{k\in\mathcal{S}} \sum_{i=1}^n I(Y_{k,i}; J_k|X_{\mathcal{L}}^n, Y_k^{i-1}, \tilde{Q}) \tag{39}$$

$$= \sum_{k\in\mathcal{S}} \sum_{i=1}^n I(Y_{k,i}; J_k, Y_k^{i-1}|X_{\mathcal{L}}^n, \tilde{Q}) \tag{40}$$

$$= \sum_{k\in\mathcal{S}} \sum_{i=1}^n I(Y_{k,i}; U_{k,i}|X_{\mathcal{L},i}, \bar{Q}_i), \tag{41}$$

where (38) follows due to the Markov chain

$$J_k \; \circlearrowleft \; Y_k^n \; \circlearrowleft \; X_{\mathcal{L}}^n \; \circlearrowleft \; Y_{\mathcal{S}\setminus k}^n \; \circlearrowleft \; J_{\mathcal{S}\setminus k} \quad \text{for } k = [1, K], \tag{42}$$

and since J_k is a function of of Y_k^n; and (40) follows due to the Markov chain $Y_{k,i} - X_{\mathcal{L}}^n - Y_k^{i-1}$ which follows since the channel is memoryless.

Then, from the relay side we have,

$$nC_{\mathcal{S}} \geq \sum_{k\in\mathcal{S}} H(J_k) \geq H(J_{\mathcal{S}}) \tag{43}$$

$$\geq I(Y_{\mathcal{S}}^n; J_{\mathcal{S}}|X_{\mathcal{T}^c}^n, J_{\mathcal{S}^c}, \tilde{Q}) \tag{44}$$

$$= I(X_{\mathcal{T}}^n, Y_{\mathcal{S}}^n; J_{\mathcal{S}}|X_{\mathcal{T}^c}^n, J_{\mathcal{S}^c}, \tilde{Q}) \tag{45}$$

$$= H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, J_{\mathcal{S}^c}, \tilde{Q}) - H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, J_{\mathcal{K}}, \tilde{Q})$$
$$+ I(Y_{\mathcal{S}}^n; J_{\mathcal{S}}|X_{\mathcal{L}}^n, J_{\mathcal{S}^c}, \tilde{Q}) \tag{46}$$

$$\geq H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, J_{\mathcal{S}^c}, \tilde{Q}) - n\Gamma_{\mathcal{T}}$$
$$+ I(Y_{\mathcal{S}}^n; J_{\mathcal{S}}|X_{\mathcal{L}}^n, J_{\mathcal{S}^c}, \tilde{Q}) \tag{47}$$

$$= \sum_{i=1}^n H(X_{\mathcal{T},i}|X_{\mathcal{T}^c}^n, J_{\mathcal{S}^c}, X_{\mathcal{T}}^{i-1}, \tilde{Q}) - n\Gamma_{\mathcal{T}}$$
$$+ I(Y_{\mathcal{S}}^n; J_{\mathcal{S}}|X_{\mathcal{L}}^n, J_{\mathcal{S}^c}, \tilde{Q}) \tag{48}$$

$$\geq \sum_{i=1}^n H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, U_{\mathcal{S}^c,i}, \bar{Q}_i) - n\Gamma_{\mathcal{T}}$$
$$+ I(Y_{\mathcal{S}}^n; J_{\mathcal{S}}|X_{\mathcal{L}}^n, J_{\mathcal{S}^c}, \tilde{Q}) \tag{49}$$

$$= nR_{\mathcal{T}} - \sum_{i=1}^n I(X_{\mathcal{T},i}; U_{\mathcal{S}^c,i}|X_{\mathcal{T}^c,i}, \bar{Q}_i)$$
$$+ \sum_{k\in\mathcal{S}} \sum_{i=1}^n I(Y_{k,i}; U_{k,i}|X_{\mathcal{L},i}, \bar{Q}_i), \tag{50}$$

where (45) follows since $J_{\mathcal{S}}$ is a function of $Y_{\mathcal{S}}^n$; (47) follows from (16); (49) follows since conditioning reduces entropy; and (50) follows from (16) and (41).

In general, $\bar{Q}_i$ is not independent of $X_{\mathcal{L},i}, Y_{\mathcal{S},i}$, and that due to Lemma 1, conditioned on $\bar{Q}_i$, we have the Markov chain

$$U_{k,i} - Y_{k,i} - X_{\mathcal{L},i} - Y_{\mathcal{K}\setminus k,i} - U_{\mathcal{K}\setminus k,i}. \tag{51}$$

VI. Concluding Remarks

We study transmission over a cloud radio access network under the framework of oblivious processing at the relays, i.e., the relays are not allowed to know, or cannot acquire, the users' codebooks. Our results shed light (and sometimes determine exactly) what operations the relay nodes should perform optimally in this case. In particular, perhaps non-surprisingly, it is shown that compress-and-forward, or variants of it, generally perform well in this case, and are optimal when the outputs at the relay nodes are conditionally independent on the users inputs. Furthermore, in addition to its relevance from a practical viewpoint, restricting the relays not to know/utilize the users' codebooks causes only a bounded rate loss in comparison with the non-oblivious setting (e.g., compress-and-forward and noisy network coding perform to within a constant gap from the cut-set bound in the Gaussian case).

References

[1] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: insights and challenges," *IEEE Wireless Communications*, vol. 22, no. 2, pp. 152–160, Apr. 2015.

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

[3] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.

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

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

[6] O. Simeone, E. Erkip, and S. Shamai, "On codebook information for interference relay channels with out-of-band relaying," *IEEE Trans. Inf. Theory,*, vol. 57, no. 5, pp. 2880–2888, May 2011.

[7] A. Sanderovich, O. Somekh, H. 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.

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

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

[10] Y. Zhou, Y. Xu, W. Yu, and J. Chen, "On the optimal fronthaul compression and decoding strategies for uplink cloud radio access networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7402–7418, Dec. 2016.

[11] S. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*.

[12] I. Estella and A. Zaidi, "Lossy compression for compute-and-forward in limited backhaul uplink multicell processing," *IEEE Trans. Comm.*, vol. PP, no. 99, pp. 1–1, 2016.

[13] A. Dytso, D. Tuninetti, and N. Devroye, "On discrete alphabets for the two-user gaussian interference channel with one receiver lacking knowledge of the interfering codebook," in *Information Theory and Applications Workshop (ITA), 2014*, Feb. 2014, pp. 1–8.

[14] Y. Tian and A. Yener, "Relaying for multiple sources in the absence of codebook information," in *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Nov. 2011, pp. 1845–1849.

[15] 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.

[16] I. Estella, A. Zaidi, G. Caire, and S. Shamai, "On the capacity of cloud radio access networks with oblivious relaying," to be submitted to *IEEE Transactions on Information Theory*, 2017.

[17] J. Korner and K. Marton, "How to encode the modulo-two sum of binary sources (corresp.)," *IEEE Transactions on Information Theory*, vol. 25, no. 2, pp. 219–221, Mar 1979.