Uplink Oblivious Cloud Radio Access Networks: An Information Theoretic Overview

Abdellatif Zaidi[†‡] Iñaki Estella Aguerri[†] Giuseppe Caire[⋆] Shlomo Shamai (Shitz)[∗]
[†] Paris 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
{abdellatif.zaidi@u-pem.fr, inaki.estella@huawei.com, caire@tu-berlin.de, sshlomo@ee.technion.ac.il}

I. Introduction

In this work, we consider transmission over a Cloud Radio Access Network (CRAN) in which the relay nodes (radio units) are constrained to operate without knowledge of the users' codebooks, i.e., are oblivious (nomadic), and only know time- or frequency-sharing protocols. The model is shown in Figure 1. 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 both relaying à-la Cover-El Gamal [1], i.e., compress-and-forward with joint decompression and decoding, as suggested in [2], are optimal. This is equivalent to noisy notwork coding [3]. 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 [4]. For memoryless Gaussian channels, we provide a full characterization of the capacity region under Gaussian signaling, i.e., when the users' channel inputs are restricted to be Gaussian. In doing so, we also discuss the suboptimality of separate decompression-decoding and the role of time-sharing. Furthermore, we elaborate on meaningful connections with the problem of distributed information bottleneck problem [5]–[7]. Finally, we evaluate and compare the performance of some oblivious, including the recent scheme [8], and non-oblivious schemes, such as [9] and [10], and cut-set bounds.

II. CRAN System Model and Main Results

Consider the discrete memoryless (DM) CRAN model shown in Figure 1. In this model, L users communicate with a common destination or central processor (CP) through K relay nodes, where $L \geq 1$ and $K \geq 1$. Relay node k, $1 \leq k \leq K$, is connected to the CP via an error-free finite-rate fronthaul link of capacity C_k. In what follows, we let $\mathcal{L} := [1\!:\!L]$ and $\mathcal{K} := [1\!:\!K]$ indicate the set of users and relays, respectively. Similar to [11], the relay nodes are constrained to operate without knowledge of the users' codebooks and only know a time-sharing sequence Q^n, i.e., a set of time instants at which users switch among different codebooks. The obliviousness of the relay nodes to the actual codebooks of the users is modeled via the notion of *randomized encoding* [2], [12]. That is, users or transmitters select their codebooks at random and the relay nodes are *not* informed about the currently selected codebooks, while the CP is given such information.

The work of S. Shamai has been supported by the European Union's Horizon 2020 Research And Innovation Programme, grant agreement no. 694630.

A. Capacity Region of a Class of CRANs

Consider the following class of DM CRANs in which the channel outputs at the relay nodes are independent conditionally on the users' inputs. That is, for all $k \in \mathcal{K}$ and all $i \in [1\!:\!n]$,

$$Y_{k,i} -\!\!\circ\!\!- X_{\mathcal{L},i} -\!\!\circ\!\!- Y_{\mathcal{K}/k,i} \tag{1}$$

forms a Markov chain in this order.

Theorem 1 ([6]). *For the class of DM CRANs with oblivious relay processing and enabled time-sharing for which* (1) *holds, the capacity region $\mathcal{C}(C_\mathcal{K})$ is given by the union of all rate tuples $(R_1, \ldots, R_L)$ which satisfy*

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

for all non-empty subsets $\mathcal{T} \subseteq \mathcal{L}$ and all $\mathcal{S} \subseteq \mathcal{K}$, for some joint measure of the form

$$p(q)\prod_{l=1}^{L} p(x_l|q)\prod_{k=1}^{K} p(y_k|x_\mathcal{L})\prod_{k=1}^{K} p(u_k|y_k, q). \tag{2}$$

The direct part of Theorem 1 can be obtained by a coding scheme in which each relay node compresses its channel output by using Wyner-Ziv binning to exploit the correlation with the channel outputs at the other relays, and forwards the bin index to the CP over its rate-limited link. The CP jointly decodes the compression indices (within the corresponding bins) and the transmitted messages, i.e., Cover-El Gamal compress-and-forward [1, Theorem 3] with joint decompression and decoding (CF-JD). Alternatively, the rate region of Theorem 1 can also be obtained by a direct application of the noisy network coding (NNC) scheme of [3, Theorem 1].

Remark 1. *The fact that the two operations of decompression and decoding are performed jointly in the scheme CF-JD is critical to achieve the full rate-region of Theorem 1, in the sense that if the CP first jointly decodes the compression indices and then jointly decodes the users' messages, i.e., the two operations are performed successively, this results in a a region that is generally strictly suboptimal. A similar observation can be found in [5], [6], [13].*

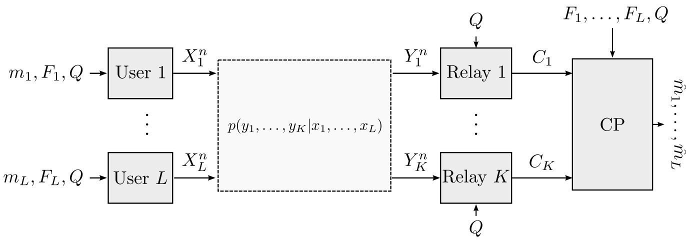

Fig. 1. CRAN model with oblivious relaying and time-sharing.

B. Memoryless MIMO Gaussian CRAN

Consider a memoryless Gaussian MIMO CRAN with oblivious relay processing and enabled time-sharing in which relay node k, $k \in \mathcal{K}$, is equipped with M_k receive antennas and has output

$$\mathbf{Y}_k = \mathbf{H}_{k,\mathcal{L}}\mathbf{X} + \mathbf{N}_k, \tag{3}$$

where $\mathbf{X} := [\mathbf{X}_1^T, \dots, \mathbf{X}_L^T]^T$, $\mathbf{X}_l \in \mathbb{C}^{N_l}$ is the channel input vector of user $l \in \mathcal{L}$, N_l is the number of antennas at user l, $\mathbf{H}_{k,\mathcal{L}} := [\mathbf{H}_{k,1}, \dots, \mathbf{H}_{k,L}]$ is the matrix obtained by concatenating the $\mathbf{H}_{k,l}$, $l \in \mathcal{L}$, horizontally, with $\mathbf{H}_{k,l} \in \mathbb{C}^{M_k \times N_l}$ being 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 memoryless Gaussian with covariance matrix $\mathbf{N}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma}_k)$ and independent from other noises and from the channel inputs $\{\mathbf{X}_l\}$. The transmission from user $l \in \mathcal{L}$ is subjected to the following covariance constraint,

$$\mathrm{E}[\mathbf{X}_l\mathbf{X}_l^H] \preceq \mathbf{K}_l, \tag{4}$$

where $\mathbf{K}_l$ is a given $N_l \times N_l$ positive semi-definite matrix, and the notation $\preceq$ indicates that the matrix $(\mathbf{K}_l - \mathrm{E}[\mathbf{X}_l\mathbf{X}_l^H])$ is positive semi-definite.

1) Capacity Region under Time-Sharing of Gaussian Inputs: Let, for all $l \in \mathcal{L}$, the input $\mathbf{X}_l$ be restricted to be distributed such that for all $Q = q$,

$$\mathbf{X}_l | Q = q \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_{l,q}), \tag{5}$$

where the matrices $\{\mathbf{K}_{l,q}\}_{q=1}^{|\mathcal{Q}|}$ are chosen to satisfy

$$\sum_{q \in \mathcal{Q}} p_Q(q)\mathbf{K}_{l,q} \preceq \mathbf{K}_l. \tag{6}$$

Theorem 2 ([6]). *The capacity region $\mathcal{C}_{\mathrm{G}}(C_{\mathcal{K}})$ of the memoryless Gaussian MIMO model with oblivious relay processing described by* (3) *and* (4) *under time-sharing of Gaussian inputs is given by the set of all rate tuples $(R_1, \dots, R_L)$ that satisfy*

$$\sum_{t \in \mathcal{T}} R_t \le \sum_{k \in \mathcal{S}} \left[C_k - \mathrm{E}_Q\left[\log \frac{|\mathbf{\Sigma}_k^{-1}|}{|\mathbf{\Sigma}_k^{-1} - \mathbf{B}_{k,Q}|} \right] \right] + \mathrm{E}_Q\left[\log \frac{|\sum_{k \in \mathcal{S}^c} \mathbf{H}_{k,\mathcal{T}}^H \mathbf{B}_{k,Q} \mathbf{H}_{k,\mathcal{T}} + \mathbf{K}_{\mathcal{T},Q}^{-1}|}{|\mathbf{K}_{\mathcal{T},Q}^{-1}|} \right], \tag{7}$$

for all non-empty $\mathcal{T} \subseteq \mathcal{L}$ and all $\mathcal{S} \subseteq \mathcal{K}$, for some pmf $p_Q(q)$ and matrices $\mathbf{K}_{q,l}$ and $\mathbf{B}_{k,q}$ such that $\mathrm{E}_Q[\mathbf{K}_{l,Q}] \preceq \mathbf{K}_l$ and $\mathbf{0} \preceq \mathbf{B}_{k,q} \preceq \mathbf{\Sigma}_k^{-1}$; and where, for $q \in Q$ and $\mathcal{T} \subseteq \mathcal{L}$, the matrix $\mathbf{K}_{\mathcal{T},q}$ is defined as $\mathbf{K}_{\mathcal{T},q} := diag[\{\mathbf{K}_{t,q}\}_{t \in \mathcal{T}}]$.

2) Price of Non-Awareness: Bounded Rate Loss: The uplink of a CRAN relates to diammond channels; and, so, its capacity remains to be found in general.

In general (i.e., without restricting to oblivious relay processing), the uplink of a CRAN relates to diammond channels [14]; and, so, its capacity remains to be found. Restricting the relay nodes not to know/utilize the users' codebooks causes only a bounded rate loss in comparison with maximum rate that would be achievable in the non-oblivious setting. The result is obtained through comparison with the cut-set bound.

Theorem 3 ([6]). *If $(R_1, \dots, R_L) \in \mathcal{C}^{uncons}(C_{\mathcal{K}})$, then there exists a constant $\Delta \ge 0$ such that $(R_1 - \Delta, \dots, R_L - \Delta) \in \mathcal{C}_G(C_{\mathcal{K}})$, with*

$$\Delta \le \begin{cases} \frac{N}{2}(2.45 + \log(\frac{KM}{N})), & \textit{for } KM > 2N, \\ \frac{KM+N}{2} & \textit{for } KM \le 2N. \end{cases} \tag{8}$$

III. Connections

A. Distributed Source Coding under Logarithmic Loss

Key element to the proof of the converse part of Theorem 1 is the connection with the Chief Executive Officer (CEO) source coding problem. For the case of $K \ge 2$ encoders, while the characterization of the optimal rate-distortion region of this problem for general distortion measures has eluded the information theory for now more than four decades, a characterization of the optimal region in the case of logarithmic loss distortion measure has been provided recently in [4]. A key step in [4] is that the log-loss distortion measure admits a lower bound in the form of the entropy of the source conditioned on the decoders input. Leveraging on this result, in our converse proof of Theorem 1 we derive a single letter upper-bound on the entropy of the channel inputs conditioned on the indices $J_{\mathcal{K}}$ that are sent by the relays, in the absence of knowledge of the codebooks indices $F_{\mathcal{L}}$. Also, the rate region of the vector Gaussian CEO problem under logarithmic loss distortion measure has been found recently in [15].

B. Distributed Information Bottleneck Method

The information bottleneck (IB) method was introduced by Tishby [16] as an information-theoretic principle for extracting the relevant information that some signal $Y \in \mathcal{Y}$ provides about another one, $X \in \mathcal{X}$, that is of interest. Beyond data classification or clustering and prediction, it has found remarkable applications in the wider field of learning [17]. Perhaps key to the analysis, and development, of the IB method is the elegant connection it has with information-theoretic rate-distortion problems. Specifically, the IB problem is essentially a remote (or indirect) source coding problem [18] with logarithmic loss distortion measure [4]; and prior information-theoretic art such as [19] and [20] is relevant for it. Also, information combining problems, as used in [21] along with the I-MMSE relation, are relevant to prove simply the Tishby Gaussian bottleneck result. Furthermore, the IB method is also related to the problem of Wyner-Ziv coding with common reconstruction [22]. Distributed architectures for IB are also studied; and the optimal rate-information region of distributed IB is developed for discrete and Gaussian sources in [7]. Extension of the IB method to general alphabets can be found in [23].

IV. Outlook

Among interesting problems that are left unaddressed in this paper that of characterizing optimal input distributions under rate-constrained compression at the relays where, e.g., discrete signaling is already known to sometimes outperform Gaussian signaling for single-user Gaussian CRAN [2]. It is conjectured that the optimal input distribution is discrete. Other issues might relate to extensions to continuous time filtered Gaussian channels, in parallel to the regular bottleneck problem [24], or extensions to settings in which fronthauls may be not available at some radio-units, and that is unknown to the systems. That is, the more radio units are connected to the central unit the higher rate could be conveyed over the CRAN uplink [25]. Alternatively, one may consider finding the worst-case noise under given input distributions, e.g., Gaussian, and rate-constrained compression at the relays. Finally, there are interesting aspects that address processing constraints of continuous waveforms, e.g., such as sampling at a given rate [26], [27] with focus on remote logarithmic distortion [4], which in turn boils down to the distributed bottleneck problem [7], [15].

References

[1] 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.

[2] 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.

[3] S. 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, 2011.

[4] 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.

[5] I. E. Aguerri, A. Zaidi, G. Caire, and S. S. (Shitz), "On the capacity of uplink cloud radio access networks with oblivious relaying," in *Proc. of IEEE Int. Symp. on Info. Theory*, Aachen, Germany, 2017.

[6] ——, "On the capacity of uplink cloud radio access networks with oblivious relaying," *IEEE Trans. on Info. Theory. Available at arxiv.org/abs/1710.09275*, 2017.

[7] I. E. Aguerri and A. Zaidi, "Distributed information bottleneck method for discrete and Gaussian sources," in *Proc. of Int. Zurich Seminar on Information and Communication, IZS*, Zurich, Switzerland, 2018.

[8] I. E. Bakoury and B. Nazer, "Integer-forcing architectures for uplink cloud radio access networks," in *IEEE Info. Theory Workshop, ITW*, Kaohsiung, 2017.

[9] I. E. Aguerri and A. Zaidi, "Lossy compression for compute-and-forward in limited backhaul uplink multicell processing," *IEEE Transactions on Communications*, vol. 64, no. 12, pp. 5227–5238, Dec 2016.

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

[11] 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.

[12] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. on Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct 1998.

[13] 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.

[14] B. Schein and R. G. Gallager, "The Gaussian parallel relay network," in *Proc. IEEE Int. Symp. Information Theory*, Sorrento, Italy, Jun. 2000, p. 22.

[15] I. E. Aguerri, A. Zaidi, and S. S. (Shitz), "Rate distortion region of the vector Gaussian CEO problem under logarithmic loss," in *Proc. of IEEE Int. Symp. on Info. Theory. Submitted for publication*, Vail, Colorado, 2018.

[16] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in *Proc. 37th Annual Allerton Conf. on Comm., Control, and Computing*, 1999, pp. 368–377.

[17] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning and games*. New York, USA: Cambridge, Univ. Press, 2006.

[18] H.-S. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. on Info. Theory*, vol. IT-26, pp. 518–521, Sep. 1980.

[19] R.-L. Dobrushin and B. S. Tsybakov, "Information transmission with additional noise," *IRE Trans. on Info. Theory*, vol. 85, pp. 293–304, 1962.

[20] H. Witsenhausen and A. Wyner, "A conditional entropy bound for a pair of discrete random variables," *IEEE Trans. on Information Theory*, vol. 21, no. 5, pp. 493–501, Sep. 1975.

[21] D. Guo, S. Shamai, and S. Verdu, "The interplay between information and estimation measures," *Foundations and Trends in Signal Processing*, vol. 6, pp. 243–429, 2013.

[22] Y. Steinberg, "Coding and common reconstruction," *IEEE Trans. Inf. Theory*, vol. IT-11, pp. 4995–5010, Nov. 2009.

[23] G. Pichler and G. Koliander, "Information bottleneck on general alphabets," in *available at arXiv:1801.01050*, 2017.

[24] A. Homri, M. Peleg, and S. S. (Shitz), "Oblivious fronthaul-constrained relay for a gaussian channel," *IEEE Trans. on Communications. Submitted for publication*, 2017.

[25] R. K. O. Simeone and S. Shamai, "Robust uplink communications over fading channels with variable backhaul connectivity," *IEEE Trans. on Communications*, vol. 12, pp. 5788–5799, Nov. 2013.

[26] Y. Chen, A.-J. Goldsmith, , and Y.-C. Eldar, "Channel capacity under sub-nyquist nonuniform sampling," *IEEE Trans. Inf. Theory*, vol. 60, pp. 4739–4756, Aug. 2014.

[27] A. K. ane Y.-C. Eldar and A.-J. Goldsmith, "Analog-to-digital compression: A new paradigm for converting signals to bits," *available at arXiv:1801.06718*, Jan. 2018.