On the Capacity of Uplink Cloud Radio Access Networks with Oblivious Relaying: Proofs

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

[†]France Research Center, Huawei Technologies, Boulogne-Billancourt, 92100, France

[‡] Université Paris-Est, Champs-sur-Marne, 77454, France

$*$ Technion Institute of Technology, Technion City, Haifa 32000, Israel

$\star$ Technische Universitat Berlin, 10587 Berlin, Germany

{inaki.estella@huawei.com, abdellatif.zaidi@u-pem.fr, sshlomo@ee.technion.ac.il, caire@tu-berlin.de}

Abstract

This document contains the proofs of the results provided in the paper "*On the Capacity of Uplink Cloud Radio Access Networks with Oblivious Relaying*".

I. Notation

This document is based on the results exposed in the document "*On the Capacity of Uplink Cloud Radio Access Networks with Oblivious Relaying*". Whenever we shall refer to one equation in such document, we refer to it as [C1]-(eq. number), e.g., [C1]-(5) refers to equation (5) of the referred manuscript.

II. Proof of Theorem 1

A. Direct Part

In this section we show that the capacity in Theorem 1 is achieved by noisy network coding (NNC) [1] and by compress-and-forward à la Cover-El Gamal with joint decoding and decompression at the CP (CoF-JD) as in [2], [3].

In NNC each relays compresses the received signal without applying Wyner-Ziv binning. At the CP, the transmitted messages are simultaneously decoded with the compression codewords. The rates achievable with NNC in O-CRAN with L users and K relays follows from [1] and are given next.

Proposition 1. *The tuple $(R_1,\dots,R_L)$ is achievable in the O-CRAN if for all $\mathcal{T}\subseteq\mathcal{L}$, and for all $\mathcal{S}\subseteq\mathcal{K}$,*

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

is satisfied for some pmf $p(q)\prod_{l=1}^{L}p(x_l|q)\prod_{k=1}^{K}p(u_k|y_k,q)$.

Alternatively to NNC, the rates region in Proposition 1 can also be achieved with CoF-JD, in which the relays compress the received signal relying on binning distributed compression techniques as stated in [2, Proposition IV.1][1]. While in the classical compress-and-forward, the CP first decodes the relay's observation, and then decodes the transmitted messages could be modified to consider the joint decoding and decompression. Such modification is motivated by the observation that by considering joint decoding, the transmitted messages can be decoded without fully decoding the compression [3].

B. Converse Part

Assume the rate tuple $(R_1,\dots,R_L)$ 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

[1]Note that [2, Proposition IV.1] requires a slight generalization to include L users and N relays and the time-sharing variable in Proposition 1.

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 $\mathcal{T} \subseteq \mathcal{L}$,

$$H(m_{\mathcal{T}}|J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) \leq H(m_{\mathcal{L}}|J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) \leq n\epsilon_n. \tag{1}$$

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

$$\begin{aligned} H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, J_{\mathcal{K}}, \tilde{Q}) &\leq \sum_{i=1}^n H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, \bar{Q}_i) - nR_{\mathcal{T}} &(2)\\ &\triangleq n\Gamma_{\mathcal{T}}. &(3)\end{aligned}$$

Inequality (3) can be shown as follows. From the destination side, we have

$$\begin{aligned}
nR_{\mathcal{T}} &= H(m_{\mathcal{T}}) &(4)\\
&=I(m_{\mathcal{T}}; J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) + H(m_{\mathcal{T}}|J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) &(5)\\
&=I(m_{\mathcal{T}}; J_{\mathcal{K}}, F_{\mathcal{T}}|F_{\mathcal{T}^c}, \tilde{Q}) + H(m_{\mathcal{T}}|J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) &(6)\\
&\leq I(m_{\mathcal{T}}; J_{\mathcal{K}}, F_{\mathcal{T}}|F_{\mathcal{T}^c}, \tilde{Q}) + n\epsilon_n &(7)\\
&=H(J_{\mathcal{K}}, F_{\mathcal{T}}|F_{\mathcal{T}^c}, \tilde{Q}) - H(J_{\mathcal{K}}, F_{\mathcal{T}}|F_{\mathcal{T}^c}, m_{\mathcal{T}}, \tilde{Q}) + n\epsilon_n &(8)\\
&=H(J_{\mathcal{K}}|F_{\mathcal{T}^c}, \tilde{Q}) + H(F_{\mathcal{T}}|F_{\mathcal{T}^c}, J_{\mathcal{K}}, \tilde{Q}) + n\epsilon_n &(9)\\
&\quad - H(F_{\mathcal{T}}|F_{\mathcal{T}^c}, m_{\mathcal{T}}, \tilde{Q}) - H(J_{\mathcal{K}}|F_{\mathcal{T}^c}, m_{\mathcal{T}}, F_{\mathcal{T}}, \tilde{Q}) &(10)\\
&=I(m_{\mathcal{T}}, F_{\mathcal{T}}; J_{\mathcal{K}}|F_{\mathcal{T}^c}, \tilde{Q}) - I(F_{\mathcal{T}}; J_{\mathcal{K}}|F_{\mathcal{T}^c}, \tilde{Q}) + n\epsilon_n &(11)\\
&\leq I(m_{\mathcal{T}}, F_{\mathcal{T}}; J_{\mathcal{K}}|F_{\mathcal{T}^c}, \tilde{Q}) + n\epsilon_n &(12)\\
&\leq I(X_{\mathcal{T}}^n; J_{\mathcal{K}}|F_{\mathcal{T}^c}, \tilde{Q}) + n\epsilon_n &(13)\\
&=H(X_{\mathcal{T}}^n|F_{\mathcal{T}^c}, \tilde{Q}) - H(X_{\mathcal{T}}^n|F_{\mathcal{T}^c}, J_{\mathcal{K}}, \tilde{Q}) + n\epsilon_n &(14)\\
&\leq H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, \tilde{Q}) - H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, F_{\mathcal{T}^c}, J_{\mathcal{K}}, \tilde{Q}) + n\epsilon_n &(15)\\
&=H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, \tilde{Q}) - H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, J_{\mathcal{K}}, \tilde{Q}) + n\epsilon_n, &(16)
\end{aligned}$$

where (4) follows since $m_{\mathcal{T}}$ are independent; (6) follows since $m_{\mathcal{T}}$ is independent of $\tilde{Q}$ and $F_{\mathcal{T}^c}$; (7) follows from (1); (11) follows since $m_{\mathcal{T}}$ is independent of $F_{\mathcal{L}}$; (13) follows from the data processing inequality;(15) follows since $X^n_{\mathcal{T}^c}, F_{\mathcal{T}^c}$ are independent from $X^n_{\mathcal{T}}$ and since conditioning reduces entropy and; (16) follows due to the Markov chain

$$X^n_{\mathcal{T}} -\!\!\circ\!\!- (X^n_{\mathcal{T}^c}, J_{\mathcal{K}}, \tilde{Q}) -\!\!\circ\!\!- F_{\mathcal{T}^c}. \tag{17}$$

Then, from (16) we have (3) as follows:

$$\begin{aligned}
H(X^n_{\mathcal{T}}|X^n_{\mathcal{T}^c}, J_{\mathcal{K}}, \tilde{Q}) &\leq H(X^n_{\mathcal{T}}|X^n_{\mathcal{T}^c}, \tilde{Q}) - nR_{\mathcal{T}} - n\epsilon_n &(18)\\
&\leq \sum_{i=1}^{n} H(X_{\mathcal{T},i}|X^n_{\mathcal{T}^c}, X^{i-1}_{\mathcal{T}}, \tilde{Q}) - nR_{\mathcal{T}} &(19)\\
&= \sum_{i=1}^{n} H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, X^{i-1}_{\mathcal{L}}, X^n_{\mathcal{L},i+1}, \tilde{Q}) - nR_{\mathcal{T}} &(20)\\
&= \sum_{i=1}^{n} H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, \bar{Q}_i) - nR_{\mathcal{T}} = n\Gamma_{\mathcal{T}}. &(21)
\end{aligned}$$

where (20) is due to Lemma 1 .

Continuing from (16), we have

$$\begin{aligned}
nR_{\mathcal{T}} &\leq \sum_{i=1}^{n} H(X_{\mathcal{T},i}|X^n_{\mathcal{T}^c}, \tilde{Q}, X^{i-1}_{\mathcal{T}}) - H(X^n_{\mathcal{T},i}|X^n_{\mathcal{T}^c}, J_{\mathcal{K}}, X^{i-1}_{\mathcal{T}}, \tilde{Q}) + n\epsilon_n &(22)\\
&= \sum_{i=1}^{n} H(X_{\mathcal{T},i}|X^n_{\mathcal{T}^c}, \tilde{Q}, X^{i-1}_{\mathcal{T}}, X^n_{\mathcal{T},i+1}) - H(X_{\mathcal{T},i}|X^n_{\mathcal{T}^c}, J_{\mathcal{K}}, X^{i-1}_{\mathcal{T}}, \tilde{Q}) + n\epsilon_n &(23)\\
&\leq \sum_{i=1}^{n} H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, \bar{Q}_i) - H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, U_{\mathcal{K},i}, \bar{Q}_i) + n\epsilon_n &(24)\\
&= \sum_{i=1}^{n} I(X_{\mathcal{T},i}; U_{\mathcal{K},i}|X_{\mathcal{T}^c,i}, \bar{Q}_i) + n\epsilon_n, &(25)
\end{aligned}$$

where (23) follows due to Lemma 1 ; and (24) follows since conditioning reduces entropy.

On the other hand, we have the following equality

$$I(Y^n_{\mathcal{S}}; J_{\mathcal{S}}|X^n_{\mathcal{L}}, J_{\mathcal{S}^c}, \tilde{Q}) = \sum_{k\in\mathcal{S}} I(Y^n_k; J_k|X^n_{\mathcal{L}}, \tilde{Q}) \tag{26}$$

$$= \sum_{k\in\mathcal{S}} \sum_{i=1}^{n} I(Y_{k,i}; J_k | X^n_{\mathcal{L}}, Y^{i-1}_k, \tilde{Q}) \tag{27}$$

$$= \sum_{k\in\mathcal{S}} \sum_{i=1}^{n} I(Y_{k,i}; J_k, Y^{i-1}_k | X^n_{\mathcal{L}}, \tilde{Q}) \tag{28}$$

$$= \sum_{k\in\mathcal{S}} \sum_{i=1}^{n} I(Y_{k,i}; U_{k,i} | X_{\mathcal{L},i}, \bar{Q}_i), \tag{29}$$

where (26) follows due to the Markov chain

$$J_k -\!\!\circ\!\!- Y^n_k -\!\!\circ\!\!- X^n_{\mathcal{L}} -\!\!\circ\!\!- Y^n_{\mathcal{S}\backslash k} -\!\!\circ\!\!- J_{\mathcal{S}\backslash k} \quad \text{for } k = [1, K], \tag{30}$$

and since J_k is a function of of Y^n_k; and (28) follows due to the Markov chain $Y_{k,i} - X^n_{\mathcal{L}} - Y^{i-1}_k$ which follows since the channel is memoryless.

Then, from the relay side we have,

$$\begin{aligned}
nC_{\mathcal{S}} \geq& \sum_{k\in\mathcal{S}} H(J_k) \geq H(J_{\mathcal{S}}) &(31)\\
\geq& H(J_{\mathcal{S}} | X^n_{\mathcal{T}^c}, J_{\mathcal{S}^c}, \tilde{Q}) &(32)\\
\geq& I(Y^n_{\mathcal{S}}; J_{\mathcal{S}} | X^n_{\mathcal{T}^c}, J_{\mathcal{S}^c}, \tilde{Q}) &(33)\\
=& I(X^n_{\mathcal{T}}, Y^n_{\mathcal{S}}; J_{\mathcal{S}} | X^n_{\mathcal{T}^c}, J_{\mathcal{S}^c}, \tilde{Q}) &(34)\\
=& H(X^n_{\mathcal{T}} | X^n_{\mathcal{T}^c}, J_{\mathcal{S}^c}, \tilde{Q}) - H(X^n_{\mathcal{T}} | X^n_{\mathcal{T}^c}, J_{\mathcal{K}}, \tilde{Q}) + I(Y^n_{\mathcal{S}}; J_{\mathcal{S}} | X^n_{\mathcal{L}}, J_{\mathcal{S}^c}, \tilde{Q}) &(35)\\
\geq& H(X^n_{\mathcal{T}} | X^n_{\mathcal{T}^c}, J_{\mathcal{S}^c}, \tilde{Q}) - n\Gamma_{\mathcal{T}} + I(Y^n_{\mathcal{S}}; J_{\mathcal{S}} | X^n_{\mathcal{L}}, J_{\mathcal{S}^c}, \tilde{Q}) &(36)\\
=& \sum_{i=1}^{n} H(X_{\mathcal{T},i} | X^n_{\mathcal{T}^c}, J_{\mathcal{S}^c}, X^{i-1}_{\mathcal{T}}, \tilde{Q}) - n\Gamma_{\mathcal{T}} + I(Y^n_{\mathcal{S}}; J_{\mathcal{S}} | X^n_{\mathcal{L}}, J_{\mathcal{S}^c}, \tilde{Q}) &(37)\\
\geq& \sum_{i=1}^{n} H(X_{\mathcal{T},i} | X_{\mathcal{T}^c,i}, U_{\mathcal{S}^c,i}, \bar{Q}_i) - n\Gamma_{\mathcal{T}} + I(Y^n_{\mathcal{S}}; J_{\mathcal{S}} | X^n_{\mathcal{L}}, J_{\mathcal{S}^c}, \tilde{Q}) &(38)\\
=& nR_{\mathcal{T}} - \sum_{i=1}^{n} I(X_{\mathcal{T},i}; U_{\mathcal{S}^c,i} | X_{\mathcal{T}^c,i}, \bar{Q}_i) + \sum_{k\in\mathcal{S}} \sum_{i=1}^{n} I(Y_{k,i}; U_{k,i} | X_{\mathcal{L},i}, \bar{Q}_i), &(39)
\end{aligned}$$

where (34) follows since $J_{\mathcal{S}}$ is a function of $Y^n_{\mathcal{S}}$; (36) follows from (3); (38) follows since conditioning reduces entropy; and (39) follows from (3) and (29).

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

$$U_{k,i} - Y_{k,i} - X_{\mathcal{L},i} - Y_{\mathcal{K}\setminus k,i} - U_{\mathcal{K}\setminus k,i}. \tag{40}$$

Finally, we define the standard time-sharing variable Q' uniformly distributed over $\{1,\dots,n\}$, $X_{\mathcal{L}} \triangleq X_{\mathcal{L},Q'}$, $Y_k \triangleq Y_{k,Q'}$, $U_k \triangleq U_{k,Q'}$ and $Q \triangleq [\bar{Q}_{Q'}, Q']$ and we have from (25)

$$nR_{\mathcal{T}} \leq \sum_{i=1}^{n} I(X_{\mathcal{T},i}; U_{\mathcal{K},i}|X_{\mathcal{T},i}, \bar{Q}_i) + n\epsilon_n \tag{41}$$

$$= nI(X_{\mathcal{T},Q'}; U_{\mathcal{K},Q'}|X_{\mathcal{T}^c,Q'}, \bar{Q}_{Q'}, Q') + n\epsilon_n \tag{42}$$

$$= nI(X_{\mathcal{T}}; U_{\mathcal{K}}|X_{\mathcal{T}^c}, Q) + n\epsilon_n, \tag{43}$$

and similarly, from (39), we have

$$R_{\mathcal{T}} \leq C_{\mathcal{S}} - \sum_{k\in\mathcal{S}} I(Y_k; U_k|X_{\mathcal{L}}, Q) + I(X_{\mathcal{L}}; U_{\mathcal{S}^c}|X_{\mathcal{T}^c}, Q).$$

This completes the proof of Theorem 1 . □

III. On the optimality of Separate Decompression and Decoding in Remark 1

The inner bound in Theorem 3 is based on a modification of the classical compress-and-forward scheme [4], [5] by letting the joint decoding of the compression codewords and the transmitted messages. In this section, we focus on the classical compress-and-forward scheme in which the decoding of the compression codewords and the decoding of the transmitted messages is performed separately; and on a low complexity version of it, in which the compressed channel outputs are decompressed successively and then, the users' message decoding is also done successively. Both strategies are shown to achieve the same sum-rate as CoF-JD in Theorem 3 .

We first define the sum-rate region. For given fronthaul tuple $(C_1,\dots,C_K)$, a sum-rate R_{sum} is said to be achievable if there exist an achievable tuple $(R_1,\dots,R_L,C_1,\dots,C_K)$ such that

$$\sum_{l=1}^{L} R_l = R_{\mathrm{sum}}. \tag{44}$$

The sum-rate region $\mathcal{R}^{\text{sum}}(C_{\mathcal{K}})$ is given by the closure of all achievable tuples $(R_{\text{sum}}, C_1, \ldots, C_K)$.

Since the partial sums over $\mathcal{T}$ in the rate constraints of Theorem 3 can be attained with equality, it follows that the achievable sum-rate region with CoF-JD and NNC, $\mathcal{R}^{\text{sum}}_{\text{JD}}$, is given as follows.

Corollary 1. *For fronthaul capacity $(C_1, \ldots, C_K)$, the sum-rate R_{sum} is achievable for the O-CRAN if for all $\mathcal{S} \subseteq \mathcal{K}$ we have*

$$R_{\text{sum}} \leq \sum_{s\in\mathcal{S}} C_s - I(Y_{\mathcal{S}}; U_{\mathcal{S}} | X_{\mathcal{L}}, U_{\mathcal{S}^c}, Q) + I(U_{\mathcal{S}^c}; X_{\mathcal{L}} | Q), \tag{45}$$

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

In the classical compress-and-forward [5], [6], the decoding complexity is reduced by considering separate decompression and user message decoding, i.e., at the CP the channel outputs are first decompressed, and then the uses messages are decoded. We denote this scheme by Compress-and-Forward with Separate Decompression and Decoding (CoF-SD). The sum-rate region achievable by CoF-SD, $\mathcal{R}^{\text{sum}}_{\text{SD}}$, is given in the next proposition.

Proposition 2. *For fronthaul capacity $(C_1, \ldots, C_K)$, the sum-rate R_{sum} is achievable for the O-CRAN if for all $\mathcal{S} \subseteq \mathcal{K}$ we have*

$$R_{\text{sum}} < I(X_{\mathcal{L}}; U_{\mathcal{K}} | Q), \tag{46}$$

$$\sum_{s\in\mathcal{S}} C_s > I(U_{\mathcal{S}}; Y_{\mathcal{S}} | U_{\mathcal{S}^c}, Q), \tag{47}$$

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

Proof: The proof follows by considering separate decompression and decoding error events in the proof in Appendix VII as in the proof of [3, Theorem 1]. □

A low complexity version of CoF-SD consists on the concatenation of successive Wyner-Ziv compression, followed by the successive decoding of the channel inputs as follows. For a given permutation of the relays $\pi : \mathcal{K} \to \mathcal{K}$:

1) Relay $\pi(1)$ employs conventional lossy source coding to compress $Y_{\pi(1)}$. Relay $\pi(k)$, $k=1,\dots,K$, employs Wyner-Ziv coding to compress $Y_{\pi(k)}$ with $(U_{\pi(1)},\dots,U_{\pi(k-1)})$ being the side information at the decoder.
2) The CP first decodes the codeword $U_{\pi(1)}$ from relay $\pi(1)$, then successively decodes the codeword $U_{\pi(k)}$, $k=1,\dots,K$, from relay $\pi(k)$ with side information $(U_{\pi(1)},\dots,U_{\pi(k-1)})$, $k=1,\dots,M$.
3) Finally, the CP decodes the codeword X_1 from transmitter 1. Then, it successively decodes the codeword X_l, $l=1,\dots,L$ from transmitter k, being $(X_1,\dots,X_{l-1})$, $l=2,\dots,L$ decoded codewords.

We denote this scheme by Compress and Forward with Successive Wyner Ziv and Successive Decoding (CoF-SWZ), and define $\mathcal{R}_{\text{SWZ}}^{\text{sum}}$ as its achievable sum-rate region.

Proposition 3. *For fronthaul capacity $(C_1,\dots,C_K)$, the sum-rate R_{sum} is achievable for the O-CRAN if for all $\mathcal{S}\subseteq\mathcal{K}$ and for some ordering $\pi:\mathcal{K}\to\mathcal{K}$, we have*

$$R_{\text{sum}} < \sum_{l=1}^{L} I(X_l;U_{\mathcal{K}}|X_1^{l-1},Q), \tag{48}$$

$$C_{\pi(k)} > I(U_{\pi(k)};Y_{\pi(k)}|U_{\pi(1)},\dots,U_{\pi(k-1)},Q), \quad k=1\dots,K, \tag{49}$$

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

Note that in general, $\mathcal{R}_{\text{SWZ}}^{\text{sum}}\subseteq\mathcal{R}_{\text{SD}}^{\text{sum}}\subseteq\mathcal{R}_{\text{JD}}^{\text{sum}}$. It is shown in [7] that joint compression decoding and separate decompression decoding, i.e., CoF-SD and CoF-JD, achieve the same sum-rate[2], for the class of channels satisfying [C1]-(5) . Next theorem shows that CoF-SWZ also achieves the same sum-rate as CoF-JD for the general O-CRAN model in Section **??**. The question of whether CoF-SWZ can achieve not only the same sum-rate but the whole rate-region achievable by CoF-JD remains as an open problem.

Theorem 1. $\mathcal{R}_{\text{JD}}^{\text{sum}}=\mathcal{R}_{\text{SD}}^{\text{sum}}=\mathcal{R}_{\text{SWZ}}^{\text{sum}}$.

[2]In fact, it follow from a minor modification of the proof in [7] that CoF-SWZ also achieves the same performance as CoF-JD in their setup.

Proof: The proof of Theorem 1 appears in Appendix IV. □

A direct consequence of Theorem 1 along with Theorem 1 is that CoF-SWZ achieves the optimal sum-rate the class of O-CRAN satisfying [C1]-(5) .

Corollary 2. *The optimal sum-rate in the O-CRAN model in Section II for the class of O-CRAN satisfying [C1]-(5) is achieved by the CoF-SWZ in Proposition 3.*

IV. Proof of Theorem 1

Since $\mathcal{R}_{\text{SWZ}}^{\text{sum}} \subseteq \mathcal{R}_{\text{SD}}^{\text{sum}} \subseteq \mathcal{R}_{\text{JD}}^{\text{sum}}$, to prove that CoF-SD and CoF-SWZ achieve same sum-rate as CoF-JD it suffices to show $\mathcal{R}_{\text{SWZ}}^{\text{sum}} \supseteq \mathcal{R}_{\text{JD}}^{\text{sum}}$.

We prove $\mathcal{R}_{\text{SWZ}}^{\text{sum}} \supseteq \mathcal{R}_{\text{JD}}^{\text{sum}}$ using the properties of submodular optimization. To this end, assume $(R_{\text{sum}}, C_1, \dots, C_K) \in \mathcal{R}_{\text{JD}}^{\text{sum}}$ for a joint pmf $p(q)\prod_{l=1}^{L} p(x_l|q)\prod_{k=1}^{K} p(u_k|y_k,q)$. For such pmf, let $\mathcal{P}_R \in \mathbb{R}_+^K$ be the polytope formed by the set of $(C_1,\dots,C_K)$ such that, for all $\mathcal{S} \subseteq \mathcal{K}$,

$$\sum_{s\in\mathcal{S}} C_s \geq \left[R_{\text{sum}} + I(U_\mathcal{S}; Y_\mathcal{S}|X_\mathcal{L}, U_{\mathcal{S}^c}, Q) - I(U_{\mathcal{S}^c}; X_\mathcal{L}|Q)\right]^+ . \tag{50}$$

Definition 1. *For a pmf $p(q)\prod_{l=1}^{L} p(x_l|q)\prod_{k=1}^{K} p(u_k|y_k,q)$ we say a point $(R_{\text{sum}}, C_1, \dots, C_K) \in \mathcal{R}_{\text{JD}}^{\text{sum}}$ is dominated by a point in $\mathcal{R}_{\text{SWZ}}^{\text{sum}}$ if there exists $(R'_{\text{sum}}, C'_1, \dots, C'_K) \in \mathcal{R}_{\text{SWZ}}^{\text{sum}}$ for which $C'_k \leq C_k$, for $k = 1,\dots,K$, and $R'_{\text{sum}} \geq R_{\text{sum}}$.*

To show $(R_{\text{sum}}, C_1, \dots, C_K) \in \mathcal{R}_{\text{SWZ}}^{\text{sum}}$, it suffices to show that each extreme point of $\mathcal{P}_R$ is dominated by a point in $\mathcal{R}_{\text{SWZ}}^{\text{sum}}$ that achieves a sum-rate $\bar{R}_{\text{sum}}$ satisfying $\bar{R}_{\text{sum}} \geq R_{\text{sum}}$.

Next, we characterize the extreme points of $\mathcal{P}_R$. Let us define the set function $g : 2^{\mathcal{K}} \to \mathbb{R}$ as follows:

$$g(\mathcal{S}) \triangleq R_{\text{sum}} + I(U_\mathcal{S}; Y_\mathcal{S}|U_{\mathcal{S}^c}, Q) - I(U_\mathcal{K}; X_\mathcal{L}|Q), \quad \text{for each} \quad \mathcal{S} \subseteq \mathcal{K}, \tag{51}$$

It can be verified that the function $g^+(\mathcal{S}) \triangleq \max\{g(\mathcal{S}), 0\}$ is a supermodular function (see [8, Appendix C, Proof of Lemma 6][3]).

[3]Note that the proof in [8, Appendix C, Proof of Lemma 6] to show that $g'(\mathcal{S}) \triangleq I(U_\mathcal{S}; Y_\mathcal{S}|U_{\mathcal{S}^c}, Q)$ is supermodular applies directly to our setup in which $Y_k -\!\!\circ\!\!- X_\mathcal{L} -\!\!\circ\!\!- Y_{\mathcal{K}/k}$ does not hold in general.

Note the following equality of (51). For each $\mathcal{S} \subseteq \mathcal{K}$, we have

$$g(\mathcal{S}) = R_{\text{sum}} + I(U_{\mathcal{S}}; Y_{\mathcal{S}}|U_{\mathcal{S}^c}, Q) - I(U_{\mathcal{K}}; X_{\mathcal{L}}|Q) \tag{52}$$

$$= R_{\text{sum}} + I(U_{\mathcal{S}}; X_{\mathcal{L}}, Y_{\mathcal{S}}|U_{\mathcal{S}^c}, Q) - I(U_{\mathcal{S}^c}; X_{\mathcal{L}}|Q) - I(U_{\mathcal{S}}; X_{\mathcal{L}}|U_{\mathcal{S}^c}, Q) \tag{53}$$

$$= R_{\text{sum}} + I(U_{\mathcal{S}}; Y_{\mathcal{S}}|X_{\mathcal{L}}, U_{\mathcal{S}^c}, Q) - I(U_{\mathcal{S}^c}; X_{\mathcal{L}}|Q), \tag{54}$$

where (53) follows due to the Markov chain

$$U_{\mathcal{S}} -\!\!\!\circ\!\!\!- Y_{\mathcal{S}} -\!\!\!\circ\!\!\!- (X_{\mathcal{L}}, U_{\mathcal{S}^c}). \tag{55}$$

Then, by construction, $\mathcal{P}_R$ is equal to the set of $(C_1, \ldots, C_K)$ satisfying for all $\mathcal{S} \subseteq \mathcal{K}$,

$$\sum_{s \in \mathcal{S}} C_s \geq g^+(\mathcal{S}). \tag{56}$$

Following the results in submodular optimization [7, 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 $\mathcal{P}_R$ can be computed as follows for $k = 1, \ldots, K$:

$$\tilde{C}_{i_k} = g^+(\{i_1, \ldots, i_k\}) - g^+(\{i_1, \ldots, i_{k-1}\}). \tag{57}$$

All the $K!$ extreme points of $\mathcal{P}_R$ can be enumerated by looking over all linear orderings $i_1 \prec i_2 \prec \cdots \prec i_K$ of $\mathcal{K}$. Each ordering of $\mathcal{K}$ is analyzed in the same manner and, therefore, for notational simplicity, the only ordering we consider is the natural ordering $i_k = k$. By construction,

$$\begin{aligned}\tilde{C}_k = &\left[R_{\text{sum}} + I(U_1^k; Y_1^k|X_{\mathcal{L}}, U_{k+1}^K, Q) - I(U_{k+1}^K; X_{\mathcal{L}}|Q)\right]^+ \\ &- \left[R_{\text{sum}} + I(U_1^{k-1}; Y_1^{k-1}|X_{\mathcal{L}}, U_k^K, Q) - I(U_k^K; X_{\mathcal{L}}|Q)\right]^+ .\end{aligned} \tag{58}$$

Let j be the first index for which $\tilde{C}_j > 0$, i.e., the first k for which $g(\{1, \ldots, j\}) > 0$. Then, it follows from (58) that

$$\tilde{C}_k = I(U_1^k; Y_1^k|X_{\mathcal{L}}, U_{k+1}^K, Q) - I(U_{k+1}^K; X_{\mathcal{L}}|Q) \tag{59}$$

$$- I(U_1^{k-1}; Y_1^{k-1}|X_{\mathcal{L}}, U_k^K, Q) + I(U_k^K; X_{\mathcal{L}}|Q) \tag{60}$$

$$=I(U_1^k;Y_1^k|U_{k+1}^K,Q)-I(U_1^{k-1};Y_1^{k-1}|U_k^K,Q) \tag{61}$$

$$=I(Y_k;U_k|U_{k+1}^K,Q), \qquad \text{for all } k>j, \tag{62}$$

where (61) follows from 52; and (62) follows due to the Markov Chain

$$U_k -\!\!\circ\!\!- Y_k -\!\!\circ\!\!- (X_{\mathcal{L}},Y_{\mathcal{K}/k},U_{\mathcal{K}/k}). \tag{63}$$

Moreover, since we must have $g(\{1,\dots,j'\})\leq 0$ for $j'<j$, $\tilde{C}_j$ can be expressed as

$$\tilde{C}_j = R_{\mathrm{sum}}+I(U_1^j;Y_1^j|X_{\mathcal{L}},U_{j+1}^K,Q)-I(U_{j+1}^K;X_{\mathcal{L}}|Q) \tag{64}$$

$$= I(Y_j;U_j|U_{j+1}^K,Q)+g(\{1,\dots,j-1\}), \tag{65}$$

$$= (1-\alpha)I(Y_j;U_j|U_{j+1}^K,Q), \tag{66}$$

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

$$\alpha \triangleq \frac{-g(\{1,\dots,j-1\})}{I(Y_j;U_j|U_{j+1}^K,Q)} \tag{67}$$

$$= \frac{I(U_j^K;X_{\mathcal{L}}|Q)-R_{\mathrm{sum}}-I(U_1^{j-1};Y_1^{j-1}|X_{\mathcal{L}},U_j^K,Q)}{I(Y_j;U_j|U_{j+1}^L,Q)}. \tag{68}$$

Therefore, for the natural ordering, the extreme point $(\tilde{C}_1,\dots,\tilde{C}_K)$ is given as

$$(\tilde{C}_1,\dots,\tilde{C}_K)=\big(0,\dots,0,(1-\alpha)I(Y_j;U_j|U_{j+1}^K,Q),I(Y_{j+1};U_{j+1}|U_{j+2}^K,Q), \tag{69}$$

$$\dots,I(Y_{K-1};U_{K-1}|U_K,Q),I(Y_K;U_K|Q)\big). \tag{70}$$

Next, we show that $(\tilde{C}_1,\dots,\tilde{C}_K)\in\mathcal{P}_R$, is dominated by a point $(\bar{R}_{\mathrm{sum}},C_1,\dots,C_K)$ in $\mathcal{R}_{\mathrm{SWZ}}^{\mathrm{sum}}$ that achieves a sum-rate $\bar{R}_{\mathrm{sum}}\geq R_{\mathrm{sum}}$.

We consider an instance of the CoF-SWZ in which for a fraction α of the time, the CP decodes $U_{j+1}^n,\dots,U_K^n$ while relays $k=1,\dots,j$ are inactive. For the remaining fraction of time $(1-\alpha)$, the CP decodes $U_j^n,\dots,U_K^n$ and relays $k=1,\dots,j-1$ are inactive. Then, the CP decodes $X_{\mathcal{L}}$.

Formally, we consider the pfm $p(q')\prod_{l=1}^L p(x_l'|q')\prod_{k=1}^K p(u_k'|y_k,q')$ for CoF-SZW as follows. Let B denote a Bernoulli random variable with parameter $\alpha\in(0,1]$, i.e., $B=1$ with probability

α and $B=0$ with probability $(1-\alpha)$. We let α as in (68). We consider the reverse ordering π such that $\pi(1)=K, \pi(2)=K-1, \ldots, \pi(K)=1$, i.e., compression is done from relay K to relay 1. Then, we let $Q'=(B,Q)$ and the tuple of random variables be distributed as

$$(Q', X'_{\mathcal{L}}, U'_{\mathcal{K}}) = \begin{cases} ((1,Q), X_{\mathcal{L}}, \emptyset, \ldots, \emptyset, U_{j+1}, \ldots, U_K) & \text{if } B=1, \\ ((0,Q), X_{\mathcal{L}}, \emptyset, \ldots, \emptyset, U_j, \ldots, U_K) & \text{if } B=0. \end{cases} \tag{71}$$

From Proposition 3, the following tuple $(\bar{R}_{\text{sum}}, C_1, \ldots, C_K) \in \mathcal{R}^{\text{sum}}_{SWZ}$ is achievable, where

$$C_k = I(Y_k; U'_k | U'_{k+1}, \ldots, U'_K, Q'), \quad \text{for} \quad k=1,\ldots,K, \tag{72}$$

$$\bar{R}_{\text{sum}} = I(X'_{\mathcal{L}}; U'_{\mathcal{K}} | Q'). \tag{73}$$

For $k=1,\ldots,j-1$, we have

$$C_k = I(Y_k; U'_k | U'_{k+1}, \ldots, U'_K, Q') \tag{74}$$

$$= 0 = \tilde{C}_k, \tag{75}$$

where (75) follows since $U'_k=\emptyset$ for $k<j$ independently of B. For $k=j+1,\ldots,K$, we have

$$C_k = I(Y_k; U'_k | U'_{k+1}, \ldots, U'_K, Q') \tag{76}$$

$$= \alpha I(Y_k; U_k | U_{k+1}, \ldots, U_K, Q, B=1) + (1-\alpha) I(Y_k; U_k | U_{k+1}, \ldots, U_K, Q, B=0) \tag{77}$$

$$= I(Y_k; U_k | U_{k+1}, \ldots, U_K, Q) = \tilde{C}_k, \tag{78}$$

where (78) follows since $U'_k = U_k$ for $k>j$ independently of B. For $k=j$, we have

$$C_j = I(Y_j; U'_j | U'_{j+1}, \ldots, U'_K, Q') \tag{79}$$

$$= \alpha I(Y_j; U_j | U_{j+1}, \ldots, U_K, Q, B=1) + (1-\alpha) I(Y_j; U_j | U_{j+1}, \ldots, U_K, Q, B=0) \tag{80}$$

$$= (1-\alpha) I(Y_j; U_j | U_{j+1}, \ldots, U_K, Q) = \tilde{C}_j; \tag{81}$$

where (81) follows since $U'_j=\emptyset$ for $B=1$ and $U'_j=U_j$ for $B=0$.

On the other hand, the sum-rate satisfies

$$\bar{R}_{\mathrm{sum}} = I(X'_{\mathcal{L}}; U'_{\mathcal{K}}|Q') \tag{82}$$

$$= \alpha I(X_{\mathcal{L}}; U_{j+1}^K|Q, B=1) + (1-\alpha) I(X_{\mathcal{L}}; U_j^K|Q, B=0) \tag{83}$$

$$= I(X_{\mathcal{L}}; U_j^K|Q) - \alpha I(X_{\mathcal{L}}; U_j|U_{j+1}^K, Q) \tag{84}$$

$$= I(X_{\mathcal{L}}; U_j^K|Q) - \frac{I(X_{\mathcal{L}}; U_j|U_{j+1}^K, Q)}{I(Y_j; U_j|U_{j+1}^L, Q)} \cdot \left[I(U_j^K; X_{\mathcal{L}}|Q) - R_{\mathrm{sum}} - I(U_1^{j-1}; Y_1^{j-1}|X_{\mathcal{L}}, U_j^K, Q) \right] \tag{85}$$

$$\geq R_{\mathrm{sum}} + I(U_1^{j-1}; Y_1^{j-1}|X_{\mathcal{L}}, U_j^K, Q) \tag{86}$$

$$\geq R_{\mathrm{sum}}, \tag{87}$$

where (85) follows from (68); and (86) follows since $I(Y_j; U_j|U_{j+1}^L, Q) \geq I(X_{\mathcal{L}}; U_j|U_{j+1}^K, Q)$ due to the Markov Chain (63).

Therefore, from (75), (78), (81) and (87), it follows that the extreme point $(\tilde{C}_1, \ldots, \tilde{C}_K) \in \mathcal{P}_R$ is dominated by the point $(\bar{R}_{\mathrm{sum}}, C_1, \ldots, C_K) \in \mathcal{R}_{\mathrm{SWZ}}^{\mathrm{sum}}$ satisfying $\bar{R}_{\mathrm{sum}} \geq R_{\mathrm{sum}}$.

Similarly, considering all possible orderings, each extreme point of $\mathcal{P}_R$ can be shown to be dominated by a point $(R_{\mathrm{sum}}, C_1, \ldots, C_K)$ which lies in $\mathcal{R}_{\mathrm{SWZ}}^{\mathrm{sum}}$ (associated to a permutation π). Since $\mathcal{R}^{\mathrm{sum}}$ is the convex hull of all such extreme points, this completes the proof.

V. Proof of Theorem 2

The proof has some similarities to [7, Proof of Theorem 4] but differs from it due to the time-sharing variable Q. Before proving the result, we provide the following lemmas.

Definition 2. *Let $(\mathbf{X}, \mathbf{Y})$ be a pair of random vectors with joint probability distribution $p(\mathbf{x}, \mathbf{y})$. The Fischer information matrix of $\mathbf{X}$ is defined as*

$$\mathbf{J}(\mathbf{X}) \triangleq \mathrm{E}[\nabla \log p(\mathbf{X}) \nabla \log p(\mathbf{X})^T]. \tag{88}$$

The Fischer information matrix of $\mathbf{X}$ conditional on $\mathbf{Y}$ is defined as

$$\mathbf{J}(\mathbf{X}|\mathbf{Y}) \triangleq \mathrm{E}[\nabla \log p(\mathbf{X}|\mathbf{Y}) \nabla \log p(\mathbf{X}|\mathbf{Y})^T]. \tag{89}$$

Lemma 1 (Fischer Information Inequality)**.** *Let* $(\mathbf{X}, \mathbf{U})$ *be an arbitrary complex random vector where the conditional Fischer information* $\mathbf{X}$ *conditioned on* $\mathbf{U}$ *exists. We have*

$$h(\mathbf{X}|\mathbf{U}) \geq \log|(\pi e)\mathbf{J}^{-1}(\mathbf{X}|\mathbf{U})|. \tag{90}$$

Lemma 2 (Brujin Identity)**.** *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* $\mathbf{\Lambda}_N$*. Assume* $(\mathbf{V}_1, \mathbf{V}_2)$ *and* $\mathbf{N}$ *are independent. We have*

$$\mathrm{mmse}(\mathbf{V}_2|\mathbf{V}_1, \mathbf{V}_2 + \mathbf{N}) = \mathbf{\Lambda}_N - \mathbf{\Lambda}_N \mathbf{J}(\mathbf{V}_2 + \mathbf{N}|\mathbf{V}_1)\mathbf{\Lambda}_N. \tag{91}$$

Let $\mathcal{H}_m$ be the set of all $m \times m$ Hermitian matrices. A positive semidefinite partial order on $\mathcal{H}_m$ is $\mathbf{B} \succeq \mathbf{A}$ if $\mathbf{B} - \mathbf{A}$ is positive semidefinite.

Definition 3. *Let* $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_n)$ *be an* n*-tuple of positive definite matrices from* $\mathcal{H}_m$*, and let* $\mathbf{w} = (w_1, \ldots, w_n)$ *be an* n*-tuple of non-negative reals whose sum is* 1*. We define the arithmetic mean of* $\mathbf{A}$ *with weight* $\mathbf{w}$ *as*

$$\mathbf{A}_n(\mathbf{A}; \mathbf{w}) \triangleq \sum_{i=1}^{n} w_i \mathbf{A}_i, \tag{92}$$

and the harmonic mean of $\mathbf{A}$ *with weights* $\mathbf{w}$ *as*

$$\mathbf{H}_n(\mathbf{A}; \mathbf{w}) \triangleq \left(\sum_{i=1}^{n} w_i \mathbf{A}_i^{-1} \right)^{-1}. \tag{93}$$

Lemma 3. *[9], [10] We have* $\mathbf{A}_n(\mathbf{A}; \mathbf{w}) \succeq \mathbf{H}_n(\mathbf{A}; \mathbf{w})$*.*

Lemma 4. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two positive-definite matrices from* $\mathcal{H}_m$ *satisfying* $\mathbf{B} \succeq \mathbf{A}$*. Then for any positive-definite matrix* $\mathbf{C}$ *in* $\mathcal{H}_m$*, we have* $|\mathbf{I} + \mathbf{BC}| \geq |\mathbf{I} + \mathbf{AC}|$*.*

Proof. The proof is given in Appendix VI. □

Next, we obtain an outer bound the capacity region in Theorem 1 for jointly Gaussian channel inputs and average covariance constraint,

$$\mathrm{E}[\|\mathbf{X}_l\|^2] \preceq \mathbf{K}_l, \quad l = [1, L]. \tag{94}$$

For convenience, we define the covariance matrix of $\mathbf{X}_{\mathcal{T}}$ as $\mathbf{K}_{\mathcal{T}} \triangleq \mathrm{diag}[\{\mathbf{K}_t\}_{t\in\mathcal{T}}]$ for $\mathcal{T} \subseteq \mathcal{L}$.

For $Q = q$, the channel input satisfies $\mathbf{X}_{\mathcal{L},q} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_{\mathcal{L},q})$. Thus, from (94), we have

$$\mathrm{E}[\|\mathbf{X}_l\|^2] = \sum_{q\in\mathcal{Q}} p(q)\mathrm{E}[\mathbf{X}_{l,q}|Q = q] = \sum_{q\in\mathcal{Q}} p(q)\mathbf{K}_{l,q} \preceq \mathbf{K}_l, \quad l = [1, L]. \tag{95}$$

On the other hand, for a fixed $Q = q$, input $\mathbf{X}_{\mathcal{L},q}$ and $\prod_{k=1}^{K} p(\hat{\mathbf{y}}_k|\mathbf{y}_k, q)$, we choose $\mathbf{B}_{k,q}$ with $\mathbf{0} \preceq \mathbf{B}_{k,q} \preceq \mathbf{\Sigma}_k^{-1}$ such that

$$\mathrm{mmse}(\mathbf{Y}_k|\mathbf{X}_{\mathcal{L},q}, \mathbf{U}_{k,q}, q) = \mathbf{\Sigma}_k - \mathbf{\Sigma}_k\mathbf{B}_{k,q}\mathbf{\Sigma}_k, \quad q = [1, |\mathcal{Q}|],\ k = [1, K]. \tag{96}$$

Such $\mathbf{B}_{k,q}$ always exists since $\mathbf{0} \preceq \mathrm{mmse}(\mathbf{Y}_k|\mathbf{X}_{\mathcal{L},q}, \mathbf{U}_{k,q}, q) \preceq \mathbf{\Sigma}_k$ for all $q = [1, |\mathcal{Q}|]$, $k = [1, K]$.

We have for $k = [1, K]$, and $q = [1, |\mathcal{Q}|]$

$$\begin{aligned} I(\mathbf{Y}_k; \mathbf{U}_k|\mathbf{X}_{\mathcal{L},q}, Q = q) &= \log|(\pi e)\mathbf{\Sigma}_k| - h(\mathbf{Y}_s|\mathbf{X}_{\mathcal{L},q}, \mathbf{U}_{s,q}, Q = q) && (97)\\ &\geq \log|(\pi e)\mathbf{\Sigma}_k| - \log|(\pi e)\ \mathrm{mmse}(\mathbf{Y}_k|\mathbf{X}_{\mathcal{L},q}, \mathbf{U}_k, q)| && (98)\\ &\geq \log\frac{|\mathbf{\Sigma}_k^{-1}|}{|\mathbf{\Sigma}_k^{-1} - \mathbf{B}_{k,q}|}. && (99) \end{aligned}$$

On the other hand, for $k = [1, K]$, and $q = [1, |\mathcal{Q}|]$

$$\begin{aligned} I(\mathbf{X}_{\mathcal{T},q}; \mathbf{U}_{S^c}|\mathbf{X}_{\mathcal{T}^c,q}, Q = q) &= h(\mathbf{X}_{\mathcal{T},q}|Q = q) - h(\mathbf{X}_{\mathcal{T},q}|\mathbf{X}_{\mathcal{T}^c,q}, \mathbf{U}_{S^c,q}, Q = q) && (100)\\ &\leq \log|\mathbf{K}_{\mathcal{T},q}| - \log|\mathbf{J}^{-1}(\mathbf{X}_{\mathcal{T},q}|\mathbf{X}_{\mathcal{T}^c,q}, \mathbf{U}_{S^c,q}, q)|, && (101)\\ &\leq \log|\mathbf{K}_{\mathcal{T},q}| + \log\left|\sum_{k\in\mathcal{S}^c} \mathbf{H}_{k,\mathcal{T}}^H\mathbf{B}_{k,q}\mathbf{H}_{k,\mathcal{T}} + \mathbf{K}_{\mathcal{T},q}^{-1}\right|, && (102) \end{aligned}$$

where (101) is due to Lemma 1; and (102) is due to

$$\mathbf{J}(\mathbf{X}_{\mathcal{T},q}|\mathbf{X}_{\mathcal{T}^c,q}, \mathbf{U}_{S^c,q}, q) = \sum_{k\in\mathcal{S}^c} \mathbf{H}_{k,\mathcal{T}}^H\mathbf{B}_{k,q}\mathbf{H}_{k,\mathcal{T}} + \mathbf{K}_{\mathcal{T},q}^{-1}. \tag{103}$$

Equality (103) is obtained as follows. Since

$$\mathbf{Y}_{\mathcal{S}^c} = \mathbf{H}_{\mathcal{S}^c,\mathcal{T}}\mathbf{X}_{\mathcal{T},q} + \mathbf{H}_{\mathcal{S}^c,\mathcal{T}^c}\mathbf{X}_{\mathcal{T}^c,q} + \mathbf{N}_{\mathcal{S}^c} \tag{104}$$

it follows from the MMSE estimation of Gaussian random vectors [11], that

$$\mathbf{X}_{\mathcal{T},q} = \mathrm{E}[\mathbf{X}_{\mathcal{T},q}|\mathbf{X}_{\mathcal{T}^c,q},\mathbf{Y}_{\mathcal{S}^c}] + \mathbf{Z}_{\mathcal{T},\mathcal{S}^c} = \sum_{k\in\mathcal{S}^c}\mathbf{G}_{\mathcal{T},k}(\mathbf{Y}_k - \mathbf{H}_{k,\mathcal{T}^c}\mathbf{X}_{\mathcal{T}^c}) + \mathbf{Z}_{\mathcal{T},\mathcal{S}^c}, \tag{105}$$

where

$$\mathbf{G}_{\mathcal{T},k} = \left(\mathbf{K}_{\mathcal{T},q}^{-1} + \sum_{j\in\mathcal{S}^c}\mathbf{H}_{j,\mathcal{T}}^H\mathbf{\Sigma}_j^{-1}\mathbf{H}_{j,\mathcal{T}}\right)^{-1}\mathbf{H}_{k,\mathcal{T}}^H\mathbf{\Sigma}_k^{-1}, \tag{106}$$

and $\mathbf{Z}_{\mathcal{T},\mathcal{S}^c} \sim \mathcal{CN}(\mathbf{0},\mathbf{\Lambda}_Z)$ with covariance matrix

$$\mathbf{\Lambda}_Z = \left(\mathbf{K}_{\mathcal{T},q}^{-1} + \sum_{k\in\mathcal{S}^c}\mathbf{H}_{k,\mathcal{T}}^H\mathbf{\Sigma}_k^{-1}\mathbf{H}_{k,\mathcal{T}}\right)^{-1}. \tag{107}$$

Note that $\mathbf{\Lambda}_Z$ is independent of $\mathbf{X}_{\mathcal{T}^c,q},\mathbf{Y}_{\mathcal{S}^c}$ due to the orthogonality principle of the MMSE and its Gaussian distribution. Hence, $\mathbf{\Lambda}_Z$ is also independent of $\mathbf{U}_{\mathcal{S}^c,q}$. Then, by Lemma 2, we have

$$\begin{aligned}
\mathbf{J}(\mathbf{X}_{\mathcal{T},q}|\mathbf{X}_{\mathcal{T}^c,q},\mathbf{U}_{S^c,q},q) &= \mathbf{\Lambda}_Z^{-1} - \mathbf{\Lambda}_Z^{-1}\mathrm{mmse}\left(\sum_{k\in\mathcal{S}^c}\mathbf{G}_{\mathcal{T},k}(\mathbf{Y}_k - \mathbf{H}_{k,\mathcal{T}^c}\mathbf{X}_{\mathcal{T}^c,q})|\mathbf{X}_{\mathcal{L},q},\mathbf{U}_{\mathcal{S}^c,q},q\right)\mathbf{\Lambda}_Z^{-1} && (108)\\
&= \mathbf{\Lambda}_Z^{-1} - \mathbf{\Lambda}_Z^{-1}\mathrm{mmse}\left(\sum_{k\in\mathcal{S}^c}\mathbf{G}_{\mathcal{T},k}\mathbf{Y}_k|\mathbf{X}_{\mathcal{L},q},\mathbf{U}_{\mathcal{S}^c,q},q\right)\mathbf{\Lambda}_Z^{-1} && (109)\\
&= \mathbf{\Lambda}_Z^{-1} - \mathbf{\Lambda}_Z^{-1}\left(\sum_{k\in\mathcal{S}^c}\mathbf{G}_{\mathcal{T},k}\mathrm{mmse}\left(\mathbf{Y}_k|\mathbf{X}_{\mathcal{L},q},\mathbf{U}_{\mathcal{S}^c,q},q\right)\mathbf{G}_{\mathcal{T},k}^H\right)\mathbf{\Lambda}_Z^{-1} && (110)\\
&= \mathbf{\Lambda}_Z^{-1} - \sum_{k\in\mathcal{S}^c}\mathbf{H}_{k,\mathcal{T}}^H\left(\mathbf{\Sigma}_k^{-1} - \mathbf{B}_k\right)\mathbf{H}_{k,\mathcal{T}} && (111)\\
&= \mathbf{K}_{\mathcal{T},q}^{-1} + \sum_{k\in\mathcal{S}^c}\mathbf{H}_{k,\mathcal{T}}^H\mathbf{B}_k\mathbf{H}_{k,\mathcal{T}}, && (112)
\end{aligned}$$

where (110) follows since the cross terms are zero due to the Markov chain,

$$(\mathbf{U}_{k,q}, \mathbf{Y}_k) -\!\!\circ\!\!- \mathbf{X}_q -\!\!\circ\!\!- (\mathbf{U}_{\mathcal{K}/k,q}, \mathbf{Y}_{\mathcal{K}/k}), \tag{113}$$

and (111) is due to (96).

Let us define the arithmetic mean $\bar{\mathbf{B}}_k \triangleq \sum_{q\in\mathcal{Q}} p(q)\mathbf{B}_{k,q}$. We have

$$\begin{aligned}
I(\mathbf{Y}_k;\mathbf{U}_k|\mathbf{X}_\mathcal{L},Q) &= \sum_{q\in\mathcal{Q}} p(q) I(\mathbf{Y}_k;\mathbf{U}_k|\mathbf{X}_\mathcal{L},Q=q) && (114)\\
&\geq \sum_{q\in\mathcal{Q}} p(q)\log\frac{|\mathbf{\Sigma}_k^{-1}|}{|\mathbf{\Sigma}_k^{-1}-\mathbf{B}_{k,q}|} && (115)\\
&\geq \log\frac{|\mathbf{\Sigma}_k^{-1}|}{|\mathbf{\Sigma}_k^{-1}-\sum_{q\in\mathcal{Q}} p(q)\mathbf{B}_{k,q}|} && (116)\\
&= \log\frac{|\mathbf{\Sigma}_k^{-1}|}{|\mathbf{\Sigma}_k^{-1}-\bar{\mathbf{B}}_k|}, && (117)
\end{aligned}$$

where (115) follows from (99) and (116) follows from the concavity of the log-det function and Jensen's Inequality [12].

For each $\mathcal{T}\subseteq L$, let $\mathbf{K}=(\mathbf{K}_{\mathcal{T},1},\ldots,\mathbf{K}_{\mathcal{T},|\mathcal{Q}|})$ and $\mathbf{p}=(p(1),\ldots,p(|\mathcal{Q}|))$, and let us consider the arithmetic mean and harmonic mean $\mathbf{A}_{|\mathcal{Q}|}(\mathbf{K},\mathbf{p})$ and $\mathbf{H}_{|\mathcal{Q}|}(\mathbf{K},\mathbf{p})$. We have

$$\begin{aligned}
&I(\mathbf{X}_\mathcal{T};\mathbf{U}_{S^c}|\mathbf{X}_{\mathcal{T}^c,q},Q)\\
&\leq \sum_{q\in\mathcal{Q}} p(q)\left(\log|\mathbf{K}_{q,\mathcal{T}}| + \log\left|\sum_{k\in\mathcal{S}^c}\mathbf{H}_{k,\mathcal{T}}^H\mathbf{B}_{k,q}\mathbf{H}_{k,\mathcal{T}}+\mathbf{K}_{\mathcal{T},q}^{-1}\right|\right) && (118)\\
&\leq \log\left|\sum_{q\in\mathcal{Q}} p(q)\mathbf{K}_{q,\mathcal{T}}\right| + \log\left|\sum_{q\in\mathcal{Q}} p(q)\sum_{k\in\mathcal{S}^c}\mathbf{H}_{k,\mathcal{T}}^H\mathbf{B}_{k,q}\mathbf{H}_{k,\mathcal{T}}+\sum_{q\in\mathcal{Q}} p(q)\mathbf{K}_{\mathcal{T},q}^{-1}\right| && (119)\\
&= \log\left(|\mathbf{A}_{|\mathcal{Q}|}(\mathbf{K},\mathbf{p})|\cdot|\mathbf{H}_{|\mathcal{Q}|}^{-1}(\mathbf{K},\mathbf{p})|\right) + \log\left|\mathbf{H}_{|\mathcal{Q}|}(\mathbf{K},\mathbf{p})\sum_{k\in\mathcal{S}^c}\mathbf{H}_{k,\mathcal{T}}^H\bar{\mathbf{B}}_k\mathbf{H}_{k,\mathcal{T}}+\mathbf{I}\right| && (120)\\
&= \log\left(|\mathbf{A}_{|\mathcal{Q}|}(\mathbf{K},\mathbf{p})|\cdot|\mathbf{H}_{|\mathcal{Q}|}^{-1}(\mathbf{K},\mathbf{p})|\right) + \log\left|\mathbf{K}_\mathcal{T}\sum_{k\in\mathcal{S}^c}\mathbf{H}_{k,\mathcal{T}}^H\bar{\mathbf{B}}_k\mathbf{H}_{k,\mathcal{T}}+\mathbf{I}\right| && (121)\\
&\leq \log\left|\mathbf{K}_\mathcal{T}\sum_{k\in\mathcal{S}^c}\mathbf{H}_{k,\mathcal{T}}^H\bar{\mathbf{B}}_k\mathbf{H}_{k,\mathcal{T}}+\mathbf{I}\right|, && (122)
\end{aligned}$$

where (118) follows from (102);(119) is due to the concavity of the log-det function and Jensen's inequality; (120) follows due to the definition of $\bar{\mathbf{B}}_k$ and the arithmetic and harmonic means; (121) follows due to Lemma 4, since $\sum_{k\in\mathcal{S}^c}\mathbf{H}_{k,\mathcal{T}}^H\bar{\mathbf{B}}_k\mathbf{H}_{k,\mathcal{T}}$ is positive-definite and the ordering $\mathbf{H}_{|\mathcal{Q}|}(\mathbf{K},\mathbf{p})\preceq\mathbf{A}_{|\mathcal{Q}|}(\mathbf{K},\mathbf{p})\preceq\mathbf{K}_{\mathcal{T}}$, which follows from (95) and Lemma 3; (122) follows since

$$
\begin{aligned}
|\mathbf{A}_{|\mathcal{Q}|}(\mathbf{K},\mathbf{p})|\cdot|\mathbf{H}_{|\mathcal{Q}|}^{-1}(\mathbf{K},\mathbf{p})| &= |\mathbf{H}_{|\mathcal{Q}|}^{-1/2}(\mathbf{K},\mathbf{p})\mathbf{A}_{|\mathcal{Q}|}(\mathbf{K},\mathbf{p})\mathbf{H}_{|\mathcal{Q}|}^{-1/2}(\mathbf{K},\mathbf{p})| && (123)\\
&= \prod_{i=1}\lambda_i(\mathbf{H}_{|\mathcal{Q}|}^{-1/2}(\mathbf{K},\mathbf{p})\mathbf{A}_{|\mathcal{Q}|}(\mathbf{K},\mathbf{p})\mathbf{H}_{|\mathcal{Q}|}^{-1/2}(\mathbf{K},\mathbf{p})) && (124)\\
&\leq \prod_{i=1}\lambda_i(\mathbf{I})\leq 1, && (125)
\end{aligned}
$$

where (125) follows since $\mathbf{H}_{|\mathcal{Q}|}^{-1/2}(\mathbf{K},\mathbf{p})\mathbf{A}_{|\mathcal{Q}|}(\mathbf{K},\mathbf{p})\mathbf{H}_{|\mathcal{Q}|}^{-1/2}(\mathbf{K},\mathbf{p})\preceq\mathbf{I}$, due to Lemma 3. Note that if $w_1=\cdots=w_{|\mathcal{Q}|}$ and $\mathbf{K}_{\mathcal{T},1}=\cdots=\mathbf{K}_{\mathcal{T},|\mathcal{Q}|}$, we have $\mathbf{A}_{|\mathcal{Q}|}(\mathbf{K},\mathbf{p})=\mathbf{H}_{|\mathcal{Q}|}(\mathbf{K},\mathbf{p})$ and the bound is satisfied with equality.

Substituting (117) and (122) in [C1]-(6) for each $\mathcal{T}\subseteq\mathcal{L}$ gives the desired outer bound.

The direct part of Theorem 2 follows by noting that this outer bound is achieved by evaluating [C1]-(6) for $Q=\emptyset$, $\mathbf{X}_{\mathcal{L}}\sim\mathcal{CN}(\mathbf{0},\mathbf{K}_{\mathcal{L}})$, and $p(\mathbf{U}_k|\mathbf{Y}_k)\sim\mathcal{CN}(\mathbf{Y}_k,\mathbf{Q}_k)$, where $\mathbf{B}_k=(\mathbf{\Sigma}_k+\mathbf{Q}_k)^{-1}$ for some $\mathbf{0}\preceq\mathbf{B}_k\preceq\mathbf{\Sigma}_k^{-1}$.

VI. Proof of Lemma 4

From $\mathbf{B}\succeq\mathbf{A}$, we have the following chain of implications

$$
\begin{aligned}
\mathbf{B}\succeq\mathbf{A} &\Rightarrow \mathbf{B}-\mathbf{A}\succeq\mathbf{0} && (126)\\
&\Rightarrow \mathbf{C}^{1/2}(\mathbf{B}-\mathbf{A})\mathbf{C}^{1/2}\succeq\mathbf{0} && (127)\\
&\Rightarrow \mathbf{C}^{1/2}\mathbf{B}\mathbf{C}^{1/2}\succeq\mathbf{C}^{1/2}\mathbf{A}\mathbf{C}^{1/2} && (128)\\
&\Rightarrow \mathbf{I}+\mathbf{C}^{1/2}\mathbf{B}\mathbf{C}^{1/2}\succeq\mathbf{I}+\mathbf{C}^{1/2}\mathbf{A}\mathbf{C}^{1/2}, && (129)
\end{aligned}
$$

where 127 follows since for any two positive-definite matrices $\mathbf{M}\succeq\mathbf{0},\mathbf{N}\succeq\mathbf{0}$ of size $N\times N$, we have $\mathbf{N}^{1/2}\mathbf{M}\mathbf{N}^{1/2}\succeq\mathbf{0}$, i.e., it is positive-definite; (128) and (129) follow due to linearity.

Due to Weyl's inequality [13], (129) implies

$$\lambda_i(\mathbf{I}+\mathbf{C}^{1/2}\mathbf{B}\mathbf{C}^{1/2}) \geq \lambda_i(\mathbf{I}+\mathbf{C}^{1/2}\mathbf{A}\mathbf{C}^{1/2}), \quad i=1,\dots,N \tag{130}$$

where $\lambda_i(\mathbf{M})$ denotes the i-th eigenvector of the positive-definite matrix $\mathbf{M} \succeq \mathbf{0}$. Then, we have

$$|\mathbf{I}+\mathbf{B}\mathbf{C}| = |\mathbf{I}+\mathbf{C}^{1/2}\mathbf{B}\mathbf{C}^{1/2}| \tag{131}$$

$$= \prod_{i=1}^{N} \lambda_i(\mathbf{I}+\mathbf{C}^{1/2}\mathbf{B}\mathbf{C}^{1/2}) \tag{132}$$

$$\geq \prod_{i=1}^{N} \lambda_i(\mathbf{I}+\mathbf{C}^{1/2}\mathbf{A}\mathbf{C}^{1/2}) \tag{133}$$

$$= |\mathbf{I}+\mathbf{C}^{1/2}\mathbf{A}\mathbf{C}^{1/2}| \tag{134}$$

$$= |\mathbf{I}+\mathbf{A}\mathbf{C}|, \tag{135}$$

where (131) follows since $|\mathbf{I}+\mathbf{N}\mathbf{M}| = |\mathbf{I}+\mathbf{M}\mathbf{N}|$; (133) follows since we have $\lambda_i(\mathbf{I}+\mathbf{C}^{1/2}\mathbf{B}\mathbf{C}^{1/2}) \geq 1$ for $i=1,\dots,N$, since $\lambda_i(\mathbf{I}+\mathbf{C}^{1/2}\mathbf{B}\mathbf{C}^{1/2}) = 1+\lambda_i(\mathbf{C}^{1/2}\mathbf{B}\mathbf{C}^{1/2})$ and $\lambda_i(\mathbf{C}^{1/2}\mathbf{B}\mathbf{C}^{1/2}) \geq 0$, since $\mathbf{C}^{1/2}\mathbf{B}\mathbf{C}^{1/2}$ is positive-definite. Similarly, $\lambda_i(\mathbf{I}+\mathbf{C}^{1/2}\mathbf{A}\mathbf{C}^{1/2}) \geq 1$ for $i=1,\dots,N$.

VII. Proof of the Inner Bound in Theorem 3

The proof of Theorem 3 is based on the proof of Theorem 3 from [3] and we provide an outline of it. The transmission is as follows. Transmitter l, $l=1,\dots,L$ sends $x_l^n(m_l)$, where $m_l \in [1, 2^{nR_l}]$. Relay k, $k=1,\dots,K$ compresses the channel output Y_k^n into U_k^n, indexed by z_k, where $i_k \in [1, 2^{n\hat{R}_k}]$ and sends a Wyner-Ziv bin index $j_t \in [1, 2^{nC_k}]$ to the CP over the error-free link. The destination receives $j_1,\dots,j_K$ and decodes jointly the compression codewords and the transmitted codewords, i.e., it jointly recovers the indices $(m_1,\dots,m_L,i_1,\dots,i_K)$. The detailed proof is given next. For simplicity of notation, we consider the case $Q=\emptyset$. Achievability for an arbitrary time-sharing random variable Q can be proved using the coded time-sharing technique.

Fix $\delta > 0$, non-negative rates $R_1,\dots,R_K$ and a joint pmf that factorizes as

$$p(q,x_{\mathcal{L}},y_{\mathcal{K}},u_{\mathcal{K}}) = p(q)\prod_{l=1}^{L} p(x_l|q)\ p(y_{\mathcal{K}}|x_{\mathcal{L}})\prod_{k=1}^{K} p(u_k|y_k,q). \tag{136}$$

Code Generation: For transmitter l, $l = 1, \ldots, L$ and every codebook realization F_l, generate a codebook $\mathcal{C}_l(F_l)$ consisting of a collection of 2^{nR_l} independent codewords $\{x_l(m_l, F_l)\}$ indexed with $m_l \in [1, 2^{nR_l}]$, where $x_l(m_l, F_l)$ has its elements generates i.i.d. according to $\prod_{i=1}^{n} p(x_i)$.

Let non-negative rates $\hat{R}_1, \ldots, \hat{R}_K$. For relay k, $k = 1, \ldots, K$, we generate a codebook $\mathcal{C}_k^r$ consisting of a collection of $2^{n\hat{R}_l}$ independent codewords $\{u_k^n(i_k)\}$ indexed with $i_k \in [1, 2^{n\hat{R}_k}]$, where codeword $u_k^n(i_k)$ has its elements generated i.i.d. according to $\prod_{i=1}^{n} p(u_i)$. Randomly and independently assign these codewords into 2^{nC_k} bins $\{\mathcal{B}_{j_k}\}$ indexed with $j_l = [1, 2^{nC_k}]$ containing $2^{n(\hat{R}_k - C_k)}$ codewords.

In the following, we drop the index F_l.

Encoding: Let $(m_1, \ldots, m_L)$ be the messages to be sent. Each node k, $k = 1, \ldots, K$ transmits the codeword $x_k^n(m_l)$.

Oblivious processing at relay k: Relay k finds an index i_k such that $u_k^n(i_k) \in \mathcal{C}_k^r$ is strongly ϵ-jointly typical with y_k^n. Using standard arguments, this can be accomplished with vanishing probability of error as long as n is large and

$$\hat{R}_l \geq I(Y_l; U_l). \tag{137}$$

Let j_k be such that $u_k(i_k) \in \mathcal{B}_{j_k}$. Relay k then forwards the bin index j_k to the CP through the error-free link.

Decoding: The CP collects all the bin indices $j_{\mathcal{K}} = (j_1, \ldots, j_K)$ from the error-free link and finds the set of indices $\hat{i}_{\mathcal{K}} = (\hat{i}_1, \ldots, \hat{i}_K)$ of the compressed vectors $u_{\mathcal{K}}^n$ and the transmitted messages $\hat{m}_{\mathcal{L}} = (\hat{m}_1, \ldots, \hat{m}_L)$, such that

$$(x_1^n(m_1), \ldots, x_L(m_L), u_1(i_1), \ldots, u_K(i_K)) \quad \text{jointly typical} \tag{138}$$

$$u_k^n(\hat{i}_k) \in \mathcal{B}_{j_k} \quad \text{for } k = 1, \ldots, K. \tag{139}$$

An error event is declared if $\hat{m}_{\mathcal{L}} \neq m_{\mathcal{L}}$. Assume that for some $\mathcal{T} \subseteq \mathcal{L}$ and $\mathcal{S} \subseteq \mathcal{K}$, we have $\hat{m}_{\mathcal{T}} \neq m_{\mathcal{T}}$ and $\hat{i}_{\mathcal{S}} \neq i_{\mathcal{S}}$ and $\hat{m}_{\mathcal{T}^c} = m_{\mathcal{T}^c}$ and $\hat{i}_{\mathcal{S}^c} = i_{\mathcal{S}^c}$. So, with high probability,

$(x^n_{\mathcal{T}}(\hat{m}_{\mathcal{T}}), x^n_{\mathcal{T}^c}(\hat{m}_{\mathcal{T}^c}), u^n_{\mathcal{S}}(i_{\mathcal{S}}), u^n_{\mathcal{S}}(i^c_{\mathcal{S}}))$ belongs to a typical set with distribution

$$\prod_{i=1}^{n}\left(P_{U_{\mathcal{S}^c},X_{\mathcal{T}^c}}(u_{\mathcal{S}^c,i}, x_{\mathcal{T}^c,i})\prod_{s\in\mathcal{S}}P_{U_s}(u_{s,i})\prod_{t\in\mathcal{T}}P_{X_t}(x_{t,i})\right). \tag{140}$$

According to [3, Lemma 3], the probability that $(x^n_{\mathcal{T}}(\hat{m}_{\mathcal{T}}), x^n_{\mathcal{T}^c}(\hat{m}_{\mathcal{T}^c}), u^n_{\mathcal{S}}(i_{\mathcal{S}}), u^n_{\mathcal{S}}(i^c_{\mathcal{S}}))$ is jointly typical is upper bounded by

$$2^{-n[H(U_{\mathcal{S}^c},X_{\mathcal{T}^c})-H(U_{\mathcal{K}},X_{\mathcal{L}})+\sum_{s\in\mathcal{S}}H(U_s)+\sum_{t\in\mathcal{T}}H(X_t)]}. \tag{141}$$

Overall, there are

$$2^{n(\sum_{j\in\mathcal{T}}R_j+\sum_{s\in\mathcal{S}}[\hat{R}_s-C_s])}-1, \tag{142}$$

such sequences $(x^n_{\mathcal{T}}(\hat{m}_{\mathcal{T}}), x^n_{\mathcal{T}^c}(\hat{m}_{\mathcal{T}^c}), u^n_{\mathcal{S}}(i_{\mathcal{S}}), u^n_{\mathcal{S}}(i^c_{\mathcal{S}}))$ in the set $\mathcal{B}_{j_1}\times\cdots\times\mathcal{B}_{j_K}$. This means that the CP is able to reliably decode $m_{\mathcal{L}}$ and $i_{\mathcal{K}}$ as long as $(R_1,\ldots,R_L)$ satisfy, for all $\mathcal{T}\subseteq\mathcal{L}$ and for all $\mathcal{S}\subseteq\mathcal{K}$

$$\begin{aligned}
\sum_{t\in\mathcal{T}}R_t &\leq \sum_{s\in\mathcal{S}}[C_s-\hat{R}_t]+H(U_{\mathcal{S}^c},X_{\mathcal{T}^c})-H(U_{\mathcal{K}},X_{\mathcal{L}})+\sum_{s\in\mathcal{S}}H(U_s)+\sum_{t\in\mathcal{T}}H(X_t) && (143)\\
&=\sum_{s\in\mathcal{S}}[C_s-H(U_s|Y_s)]+H(U_{\mathcal{S}^c},X_{\mathcal{T}^c})-H(U_{\mathcal{K}},X_{\mathcal{L}})+\sum_{t\in\mathcal{T}}H(X_t) && (144)\\
&=\sum_{s\in\mathcal{S}}[C_s-H(U_s|Y_s)]+H(U_{\mathcal{S}^c},X_{\mathcal{T}^c})-H(U_{\mathcal{K}},X_{\mathcal{L}})+H(X_{\mathcal{T}}) && (145)\\
&=\sum_{s\in\mathcal{S}}[C_s-H(U_s|Y_s)]+H(U_{\mathcal{S}^c},X_{\mathcal{T}^c})-H(U_{\mathcal{K}},X_{\mathcal{T}^c}|X_{\mathcal{T}}) && (146)\\
&=\sum_{s\in\mathcal{S}}[C_s-H(U_s|Y_s)]+H(X_{\mathcal{T}^c})+H(U_{\mathcal{S}^c}|X_{\mathcal{T}^c})-H(X_{\mathcal{T}^c}|X_{\mathcal{T}})-H(U_{\mathcal{K}}|X_{\mathcal{L}}) && (147)\\
&=\sum_{s\in\mathcal{S}}[C_s-H(U_s|Y_s)]+H(U_{\mathcal{S}^c}|X_{\mathcal{T}^c})-H(U_{\mathcal{K}}|X_{\mathcal{L}}) && (148)\\
&=\sum_{s\in\mathcal{S}}[C_s-H(U_s|Y_s)]+I(U_{\mathcal{S}^c};X_{\mathcal{T}}|X_{\mathcal{T}^c})-H(U_{\mathcal{S}}|X_{\mathcal{L}},U_{\mathcal{S}^c}) && (149)\\
&=\sum_{s\in\mathcal{S}}C_s-H(U_{\mathcal{S}}|Y_{\mathcal{S}},X_{\mathcal{L}},U_{\mathcal{S}^c})+I(U_{\mathcal{S}^c};X_{\mathcal{T}}|X_{\mathcal{T}^c})-H(U_{\mathcal{S}}|X_{\mathcal{L}},U_{\mathcal{S}^c}) && (150)
\end{aligned}$$

$$= \sum_{s\in\mathcal{S}} C_s - I(U_{\mathcal{S}}; Y_{\mathcal{S}}|X_{\mathcal{L}}, U_{\mathcal{S}^c}) + I(U_{\mathcal{S}^c}; X_{\mathcal{T}}|X_{\mathcal{T}^c}), \tag{151}$$

where (144) follows from (137); (145) follows due to the independence of $\mathcal{X}_{\mathcal{T}}$; (148) follows due to the independence of $X_{\mathcal{T}^c}$ and $X_{\mathcal{T}}$; (150) follows due to the Markov chain

$$U_k -\!\!\circ\!\!- Y_k - (X_{\mathcal{L}}, U_{\mathcal{K}/k}) \tag{152}$$

This completes the proof of Theorem 3 .

VIII. Proof of the Outer Bound in Theorem 4

Suppose the point $(R_1, \ldots, R_L)$ 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 notation simplicity we define $X_{\mathcal{L}}^n \triangleq (X_1^n, \ldots, X_L^n)$. Define

$$U_{i,k} \triangleq (J_k, Y_{\mathcal{K}}^{i-1}) \quad \text{and} \quad \bar{Q}_i \triangleq (X_{\mathcal{L}}^{i-1}, X_{\mathcal{L},i+1}^n, \tilde{Q}). \tag{153}$$

From Fano's inequality, we have with $\epsilon_n \to 0$ for $n \to \infty$ (for vanishing probability of error),

$$H(m_{\mathcal{T}}|J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) \leq H(m_{\mathcal{L}}|J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) \leq n\epsilon_n. \tag{154}$$

First, we show the following inequality, which will be used in the proof.

$$H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, J_{\mathcal{K}}, \tilde{Q}) \leq n\Gamma_{\mathcal{T}} \triangleq \sum_{i=1}^n H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, \bar{Q}_i) - n\sum_{t\in\mathcal{T}} R_t. \tag{155}$$

Inequality (155) can be shown as follows. From the destination side, we have

$$n\sum_{t\in\mathcal{T}} R_t = H(m_{\mathcal{T}}) \tag{156}$$

$$= I(m_{\mathcal{T}}; J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) + H(m_{\mathcal{T}}|J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) \tag{157}$$

$$= I(m_{\mathcal{T}}; J_{\mathcal{K}}, F_{\mathcal{T}}|F_{\mathcal{T}^c}, \tilde{Q}) + H(m_{\mathcal{T}}|J_{\mathcal{K}}, F_{\mathcal{L}}, \tilde{Q}) \tag{158}$$

$$\leq I(m_{\mathcal{T}}; J_{\mathcal{K}}, F_{\mathcal{T}}|F_{\mathcal{T}^c}, \tilde{Q}) + n\epsilon_n \tag{159}$$

$$= H(J_{\mathcal{K}}, F_{\mathcal{T}}|F_{\mathcal{T}^c}, \tilde{Q}) - H(J_{\mathcal{K}}, F_{\mathcal{T}}|F_{\mathcal{T}^c}, m_{\mathcal{T}}, \tilde{Q}) + n\epsilon_n \tag{160}$$

$$=H(J_{\mathcal{K}}|F_{\mathcal{T}^c},\tilde{Q})+H(F_{\mathcal{T}}|F_{\mathcal{T}^c},J_{\mathcal{K}},\tilde{Q})-H(F_{\mathcal{T}}|F_{\mathcal{T}^c},m_{\mathcal{T}},\tilde{Q})-H(J_{\mathcal{K}}|F_{\mathcal{T}^c},m_{\mathcal{T}},F_{\mathcal{T}},\tilde{Q})+n\epsilon_n \tag{161}$$

$$=I(m_{\mathcal{T}},F_{\mathcal{T}};J_{\mathcal{K}}|F_{\mathcal{T}^c},\tilde{Q})-I(F_{\mathcal{T}};J_{\mathcal{K}}|F_{\mathcal{T}^c},\tilde{Q})+n\epsilon_n \tag{162}$$

$$\leq I(m_{\mathcal{T}},F_{\mathcal{T}};J_{\mathcal{K}}|F_{\mathcal{T}^c},\tilde{Q})+n\epsilon_n \tag{163}$$

$$\leq I(X^n_{\mathcal{T}};J_{\mathcal{K}}|F_{\mathcal{T}^c},\tilde{Q})+n\epsilon_n \tag{164}$$

$$=H(X^n_{\mathcal{T}}|F_{\mathcal{T}^c},\tilde{Q})-H(X^n_{\mathcal{T}}|F_{\mathcal{T}^c},J_{\mathcal{K}},\tilde{Q})+n\epsilon_n \tag{165}$$

$$\leq H(X^n_{\mathcal{T}}|X^n_{\mathcal{T}^c},\tilde{Q})-H(X^n_{\mathcal{T}}|X^n_{\mathcal{T}^c},F_{\mathcal{T}^c},J_{\mathcal{K}},\tilde{Q})+n\epsilon_n \tag{166}$$

$$=H(X^n_{\mathcal{T}}|X^n_{\mathcal{T}^c},\tilde{Q})-H(X^n_{\mathcal{T}}|X^n_{\mathcal{T}^c},J_{\mathcal{K}},\tilde{Q})+n\epsilon_n, \tag{167}$$

where (156) follows since $m_{\mathcal{T}}$ are independent; (158) follows since $m_{\mathcal{T}}$ is independent of $\tilde{Q}$ and $F_{\mathcal{T}^c}$; (159) follows from (154); (162) follows since $m_{\mathcal{T}}$ is independent of $F_{\mathcal{L}}$; (164) follows from the data processing inequality;(166) follows since $X^n_{\mathcal{T}^c}, F_{\mathcal{T}^c}$ are independent from $X^n_{\mathcal{T}}$ and since conditioning reduces entropy and (167) follows due to the Markov chain

$$X^n_{\mathcal{T}} -\!\!\circ\!\!- (X^n_{\mathcal{T}^c},J_{\mathcal{K}},\tilde{Q}) -\!\!\circ\!\!- F_{\mathcal{T}^c}. \tag{168}$$

Then, from (167) we have (155) as follows:

$$\begin{aligned}
H(X^n_{\mathcal{T}}|X^n_{\mathcal{T}^c},J_{\mathcal{K}},\tilde{Q}) &\leq H(X^n_{\mathcal{T}}|X^n_{\mathcal{T}^c},\tilde{Q})-n\sum_{t\in\mathcal{T}}R_t-n\epsilon_n && (169)\\
&\leq \sum_{i=1}^{n}H(X_{\mathcal{T},i}|X^n_{\mathcal{T}^c},X^{i-1}_{\mathcal{T}},\tilde{Q})-n\sum_{t\in\mathcal{T}}R_t && (170)\\
&= \sum_{i=1}^{n}H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i},X^{i-1}_{\mathcal{L}},X^n_{\mathcal{L},i+1},\tilde{Q})-n\sum_{t\in\mathcal{T}}R_t && (171)\\
&= \sum_{i=1}^{n}H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i},\bar{Q}_i)-n\sum_{t\in\mathcal{T}}R_t=n\Gamma_{\mathcal{T}}. && (172)
\end{aligned}$$

where (171) is due to Lemma 1 .

Continuing from (167), we have

$$\begin{aligned}
n\sum_{t\in\mathcal{T}} R_t \leq & H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n,\tilde{Q}) - H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, J_{\mathcal{K}},\tilde{Q}) + n\epsilon_n && (173)\\
=& \sum_{i=1}^n H(X_{\mathcal{T},i}|X_{\mathcal{T}^c}^n,\tilde{Q},X_{\mathcal{T}}^{i-1}) - H(X_{\mathcal{T},i}^n|X_{\mathcal{T}^c}^n, J_{\mathcal{K}}, X_{\mathcal{T}}^{i-1},\tilde{Q}) + n\epsilon_n && (174)\\
=& \sum_{i=1}^n H(X_{\mathcal{T},i}|X_{\mathcal{T}^c}^n,\tilde{Q},X_{\mathcal{T}}^{i-1},X_{\mathcal{T},i+1}^n) - H(X_{\mathcal{T},i}|X_{\mathcal{T}^c}^n, J_{\mathcal{K}}, X_{\mathcal{T}}^{i-1},\tilde{Q}) + n\epsilon_n && (175)\\
\leq& \sum_{i=1}^n H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i},\bar{Q}_i) - H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, J_{\mathcal{K}}, Y_{\mathcal{K}}^{i-1}, X_{\mathcal{L}}^{i-1}, X_{\mathcal{L},i+1}^n,\tilde{Q}) + n\epsilon_n && (176)\\
=& \sum_{i=1}^n H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i},\bar{Q}_i) - H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, U_{\mathcal{K},i},\bar{Q}_i) + n\epsilon_n && (177)\\
=& \sum_{i=1}^n I(X_{\mathcal{T},i}; U_{\mathcal{K},i}|X_{\mathcal{T}^c,i},\bar{Q}_i) + n\epsilon_n, && (178)
\end{aligned}$$

where (175) follows due to Lemma 1 and (176) follows since conditioning reduces entropy.

On the other hand, we have the following inequality

$$\begin{aligned}
I(Y_{\mathcal{K}}^n; J_{\mathcal{S}}|X_{\mathcal{L}}^n, J_{\mathcal{S}^c},\tilde{Q}) &= \sum_{i=1}^n I(Y_{\mathcal{K},i}; J_{\mathcal{S}}|X_{\mathcal{L}}^n, J_{\mathcal{S}^c},\tilde{Q}, Y_{\mathcal{K}}^{i-1}) && (179)\\
&= \sum_{i=1}^n I(Y_{\mathcal{K},i}; J_{\mathcal{S}}, Y_{\mathcal{K}}^{i-1}|X_{\mathcal{L}}^n, J_{\mathcal{S}^c},\tilde{Q}, Y_{\mathcal{K}}^{i-1}) && (180)\\
&= \sum_{i=1}^n I(Y_{\mathcal{K},i}; U_{\mathcal{S},i}|X_{\mathcal{L},i}, U_{\mathcal{S}^c,i},\bar{Q}_i) && (181)\\
&\geq \sum_{i=1}^n I(Y_{\mathcal{S},i}; U_{\mathcal{S},i}|X_{\mathcal{L},i}, U_{\mathcal{S}^c,i},\bar{Q}_i). && (182)
\end{aligned}$$

Then, from the relay side we have,

$$\begin{aligned}
n\sum_{k\in\mathcal{S}} C_k \geq & \sum_{k\in\mathcal{S}} H(J_k) && (183)\\
\geq & H(J_{\mathcal{S}}) && (184)\\
\geq & H(J_{\mathcal{S}}|X_{\mathcal{T}^c}^n, J_{\mathcal{S}^c},\tilde{Q}) && (185)
\end{aligned}$$

$$\geq I(Y_{\mathcal{K}}^n; J_{\mathcal{S}}|X_{\mathcal{T}^c}^n, J_{\mathcal{S}^c}, \tilde{Q}) \tag{186}$$

$$= I(X_{\mathcal{T}}^n, Y_{\mathcal{K}}^n; J_{\mathcal{S}}|X_{\mathcal{T}^c}^n, J_{\mathcal{S}^c}, \tilde{Q}) \tag{187}$$

$$= I(X_{\mathcal{T}}^n; J_{\mathcal{S}}|X_{\mathcal{T}^c}^n, J_{\mathcal{S}^c}, \tilde{Q}) + I(Y_{\mathcal{K}}^n; J_{\mathcal{S}}|X_{\mathcal{L}}^n, J_{\mathcal{S}^c}, \tilde{Q}) \tag{188}$$

$$= H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, J_{\mathcal{S}^c}, \tilde{Q}) - H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, J_{\mathcal{K}}, \tilde{Q}) + I(Y_{\mathcal{K}}^n; J_{\mathcal{S}}|X_{\mathcal{L}}^n, J_{\mathcal{S}^c}, \tilde{Q}) \tag{189}$$

$$\geq H(X_{\mathcal{T}}^n|X_{\mathcal{T}^c}^n, J_{\mathcal{S}^c}, \tilde{Q}) - n\Gamma_{\mathcal{T}} + I(Y_{\mathcal{K}}^n; J_{\mathcal{S}}|X_{\mathcal{L}}^n, J_{\mathcal{S}^c}, \tilde{Q}) \tag{190}$$

$$= \sum_{i=1}^n H(X_{\mathcal{T},i}|X_{\mathcal{T}^c}^n, J_{\mathcal{S}^c}, X_{\mathcal{T}}^{i-1}, \tilde{Q}) - n\Gamma_{\mathcal{T}} + I(Y_{\mathcal{K}}^n; J_{\mathcal{S}}|X_{\mathcal{L}}^n, J_{\mathcal{S}^c}, \tilde{Q}) \tag{191}$$

$$\geq \sum_{i=1}^n H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, U_{\mathcal{S}^c,i}, \bar{Q}_i) - n\Gamma_{\mathcal{T}} + I(Y_{\mathcal{K}}^n; J_{\mathcal{S}}|X_{\mathcal{L}}^n, J_{\mathcal{S}^c}, \tilde{Q}) \tag{192}$$

$$= \sum_{i=1}^n H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, U_{\mathcal{S}^c,i}, \bar{Q}_i) - H(X_{\mathcal{T},i}|X_{\mathcal{T}^c,i}, \bar{Q}_i) \tag{193}$$

$$+ n\sum_{t\in\mathcal{T}} R_t + \sum_{i=1}^n I(Y_{\mathcal{S},i}; U_{\mathcal{S},i}|X_{\mathcal{L},i}, U_{\mathcal{S}^c,i}, \bar{Q}_i) \tag{194}$$

$$= \sum_{i=1}^n I(Y_{\mathcal{S},i}; U_{\mathcal{S},i}|X_{\mathcal{L},i}, U_{\mathcal{S}^c,i}, \bar{Q}_i) + n\sum_{t\in\mathcal{T}} R_t - \sum_{i=1}^n I(X_{\mathcal{T},i}; U_{\mathcal{S}^c,i}|X_{\mathcal{T}^c,i}, \bar{Q}_i) \tag{195}$$

where: (187) follows since $J_{\mathcal{S}}$ is a function of $Y_{\mathcal{S}}^n$; (190) follows from (155); (192) follows since conditioning reduces entropy; and (194) follows from (155) and (182).

We define the standard time-sharing variable Q' uniformly distributed over $\{1, \dots, n\}$, $X_{\mathcal{L}} \triangleq X_{\mathcal{L},Q'}$, $Y_k \triangleq Y_{k,Q'}$, $U_k \triangleq U_{k,Q'}$ and $Q \triangleq [\bar{Q}_{Q'}, Q']$ and we have from (178)

$$n\sum_{t\in\mathcal{T}} R_t \leq \sum_{i=1}^n I(X_{\mathcal{T},i}; U_{\mathcal{K},i}|X_{\mathcal{T},i}, \bar{Q}_i) + n\epsilon_n \tag{196}$$

$$= nI(X_{\mathcal{T},Q'}; U_{\mathcal{K},Q'}|X_{\mathcal{T}^c,Q'}, \bar{Q}_{Q'}, Q') + n\epsilon_n \tag{197}$$

$$= nI(X_{\mathcal{T}}; U_{\mathcal{K}}|X_{\mathcal{T}^c}, Q) + n\epsilon_n \tag{198}$$

and similarly, from (195), we have

$$\sum_{t\in\mathcal{T}} R_t \leq \sum_{k\in\mathcal{S}} C_k - I(Y_{\mathcal{S}}; U_{\mathcal{S}}|X_{\mathcal{L}}, U_{\mathcal{S}^c}, Q) + I(X_{\mathcal{L}}; U_{\mathcal{S}^c}|X_{\mathcal{T}^c}, Q). \tag{199}$$

Define $W_{Q'} \triangleq (Y_{\mathcal{K}}^{Q'-1}, Y_{\mathcal{K},Q'+1}^{n})$, so that from Lemma 1 , $X_{\mathcal{L},Q'}$ and $Y_{\mathcal{K},Q'}$ are independent with $W \triangleq W_{Q'}$ when not conditioned on $F_{\mathcal{L}}$. Note that in general, $\bar{Q}_{Q'}$ is not independent of $X_{\mathcal{L},Q'}, Y_{\mathcal{K},Q'}$. Then, conditioned on Q, the auxiliary variables $U_{k,Q'}$ can be represented as

$$U_{k,Q'} = (J_k, Y_{\mathcal{K}}^{Q'-1}) \tag{200}$$

$$= f'_k(W_{Q'}, Y_{k,Q'}) \tag{201}$$

$$= f_k(W, Y_k, Q) \tag{202}$$

where $f'_k(W_{Q'}, Y_{k,Q'}) = (J_k, Y_{\mathcal{K}}^{Q'-1})$. Note that this implies that, conditioned on $\bar{Q}_i$, we have the Markov chain

$$U_k -\!\!\circ\!\!- Y_k -\!\!\circ\!\!- (X_{\mathcal{L}}, Y_{\mathcal{K}\setminus k}). \tag{203}$$

and the Markov chain

$$U_k -\!\!\circ\!\!- (Y_k, W) -\!\!\circ\!\!- (X_{\mathcal{L}}, Y_{\mathcal{K}\setminus k}, U_{\mathcal{K}\setminus k}). \tag{204}$$

This completes the proof of Theorem 4 .

REFERENCES

[1] 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, May 2011.

[2] 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.

[3] 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.

[4] 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.

[5] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. on Information Theory*, Feb. 2004.

[6] T. M. Cover, "Broadcast channels," *IEEE Trans. on Information Theory*, vol. 18, no. 1, pp. 2–14, January 1972.

[7] 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.

[8] 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.

[9] K. V. Bhagwat and R. Subramanian, “Inequalities between means of positive operators,” *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 83, no. 3, pp. 393–401, 005 1978.

[10] B. Mond and J. Pečarić, “A mixed arithmetic-mean-harmonic-mean matrix inequality,” *Linear Algebra and its Applications*, vol. 237, pp. 449 – 454, 1996.

[11] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[12] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge Univ Pr, 2004.

[13] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.