Broadcast Capacity Regions with Three Receivers and Message Cognition

Tobias J. Oechtering
KTH Royal Institute of Technology
ACCESS Linnaeus Center
Stockholm, Sweden

Michèle Wigger
Telecom ParisTech
Paris, France
michele.wigger@telecom-paristech.fr

Roy Timo
Institute for Telecommunications Research
University of South Australia
roy.timo@unisa.edu.au

***Abstract*—We consider the capacity region of a three receiver broadcast channel with some message cognition at two receivers. The problem generalizes the bi-directional broadcast channel to include a third receiver, a common message, and (partial) message cognition. We characterize the capacity region for several classes of less noisy, more capable, and deterministic broadcast channels.**

I. Introduction

A canonical cooperative-communications problem is the bi-directional broadcast channel (BC) [1], which is a special case of the almost lossless joint source-channel coding problem in [2]. In this paper we generalize the bi-directional BC to the problem in Fig. 1, i.e., to include a third receiver and a common message. Unlike the bi-directional BC [1], our setup of Fig. 1 is not a special case of [2], but rather closely related to the lossy joint source-channel coding problem in [3].

We will also extend the setup in Fig. 1 to a setup where Receivers 1 and 2 have only partial cognition of each other's messages. The capacity region for the two-user BC with degraded message sets and partial message cognition was first studied in [4].

Our goal is to determine the capacity region of the setup in Fig. 1 as well as the capacity region of the extended setup with partial message cognition.

In Fig. 1, the Transmitter wishes to send three messages, M_0, M_1, and M_2, to the receivers. Receiver 1 requires the *private* message M_1 and the *common* message M_0; Receiver 2 requires the *private* message M_2 and M_0; and Receiver 3 requires only M_0. The messages are modelled as independent random variables, where each M_k is uniformly distributed over the set $\mathcal{M}_k \triangleq \{1, 2, \ldots, \lfloor 2^{nR_k} \rfloor\}$, for $k \in \{0, 1, 2\}$. Here R_k denotes the transmission rate of message M_k, and n denotes the blocklength.

The BC is discrete and memoryless. Denoting by $\mathcal{X}$ the channel input alphabet and by $\mathcal{Y}_k$ the channel output alphabet at Receiver k, the channel input $X^n \triangleq (X_1, \ldots, X_n)$ takes value in $\mathcal{X}^n$ and Receiver k's outputs $Y_k^n \triangleq (Y_{k,1}, \ldots, Y_{k,n})$ take value in $\mathcal{Y}_k^n$, for $k \in \{1, 2, 3\}$. We consider a memoryless BC so that the conditional distribution of (Y_1^n, Y_2^n, Y_3^n) given X^n is defined by $P_{Y_1,Y_2,Y_3|X}(y_1, y_2, y_3|x)$.

Paper Outline: In Section II, we consider the setup in Fig. 1, where Receiver 1 (resp. 2) is fully cognizant of Message M_2 (resp. M_1). In Section III, we consider the extended setup with partial message cognition; i.e., Receiver 1 (resp. 2) knows only part of the message M_2 (resp. M_1).

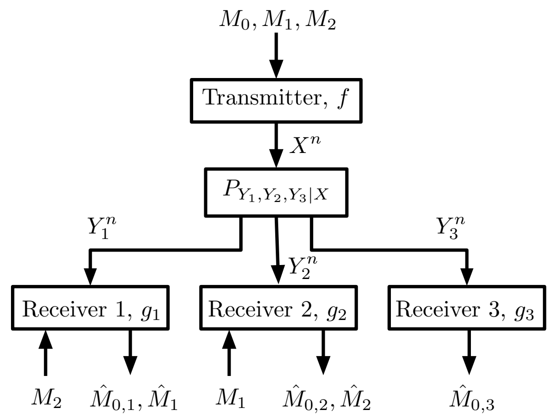

Fig. 1. Broadcast channel with three receivers and message cognition.

II. Full Cognition at Receivers 1 and 2

Throughout this section we assume full cognition at Receivers 1 and 2, i.e., that Receiver 1 is cognizant of the entire message M_2 and Receiver 2 of the entire message M_1.

A code consists of four maps: An encoder at the Transmitter,

$$f : \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{X}^n$$

and a decoder at each receiver,

$$\begin{aligned} g_1 &: \mathcal{Y}_1^n \times \mathcal{M}_2 \to \mathcal{M}_0 \times \mathcal{M}_1 \\ g_2 &: \mathcal{Y}_2^n \times \mathcal{M}_1 \to \mathcal{M}_0 \times \mathcal{M}_2 \\ g_3 &: \mathcal{Y}_3^n \to \mathcal{M}_0. \end{aligned}$$

The Transmitter sends $X^n = f(M_0, M_1, M_2)$; Receiver 1 decodes $(\hat{M}_{0,1}, \hat{M}_1) = g_1(Y_1^n, M_2)$; Receiver 2 decodes $(\hat{M}_{0,2}, \hat{M}_2) = g_2(Y_2^n, M_1)$; and Receiver 3 decodes $\hat{M}_{0,3} = g_3(Y_3^n)$. The *average joint probability of error* is

$$P_{\mathrm{e}} \triangleq \mathbb{P}\big[\hat{M}_{0,3} \neq M_0 \text{ or } (\hat{M}_{0,1}, \hat{M}_1) \neq (M_0, M_1) \\ \text{or } (\hat{M}_{0,2}, \hat{M}_2) \neq (M_0, M_2)\big].$$

The rates (R_0, R_1, R_2) are said to be *achievable* if for each $\epsilon > 0$ there exists a code (f, g_1, g_2, g_3) with $P_{\mathrm{e}} \leq \epsilon$ for some sufficiently large blocklength n. The *capacity region* $\mathcal{C}$ is the closure of the set of all achievable rates.

We first give an inner bound for $\mathcal{C}$. This bound is achieved using a combination of superposition coding, rate-splitting, and bi-directional coding. Roughly, the superposition cloud-centers carry the common message M_0 and the satellites simultaneously carry the bi-directional messages (M_1, M_2). Rate-splitting is used to transfer rate from the satellites to the cloud-centers.

Let $\mathcal{R}_{\text{in}}^*$ denote the set of rate tuples (R_0, R_1, R_2) satisfying

$$R_0 \leq I(U; Y_3) \tag{1a}$$

$$R_0 + R_1 \leq \min\{I(X; Y_1), I(U; Y_3) + I(X; Y_1|U)\} \tag{1b}$$

$$R_0 + R_2 \leq \min\{I(X; Y_2), I(U; Y_3) + I(X; Y_2|U)\}, \tag{1c}$$

for some (U, X) with $U \multimap X \multimap (Y_1, Y_2, Y_3)$.

Proposition 1: $\mathcal{C} \supseteq \mathcal{R}_{\text{in}}^*$, and $\mathcal{R}_{\text{in}}^*$ is convex.

Proof: The proposition is a corollary of Theorem 2, which is given later in Section III. See Remark 3. ■

Remark 1: The capacity region in [1, Thm. 2.5] can be recovered from Proposition 1 by setting U to be constant.

The next theorem gives five non-trivial settings for which the inclusion in Proposition 1 is an equality. We first need two definitions from [5, p. 121]. A channel output Y_i is said to be *less noisy* than another output Y_j (abbreviated $Y_i \succeq Y_j$) if $I(U; Y_i) \geq I(U; Y_j)$ holds for every auxiliary random variable U with $U \multimap X \multimap (Y_i, Y_j)$. A channel output Y_i is said to be *more capable* than another output Y_j if $I(X; Y_i) \geq I(X; Y_j)$ holds for every X.

Theorem 1: $\mathcal{R}_{\text{in}}^* = \mathcal{C}$ in each of the following settings.

(i) If $Y_1 \succeq Y_3$ and $Y_2 \succeq Y_3$, then $\mathcal{C}$ is equal to the set of rate tuples (R_0, R_1, R_2) satisfying

$$R_1 \leq I(X; Y_1|U)$$
$$R_2 \leq I(X; Y_2|U)$$
$$R_0 \leq I(U; Y_3)$$

for some (U, X) with $U \multimap X \multimap (Y_1, Y_2, Y_3)$.

(ii) If Y_3 is more capable than Y_1 and Y_2, then $\mathcal{C}$ is equal to the set of all rate tuples (R_0, R_1, R_2) satisfying

$$R_0 + R_1 \leq I(X; Y_1)$$
$$R_0 + R_2 \leq I(X; Y_2)$$

for some X.

(iii) If $Y_3 \succeq Y_1$, then $\mathcal{C}$ is equal to the set of rate tuples (R_0, R_1, R_2) satisfying

$$R_0 + R_1 \leq I(X; Y_1)$$
$$R_0 + R_2 \leq I(X; Y_2)$$
$$R_0 + R_2 \leq I(U; Y_3) + I(X; Y_2|U)$$
$$R_0 \leq I(U; Y_3)$$

for some (U, X) with $U \multimap X \multimap (Y_1, Y_2, Y_3)$.

(iv) If the marginal conditional distributions $p_{Y_1|X}$ and $p_{Y_2|X}$ are the same, then $\mathcal{C}$ is equal to the set of rate tuples (R_0, R_1, R_2) satisfying

$$R_0 + R_1 \leq I(X; Y_1)$$
$$R_0 + R_2 \leq I(X; Y_2)$$
$$R_0 + R_1 \leq I(U; Y_3) + I(X; Y_1|U)$$
$$R_0 + R_2 \leq I(U; Y_3) + I(X; Y_2|U)$$
$$R_0 \leq I(U; Y_3)$$

for some (U, X) with $U \multimap X \multimap (Y_1, Y_2, Y_3)$.

(v) If Y_3 is a deterministic function of X, then $\mathcal{C}$ is equal to the set of all rate tuples (R_0, R_1, R_2) satisfying

$$R_0 \leq H(Y_3)$$
$$R_0 + R_1 \leq I(X; Y_1)$$
$$R_0 + R_2 \leq I(X; Y_2)$$

for some X.

Proof: The proof for (i), (iii), and (iv) are omitted due to space constraints. A sketch of the proof for case (v) can be found in Section IV. The direct part for (ii) follows by setting $U = X$ in (1) and using the *more capable* condition. The converse to (ii) is trivial. ■

III. Partial Cognition at Receivers 1 and 2

A natural generalization of the setup in Fig. 1 is to vary the quantity of side information at Receivers 1 and 2, as it was done for the two-receiver BC setup in [4]. Specifically, suppose that the bi-directional messages take the form

$$M_k = (M_{k,\text{c}}, M_{k,\text{p}}), \quad k = 1, 2,$$

where $M_{k,\text{c}}$ and $M_{k,\text{p}}$ are independent and uniformly distributed on $\{1, 2, \ldots, \lfloor 2^{nR_{k,\text{c}}} \rfloor\}$ and $\{1, 2, \ldots, \lfloor 2^{nR_{k,\text{p}}} \rfloor\}$, respectively. Receiver 1 is now cognizant of message $M_{2,\text{c}}$ – instead of M_2 – and is ignorant of $M_{2,\text{p}}$. Similarly, Receiver 2 is cognizant of $M_{1,\text{c}}$ and ignorant of $M_{1,\text{p}}$. The capacity region for the setup with partial cognition is defined analogously to $\mathcal{C}$; i.e., it is the set of all achievable rates $(R_0, R_{1,\text{c}}, R_{1,\text{p}}, R_{2,\text{c}}, R_{2,\text{p}})$. We let $\mathcal{C}_{\text{part}}$ denote this region. For brevity, we retain $R_1 = R_{1,\text{c}} + R_{1,\text{p}}$ and $R_2 = R_{2,\text{c}} + R_{2,\text{p}}$.

Remark 2: The partial-cognition setup includes the general two-receiver BC [5, Sect. 8] as a special case. Hence, we do not expect to completely characterise $\mathcal{C}_{\text{part}}$.

The next theorem is proved using a combination of superposition coding, rate-splitting, and bi-directional coding. Let $\mathcal{R}_{\text{in,part}}^{(1)}$ denote the set of all rates $(R_0, R_{1,\text{c}}, R_{1,\text{p}}, R_{2,\text{c}}, R_{2,\text{p}})$ satisfying

$$R_0 \leq I(U; Y_3) \tag{7a}$$

$$R_{1,\text{p}} \leq I(X; Y_1|V) \tag{7b}$$

$$R_0 + R_2 \leq \min\{I(V; Y_2), I(U; Y_3) + I(V; Y_2|U)\} \tag{7c}$$

$$R_0 + R_1 + R_{2,\text{p}} \leq \min\{I(X; Y_1), I(U; Y_3) + I(X; Y_1|U)\} \tag{7d}$$

for some (U, V, X) with $U \multimap V \multimap X \multimap (Y_1, Y_2, Y_3)$. Let $\mathcal{R}_{\text{in,part}}^{(2)}$ denote the set of all rates $(R_0, R_{1,\text{c}}, R_{1,\text{p}}, R_{2,\text{c}}, R_{2,\text{p}})$ satisfying (7) with indices 1 and 2 interchanged. Let

$$\mathcal{R}_{\text{in,part}} \triangleq \text{convex hull}\big(\mathcal{R}_{\text{in,part}}^{(1)} \cup \mathcal{R}_{\text{in,part}}^{(2)}\big). \tag{8}$$

Theorem 2: $\mathcal{C}_{\text{part}} \supseteq \mathcal{R}_{\text{in,part}}$.

Remark 3: The inner bound of Proposition 1 follows directly from Theorem 2 by setting $V = X$.

Proof of Theorem 2: We now sketch the coding theorem. *Code Construction:* Split the messages $M_{1,\mathrm{c}}, M_{2,\mathrm{c}}, M_{2,\mathrm{p}}$ as

$$M_{1,\mathrm{c}} = (M_{1,\mathrm{c}}^{(1)}, M_{1,\mathrm{c}}^{(2)}),\ M_{2,\mathrm{c}} = (M_{2,\mathrm{c}}^{(1)}, M_{2,\mathrm{c}}^{(2)}), \text{ and } M_{2,\mathrm{p}} = (M_{2,\mathrm{p}}^{(1)}, M_{2,\mathrm{p}}^{(2)})$$

with rates $R_{1,\mathrm{c}}^{(k)}$, $R_{2,\mathrm{c}}^{(k)}$, and $R_{2,\mathrm{p}}^{(k)}$, $k \in \{1,2\}$. Construct two new messages $M_{\oplus}^{(1)}$ and $M_{\oplus}^{(2)}$ as follows

$$M_{\oplus}^{(k)} = \left(M_{1,\mathrm{c}}^{(k)} + M_{2,\mathrm{c}}^{(k)}\right) \text{ modulo } 2^{n \max\{R_{1,\mathrm{c}}^{(k)}, R_{2,\mathrm{c}}^{(k)}\}}.$$

We use a three-layer superposition coding scheme. The cloud-center encodes M_0, $M_{\oplus}^{(1)}$, $M_{2,\mathrm{p}}^{(1)}$, the first satellite encodes $M_{\oplus}^{(2)}, M_{2,\mathrm{p}}^{(2)}$, and the top-most satellite encodes $M_{1,\mathrm{p}}$. For the random code construction we use the law P_U to generate the cloud centers, the conditional law $P_{V|U}$ for the first satellites, and the conditional law $P_{X|V}$ for the top-most satellites.

Decoding: Receiver 3 decodes the cloud center, Receiver 2 decodes the cloud center and the first satellite, and Receiver 1 decodes the cloud center and both satellites. Arbitrary small probability of error is achieved if

$$R_0 + R_{1,\mathrm{c}}^{(1)} + R_{2,\mathrm{p}}^{(1)} \le I(U;Y_3) \tag{9a}$$
$$R_0 + R_{2,\mathrm{c}}^{(1)} + R_{2,\mathrm{p}}^{(1)} \le I(U;Y_3) \tag{9b}$$
$$R_0 + R_2 \le I(V;Y_2) \tag{9c}$$
$$R_{2,\mathrm{c}}^{(2)} + R_{2,\mathrm{p}}^{(2)} \le I(V;Y_2|U) \tag{9d}$$
$$R_0 + R_1 + R_{2,\mathrm{p}} \le I(X;Y_1) \tag{9e}$$
$$R_{1,\mathrm{c}}^{(2)} + R_{1,\mathrm{p}} + R_{2,\mathrm{p}}^{(2)} \le I(X;Y_1|U) \tag{9f}$$
$$R_{1,\mathrm{p}} \le I(X;Y_1|V). \tag{9g}$$

Applying the Fourier-Motzkin elimination algorithm results in the rate constraints in (7). ■

Remark 4: For the region defined in (7), it can be shown following [5, Appendix C] that it suffices to consider auxiliary random variables $(U,V) \in \mathcal{U} \times \mathcal{V}$ with cardinality $|\mathcal{U}| \le |\mathcal{X}| + 4$ and $|\mathcal{V}| \le (|\mathcal{X}|+4)(|\mathcal{X}|+1)$. Tighter constraints can be obtained for some special cases.

Theorem 3: $\mathcal{C}_{\text{part}} = \mathcal{R}_{\text{in,part}}$ in each of the following settings.

(i) If $Y_1 \succeq Y_2 \succeq Y_3$, then $\mathcal{C}_{\text{part}}$ is the set of rates $(R_0, R_{1,\mathrm{c}}, R_{1,\mathrm{p}}, R_{2,\mathrm{c}}, R_{2,\mathrm{p}})$ satisfying

$$R_0 \le I(U;Y_3) \tag{10a}$$
$$R_{1,\mathrm{p}} \le I(X;Y_1|V) \tag{10b}$$
$$R_2 \le I(V;Y_2|U) \tag{10c}$$
$$R_1 + R_{2,\mathrm{p}} \le I(X;Y_1|U) \tag{10d}$$

for some (U,V,X) with $U -\!\!\circ\!\!- V -\!\!\circ\!\!- X -\!\!\circ\!\!- (Y_1,Y_2,Y_3)$.

(ii) If $Y_1 \succeq Y_3 \succeq Y_2$, then $\mathcal{C}_{\text{part}}$ is the set of rates $(R_0, R_{1,\mathrm{c}}, R_{1,\mathrm{p}}, R_{2,\mathrm{c}}, R_{2,\mathrm{p}})$ satisfying

$$R_0 \le I(U;Y_2) \tag{11a}$$
$$R_{1,\mathrm{p}} \le I(X;Y_1|V) \tag{11b}$$
$$R_2 \le I(V;Y_2|U) \tag{11c}$$
$$R_1 + R_{2,\mathrm{p}} \le I(X;Y_1|U) \tag{11d}$$

for some (U,V,X) with $U -\!\!\circ\!\!- V -\!\!\circ\!\!- X -\!\!\circ\!\!- (Y_1,Y_2,Y_3)$.

(iii) If $Y_3 \succeq Y_1 \succeq Y_2$, then $\mathcal{C}_{\text{part}}$ is the set of rates $(R_0, R_{1,\mathrm{c}}, R_{1,\mathrm{p}}, R_{2,\mathrm{c}}, R_{2,\mathrm{p}})$ satisfying

$$R_{1,\mathrm{p}} \le I(X;Y_1|U) \tag{12a}$$
$$R_0 + R_2 \le I(U;Y_2) \tag{12b}$$
$$R_0 + R_1 + R_{2,\mathrm{p}} \le I(X;Y_1) \tag{12c}$$

for some (U,X) with $U -\!\!\circ\!\!- X -\!\!\circ\!\!- (Y_1,Y_2,Y_3)$.

Proof: See Section IV. ■

When $Y_1 \succeq Y_3 \succeq Y_2$ or $Y_3 \succeq Y_1 \succeq Y_2$, the capacity region depends on the channel law $P_{Y_3|X}$ to Receiver 3 only through the fact that it must satisfy the less-noisy conditions.

We observe that when $Y_3 \succeq Y_1 \succeq Y_2$, a two-layer superposition coding scheme suffices to achieve capacity. Moreover, in this case, the result remains valid also when Y_3 is more capable than Y_1, but not less noisy.

Proposition 2: If Y_3 is more capable than Y_1 and $Y_3, Y_1 \succeq Y_2$, then $\mathcal{C}_{\text{part}}$ is the set of rates $(R_0, R_{1,\mathrm{c}}, R_{1,\mathrm{p}}, R_{2,\mathrm{c}}, R_{2,\mathrm{p}})$ satisfying (12) for some (U,X) with $U -\!\!\circ\!\!- X -\!\!\circ\!\!- (Y_1,Y_2,Y_3)$.

Proof: The converse follows by noting that our converse in Section IV-B for case (iii) only requires that $Y_1 \succeq Y_2$, and thus remains valid in this slightly weaker setup. The achievability follows by modifying our scheme achieving $\mathcal{R}_{\text{in,part}}^{(1)}$ so that Receiver 3 also decodes the two satellites, in addition to the cloud center. ■

Corollary 3.1: The capacity regions for the cases $Y_2 \succeq Y_1 \succeq Y_3$, $Y_2 \succeq Y_3 \succeq Y_1$, and $Y_3 \succeq Y_2 \succeq Y_1$ directly follow from the previous theorem. The regions are given by (10), (11), and (12) where we have to exchange the indices 1 and 2. Similarly, the capacity region for the case Y_3 more capable than Y_2 and $Y_1, Y_3 \succeq Y_2$ is given by (12) where again the indices 1 and 2 have to be exchanged.

In the usual way, Theorem 3 can be adapted to the Gaussian BC, where $Y_k = X + Z_k$ with $Z_k \sim \mathcal{N}(0, \sigma_k^2)$. For Gaussian BCs the capacity region takes on a particularly simple form. This can be proved with the entropy-power inequality and the maximal entropy property for a fixed second moment.

Corollary 3.2: Depending on the variances $\sigma_1^2, \sigma_2^2, \sigma_3^2 > 0$, the capacity $\mathcal{C}_{\text{part}}$ for Gaussian channels is given as follows.

(i) If $\sigma_3^2 \ge \sigma_2^2 \ge \sigma_1^2$, then $\mathcal{C}_{\text{part}}$ is the set of all rates $(R_0, R_{1,\mathrm{c}}, R_{1,\mathrm{p}}, R_{2,\mathrm{c}}, R_{2,\mathrm{p}})$ satisfying

$$R_0 \le \frac{1}{2}\log\left(1 + \frac{\alpha P}{(1-\alpha)P + \sigma_3^2}\right) \tag{13a}$$
$$R_1 \le \frac{1}{2}\log\left(1 + \frac{(1-\alpha-\beta)P}{\sigma_1^2}\right) + R_{1,\mathrm{c}} \tag{13b}$$
$$R_2 \le \frac{1}{2}\log\left(1 + \frac{\beta P}{(1-\alpha-\beta)P + \sigma_2^2}\right) \tag{13c}$$
$$R_1 + R_2 \le \frac{1}{2}\log\left(1 + \frac{(1-\alpha)P}{\sigma_1^2}\right) + R_{2,\mathrm{c}} \tag{13d}$$

for some $\alpha, \beta \in [0,1]$ such that $\alpha + \beta \le 1$.

(ii) If $\sigma_2^2 \ge \sigma_3^2 \ge \sigma_1^2$, then $\mathcal{C}_{\text{part}}$ is the set of rates

$(R_0, R_{1,\mathrm{c}}, R_{1,\mathrm{p}}, R_{2,\mathrm{c}}, R_{2,\mathrm{p}})$ satisfying

$$R_0 \leq \frac{1}{2}\log\left(1+\frac{\alpha P}{(1-\alpha)P+\sigma_2^2}\right) \quad (14a)$$

$$R_1 \leq \frac{1}{2}\log\left(1+\frac{(1-\alpha-\beta)P}{\sigma_1^2}\right)+R_{1,\mathrm{c}} \quad (14b)$$

$$R_2 \leq \frac{1}{2}\log\left(1+\frac{\beta P}{(1-\alpha-\beta)P+\sigma_2^2}\right) \quad (14c)$$

$$R_1+R_2 \leq \frac{1}{2}\log\left(1+\frac{(1-\alpha)P}{\sigma_1^2}\right)+R_{2,\mathrm{c}} \quad (14d)$$

for some $\alpha, \beta \in [0,1]$ such that $\alpha+\beta \leq 1$.

(iii) If $\sigma_2^2 \geq \sigma_1^2 \geq \sigma_3^2$, then $\mathcal{C}_{\text{part}}$ is the set of rates $(R_0, R_{1,\mathrm{c}}, R_{1,\mathrm{p}}, R_{2,\mathrm{c}}, R_{2,\mathrm{p}})$ satisfying

$$R_1 \leq \frac{1}{2}\log\left(1+\frac{(1-\alpha)P}{\sigma_1^2}\right)+R_{1,\mathrm{c}} \quad (15a)$$

$$R_0+R_2 \leq \frac{1}{2}\log\left(1+\frac{\alpha P}{(1-\alpha)P+\sigma_2^2}\right) \quad (15b)$$

$$R_0+R_1+R_2 \leq \frac{1}{2}\log\left(1+\frac{P}{\sigma_1^2}\right)+R_{2,\mathrm{c}}. \quad (15c)$$

for some $\alpha \in [0,1]$.

Remark 5: From the above capacity expressions we notice the following. In the above setups, when $R_{1,\mathrm{c}}=0$, i.e., Receiver 2 does not have any knowledge about Message M_1, then providing $M_{2,\mathrm{c}}$ (even when $M_{2,\mathrm{c}}=M_2$) to Receiver 1 does not increase capacity. In fact, providing $M_{2,\mathrm{c}}$ to Receiver 1 only increases the capacity when $R_{1,\mathrm{c}}$ is above a certain threshold that depends on the channel parameters. In contrast, providing $M_{1,\mathrm{c}}$ to Receiver 2 is always beneficial.

IV. Proofs

A. Proof of Deterministic Part of Theorem 1

We have $Y_3=\phi(X)$ for some deterministic $\phi: \mathcal{X} \to \mathcal{U}$. Recall Proposition 1 and (1). Choose $U=Y_3=\phi(X)$, so that

$$\begin{aligned}
R_0 &\leq H(Y_3)\\
R_0+R_1 &\leq \min\{I(X;Y_1), H(Y_3)+I(X;Y_1|Y_3)\}\\
R_0+R_2 &\leq \min\{I(X;Y_2), H(Y_3)+I(X;Y_2|Y_3)\}.
\end{aligned}$$

The first term in each min is larger because $Y_3 -\!\circ\!- X -\!\circ\!- (Y_1,Y_2)$, and so we have $R_0 \leq H(Y_3)$, $R_0+R_1 \leq I(X;Y_1)$ and $R_0+R_2 \leq I(X;Y_2)$. The converse is obvious.

B. Proof of Theorem 3

We only present the essential parts of the proof. Standard arguments finalize the converses. In what follows, inequalities (a) are justified by Fano's inequality, equalities (b) by Csiszár's sum identity, and inequalities (c) by [6, Lemma 1].

(i) $Y_1 \succeq Y_2 \succeq Y_3$.

Converse: For any achievable rate tuple we have

$$\begin{aligned}
nR_0-n\epsilon_n &\overset{(a)}{\leq} I(M_0;Y_3^n) \leq \sum_{i=1}^n I(M_0,Y_3^{i-1};Y_{3,i})\\
&= \sum_{i=1}^n I(M_0;Y_{3,i})+I(Y_3^{i-1};Y_{3,i}|M_0)\\
&\overset{(c)}{\leq} \sum_{i=1}^n I(M_0;Y_{3,i})+I(Y_2^{i-1};Y_{3,i}|M_0)\\
&= \sum_{i=1}^n I(M_0,Y_2^{i-1};Y_{3,i}) = \sum_{i=1}^n I(U_i;Y_{3,i})
\end{aligned}$$

with $U_i=(M_0,Y_2^{i-1})$. In (c) we can apply [6, Lemma 1] because $Y_2 \succeq Y_3$ and $(Y_2^n,Y_3^n) -\!\circ\!- X^n -\!\circ\!- M_0$.

We also have

$$\begin{aligned}
nR_2-n\epsilon_n &\overset{(a)}{\leq} I(M_2;Y_2^n|M_0,M_{1c})\\
&\leq \sum_{i=1}^n I(M_0,M_{1,\mathrm{c}},M_2,Y_2^{i-1};Y_{2,i}|M_0,Y_2^{i-1})\\
&= \sum_{i=1}^n I(V_i;Y_{2,i}|U_i) \qquad (16)
\end{aligned}$$

with $V_i=(M_0,M_{1,\mathrm{c}},M_2,Y_2^{i-1}) \equiv (U_i,M_{1,\mathrm{c}},M_2)$ which satisfies $U_i -\!\circ\!- V_i -\!\circ\!- X_i -\!\circ\!- (Y_{1,i},Y_{2,i},Y_{3,i})$.

Next,

$$\begin{aligned}
nR_{1,\mathrm{p}}-n\epsilon_n &\overset{(a)}{\leq} I(M_{1,\mathrm{p}};Y_1^n|M_0,M_{1,\mathrm{c}},M_2)\\
&= \sum_{i=1}^n I(X_i;Y_{1,i}|M_0,M_{1,\mathrm{c}},M_2,Y_1^{i-1})\\
&= \sum_{i=1}^n I(X_i;Y_{1,i}|M_0,M_{1,\mathrm{c}},M_2)\\
&\qquad -I(Y_1^{i-1};Y_{1,i}|M_0,M_{1,\mathrm{c}},M_2)\\
&\overset{(c)}{\leq} \sum_{i=1}^n I(X_i;Y_{1,i}|M_0,M_{1,\mathrm{c}},M_2)\\
&\qquad -I(Y_2^{i-1};Y_{1,i}|M_0,M_{1,\mathrm{c}},M_2)\\
&= \sum_{i=1}^n I(X_i;Y_{1,i}|V_i) \qquad (17)
\end{aligned}$$

where in (c) we can apply [6, Lemma 1] because $Y_1 \succeq Y_2$ and $(Y_1^n,Y_2^n) -\!\circ\!- X^n -\!\circ\!- (M_0,M_{1,\mathrm{c}},M_2)$. Finally,

$$\begin{aligned}
&n(R_1+R_{2,\mathrm{p}})-n\epsilon_n\\
&\overset{(a)}{\leq} I(M_1;Y_1^n|M_0,M_{2,\mathrm{c}})+I(M_{2,\mathrm{p}};Y_2^n|M_0,M_1,M_{2,\mathrm{c}})\\
&\leq \sum_{i=1}^n I(M_1,Y_{2,i+1}^n;Y_{1,i}|M_0,M_{2,\mathrm{c}},Y_1^{i-1})\\
&\qquad -I(Y_{2,i+1}^n;Y_{1,i}|M_0,M_1,M_{2,\mathrm{c}},Y_1^{i-1})\\
&\qquad +I(M_{2,\mathrm{p}};Y_{2,i}|M_0,M_1,M_{2,\mathrm{c}},Y_{2,i+1}^n,Y_1^{i-1})\\
&\qquad +I(Y_1^{i-1};Y_{2,i}|M_0,M_1,M_{2,\mathrm{c}},Y_{2,i+1}^n)\\
&\overset{(b)}{=} \sum_{i=1}^n I(M_1,M_{2,\mathrm{c}},Y_{2,i+1}^n;Y_{1,i}|M_0,Y_1^{i-1})\\
&\qquad +I(X_i;Y_{2,i}|M_0,M_1,M_{2,\mathrm{c}},Y_{2,i+1}^n,Y_1^{i-1})\\
&\overset{(d)}{\leq} \sum_{i=1}^n I(M_1,M_{2,\mathrm{c}},Y_{2,i+1}^n;Y_{1,i}|M_0,Y_1^{i-1})
\end{aligned}$$

$$
\begin{aligned}
&+I(X_i;Y_{1,i}|M_0,M_1,M_{2,\mathrm{c}},Y_{2,i+1}^n,Y_1^{i-1})\\
&=\sum_{i=1}^n I(X_i;Y_{1,i}|M_0,Y_1^{i-1})\\
&=\sum_{i=1}^n I(X_i;Y_{1,i}|M_0)-I(Y_{1,i};Y_1^{i-1}|M_0)\\
&\overset{(c)}{\leq}\sum_{i=1}^n I(X_i;Y_{1,i}|M_0)-I(Y_{1,i};Y_2^{i-1}|M_0)\\
&=\sum_{i=1}^n I(X_i;Y_{1,i}|Y_2^{i-1},M_0)=\sum_{i=1}^n I(X_i;Y_{1,i}|U_i) \qquad (18)
\end{aligned}
$$

where in (c) and (d) we used that $Y_1 \succeq Y_2$ and the Markov chains $(Y_{1,i},Y_{2,i}) -\!\!\circ\!\!- X_i -\!\!\circ\!\!- (M_0,M_1,M_2,Y_1^{i-1},Y_{2,i+1}^n)$ and $(Y_1^n,Y_2^n) -\!\!\circ\!\!- X^n -\!\!\circ\!\!- M_0$.

Direct part: Follows from (7) and from

$$
\begin{aligned}
I(V;Y_2)&\geq I(U;Y_3)+I(V;Y_2|U)\\
I(X;Y_1)&\geq I(U;Y_3)+I(X;Y_1|U),
\end{aligned}
$$

where the latter inequalities hold because $Y_1 \succeq Y_2 \succeq Y_3$ implies $I(U;Y_3) \leq \min\{I(U;Y_1),I(U;Y_2)\}$ for any $U -\!\!\circ\!\!- V -\!\!\circ\!\!- X -\!\!\circ\!\!- (Y_1,Y_2,Y_3)$.

(ii) $Y_1 \succeq Y_3 \succeq Y_2$.

Converse: For any achievable rate tuple we have

$$
\begin{aligned}
nR_0-n\epsilon_n &\overset{(a)}{\leq} I(M_0;Y_2^n)\\
&\leq \sum_{i=1}^n I(M_0,Y_2^{i-1};Y_{2,i})=\sum_{i=1}^n I(U_i;Y_{2,i})
\end{aligned}
$$

with $U_i=(M_0,Y_2^{i-1})$. In the same way as before, we can prove bounds (16), (17), and (18) with $V_i=(U_i,M_{1,\mathrm{c}},M_2)$, which satisfies $U_i -\!\!\circ\!\!- V_i -\!\!\circ\!\!- X_i -\!\!\circ\!\!- (Y_{1,i},Y_{2,i},Y_{3,i})$.

Direct part: Follows from (7) and from

$$
\begin{aligned}
I(V;Y_2)&\leq I(U;Y_3)+I(V;Y_2|U)\\
I(X;Y_1)&\geq I(U;Y_3)+I(X;Y_1|U),
\end{aligned}
$$

which hold because $Y_1 \succeq Y_3 \succeq Y_2$ implies $I(U;Y_1) \geq I(U;Y_3) \geq I(U;Y_2)$ for any $U -\!\!\circ\!\!- V -\!\!\circ\!\!- X -\!\!\circ\!\!- (Y_1,Y_2,Y_3)$.

(iii) $Y_3 \succeq Y_1 \succeq Y_2$.

Converse: For any achievable rate tuple we have

$$
\begin{aligned}
&n(R_0+R_2)-n\epsilon_n\\
&\overset{(a)}{\leq} I(M_0,M_2;Y_2^n|M_{1,\mathrm{c}})\\
&\leq \sum_{i=1}^n I(M_0,M_{1,\mathrm{c}},M_2,Y_2^{i-1};Y_{2,i})\\
&\overset{(e)}{\leq} \sum_{i=1}^n I(M_0,M_{1,\mathrm{c}},M_2,Y_2^{i-1};Y_{1,i})\\
&=\sum_{i=1}^n I(M_0,M_{1,\mathrm{c}},M_2;Y_{1,i})\\
&\qquad +I(Y_2^{i-1};Y_{1,i}|M_0,M_{1,\mathrm{c}},M_2;)\\
&\overset{(c)}{\leq} \sum_{i=1}^n I(M_0,M_{1,\mathrm{c}},M_2,Y_1^{i-1};Y_{1,i})=\sum_{i=1}^n I(U_i;Y_{1,i})
\end{aligned}
$$

with $U_i=(M_0,M_{1,\mathrm{c}},M_2,Y_1^{i-1})$. Here, (e) and (c) use that $Y_1 \succeq Y_2$ and that $(Y_{1,i},Y_{2,i}) -\!\!\circ\!\!- X_i -\!\!\circ\!\!- (M_0,M_{1,\mathrm{c}},M_2,Y_2^{i-1})$ and $(Y_1^n,Y_2^n) -\!\!\circ\!\!- X^n -\!\!\circ\!\!- (M_0,M_{1,\mathrm{c}},M_2)$. Moreover,

$$
\begin{aligned}
nR_{1,\mathrm{p}}-n\epsilon_n &\overset{(a)}{\leq} I(M_{1,\mathrm{p}};Y_1^n|M_0,M_{1,\mathrm{c}},M_2)\\
&=\sum_{i=1}^n I(M_{1,p};Y_{1,i}|M_0,M_{1,\mathrm{c}},M_2,Y_1^{i-1})\\
&=\sum_{i=1}^n I(X_i;Y_{1,i}|U_i),
\end{aligned}
$$

and

$$
\begin{aligned}
&n(R_0+R_1+R_{2,\mathrm{p}})-n\epsilon_n\\
&\overset{(a)}{\leq} I(M_0,M_1,M_{2,\mathrm{c}};Y_1^n)+I(M_{2,\mathrm{p}};Y_2^n|M_0,M_1,M_{2,\mathrm{c}})\\
&\leq \sum_{i=1}^n I(M_0,M_1,M_{2,\mathrm{c}},Y_1^{i-1},Y_{2,i+1}^n;Y_{1,i})\\
&\qquad -I(Y_{2,i+1}^n;Y_{1,i}|M_0,M_1,M_{2,\mathrm{c}},Y_1^{i-1})\\
&\qquad +I(M_{2,\mathrm{p}};Y_{2,i}|M_0,M_1,M_{2\mathrm{c}},Y_{2,i+1}^n,Y_1^{i-1})\\
&\qquad +I(Y_1^{i-1};Y_{2,i}|M_0,M_1,M_{2,\mathrm{c}},Y_{2,i+1}^n)\\
&\overset{(b)}{=} \sum_{i=1}^n I(M_0,M_1,M_{2,\mathrm{c}},Y_1^{i-1},Y_{2,i+1}^n;Y_{1,i})\\
&\qquad +I(X_i;Y_{2,i}|M_0,M_1,M_{2,\mathrm{c}},Y_1^{i-1},Y_{2,i+1}^n)\\
&\overset{(f)}{\leq} \sum_{i=1}^n I(M_0,M_1,M_{2,\mathrm{c}},Y_1^{i-1},Y_{2,i+1}^n;Y_{1,i})\\
&\qquad +I(X_i;Y_{1,i}|M_0,M_1,M_{2,\mathrm{c}},Y_1^{i-1},Y_{2,i+1}^n)\\
&=\sum_{i=1}^n I(X_i,Y_{1,i})
\end{aligned}
$$

where in (f) we used $Y_1 \succeq Y_2$ and the Markov chain $(Y_{1,i},Y_{2,i}) -\!\!\circ\!\!- X_i -\!\!\circ\!\!- (M_0,M_1,M_{2,\mathrm{c}},Y_1^{i-1},Y_{2,i+1}^n)$.

Direct part: Follows from (7) by setting $V=U$ and from

$$
\begin{aligned}
I(U;Y_2)&\leq I(U;Y_3)+I(U;Y_2|U)=I(U;Y_3)\\
I(X;Y_1)&\leq I(U;Y_3)+I(X;Y_1|U),
\end{aligned}
$$

which hold because $Y_3 \succeq Y_1 \succeq Y_2$ implies $I(U;Y_3) \geq \max\{I(U;Y_1),I(U;Y_2)\}$ for any $U -\!\!\circ\!\!- X -\!\!\circ\!\!- (Y_1,Y_2,Y_3)$.

References

[1] T. J. Oechtering, C. Schnurr, I. Bjelaković, and H. Boche, "Broadcast capacity region of two-phase bidirectional relaying," *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 454–458, 2008.

[2] E. Tuncel, "Slepian-Wolf coding over broadcast channels," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1469–1482, 2006.

[3] J. Nayak, E. Tuncel, and D. Gündüz, "Wyner-Ziv coding over broadcast channels: Digital schemes," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1782–1799, 2010.

[4] G. Kramer and S. Shamai, "Capacity for classes of broadcast channels with receiver side information," in *Proceedings IEEE Information Theory Workshop*, Lake Tahoe, California, 2007.

[5] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.

[6] C. Nair and Z. V. Wang, "The capacity region of the three receiver less noisy broadcast channel," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4058–4062, 2011.