entropy

Article

Cognition and Cooperation in Interfered Multiple Access Channels †

Jonathan Shimonovich [1,3], Anelia Somekh-Baruch [2,*] and Shlomo Shamai (Shitz) [3]

1 Check Point Software Technologies Ltd., Tel Aviv 67897, Israel; sonatans@gmail.com
2 Faculty of Engineering, Bar-Ilan University, Ramat-Gan 52900, Israel
3 Electrical Engineering Department, Technion, Haifa 32000, Israel; sshlomo@ee.technion.ac.il
* Correspondence: somekha@biu.ac.il; Tel.: +972-3-5317733
† This paper is an extened version of our paper published in 2013 IEEE Information Theory Workshop (ITW), Sevilla, Spain, 9–13 September 2013 and 2012 IEEE 27th Convention of Electrical and Electronics Engineers, Eilat, Israel, 14–17 November 2012.

Received: 31 May 2017; Accepted: 20 July 2017; Published: 24 July 2017

Abstract: In this work, we investigate a three-user cognitive communication network where a primary two-user multiple access channel suffers interference from a secondary point-to-point channel, sharing the same medium. While the point-to-point channel transmitter—transmitter 3—causes an interference at the primary multiple access channel receiver, we assume that the primary channel transmitters—transmitters 1 and 2—do not cause any interference at the point-to-point receiver. It is assumed that one of the multiple access channel transmitters has cognitive capabilities and cribs causally from the other multiple access channel transmitter. Furthermore, we assume that the cognitive transmitter knows the message of transmitter 3 in a non-causal manner, thus introducing the three-user multiple access cognitive Z-interference channel. We obtain inner and outer bounds on the capacity region of the this channel for both causal and strictly causal cribbing cognitive encoders. We further investigate different variations and aspects of the channel, referring to some previously studied cases. Attempting to better characterize the capacity region we look at the vertex points of the capacity region where each one of the transmitters tries to achieve its maximal rate. Moreover, we find the capacity region of a special case of a certain kind of *more-capable* multiple access cognitive Z-interference channels. In addition, we study the case of full unidirectional cooperation between the 2 multiple access channel encoders. Finally, since direct cribbing allows us full cognition in the case of continuous input alphabets, we study the case of partial cribbing, i.e., when the cribbing is performed via a deterministic function.

Keywords: cognitive radio; multiple access channel; interference channel; capacity region; cognition; cribbing; cooperative communication

1. Introduction

Two of the most fundamental multi-terminal communication channels are the Multiple-Access Channel (MAC) and the Interference Channel (IFC). The MAC, sometimes referred to as the uplink channel, consists of multiple transmitters, sending messages to a single receiver (base station). The capacity region of the two-user MAC channel was determined, early on, by Ahlswede [1], and Liao [2]. However, the capacity regions of many other fundamental multi-terminal channels are yet unknown. One of these channels is the Interference Channel (IFC). The two-user IFC consists of two point to point transmitter-receiver pairs, where each of the transmitters has its own intended receiver and serves as an interference to the other transmitter-receiver. The study of this channel was initiated by C.E. Shannon [3], and extended by R. Ahlswede [4] who gave simple but fundamental inner and outer bounds to the capacity region. The fundamental achievable region of the discrete memoryless two-user IC is the Han-Kobayashi (HK) region [5] which can be exressed by a simplified expression [6]. Much progress has been made toward understanding this channel (see, e.g., [7–13] and the references therein). Although widely investigated, this problem remains unsolved except for some specific channel configurations, enforcing various constraints on the channel [14].

A common scenario of multi-terminal network is comprised of these two channels. For instance, looking at Wi-Fi or cellular communication, there are usually several portable devices (i.e., laptops, mobile phones, etc.) "talking" to a single end point (i.e., base station, Access Point, etc.). Moreover, the same frequencies are frequently used by nearby base stations, causing interferences at adjacent receivers. This increasing usage of wireless services and constant reuse of frequencies imply an ever increasing problem of optimizing the wireless medium for achieving better transmission rates. *Cognitive radio* technology is one of the novel strategies for overcoming the problem of inefficient spectrum usage which has been receiving a lot of attention [15–17].

Cognition stands for awareness of system paramters, such as operative frequencies, time schedules, space directivity, and actual transmission. The latter refers to transmitted messages of interfering transmitters, which are either monitored by receiving the interfering signals (cribbing), or on a network scale (a-priori available transmitted messages). Examples of signal awareness are reflected by Dynamic Spectrum Access (DSA) (see the tutorial [18], and references therein), as well as a variety of techniques for spectrum and activity sensing (see [19] and references therein). The timely relevance of cognitive radios and the information theoretic framework that can assess the potential benefits and limitations are reflected in recent literature (see [15] and references therein).

In our study, we focus on aspects of cognition in terms of the ability to recognize the primary (licensed) user and adapt its communication strategy to minimize the interference that it generates, while maximizing its own Quality of Service (QoS). Furthermore, cognition allows *cooperation* between transmitters in relaying information to improve network capacity. The shared information used by the cognitive transmitter might be achieved through a noisy observation of the channel or via a dedicated link. The cognitive transmitter may apply different strategies such as decode-and-forward (DF) or amplify-and-forward (AF) for relaying the other transmitter information.

To obtain information theoretical limits of cognitive radios, the Cognitive Interference Channel (CIFC) is defined in [20]. CIFC refers to a two-user Interference Channel (IFC) in which the cognitive user (secondary user) is cognizant of the message being transmitted by the other user (primary user), either in a non-causal or causal manner. The two-user CIFC was further studied in [21–25]. Cognitive radio was applied to the MAC in 1985, when Willems and Van Der Meulen established the capacity region of the MAC with cribbing encoders [26]. Cribbing encoders means that one or both encoders crib from the other encoder and learn the channel input(s) (to be) emitted by this encoder in a causal manner. Since then, the cognitive MAC has received much attention, recently characterizing capacity regions for various extensions [27–32]. Today, there are already practical implications of advanced processing techniques in the cognitive arena. For example, [33], shows coding techniques for an Orthogonal Frequency-Division Multiple Access (OFDMA)-based secondary service in cognitive networks that outperform traditional coding schemes, see also [34]. Hence, aspects of binning (dirty-paper coding [35]), as well as rate splitting [5], used in cognitive coding schemes, do have even stronger practical implications.

In this paper, we study a common wireless scenario in which a Multiple Access Channel (MAC) suffers interferences from a point-to-point (P2P) channel sharing the same medium. The main motivation behind this model is trying to interweave a MAC channel on top of a licensed P2P channel. The P2P licensed user must not suffer interference while the other users may use cognitive radio to improve performance. Adding cognition capabilities to one of the MAC transmitters, we investigate the case in which it has knowledge of signals transmitted by another user intended for the same receiver as well as signals transmitted by the P2P user on a separate channel resulting in an interference at the MAC receiver. We introduce Multiple Access Cognitive Z-Interference Channel (MA-CZIC) which consists of three transmitters and two receivers; two-user MAC as a primary network and a point-to-point channel as a secondary channel. The communication system, including the primary and secondary channels (whose outputs are Y and Z, respectively), is depicted in Figure 1. The signal X_1 is generated by Encoder 1. Encoder 2 is assumed to be a cognitive cribbing encoder, that is, it has *causal* knowledge of Encoder 1's signal, as well as *non-causal* knowledge of Encoder 3's signal. We note that while the signal X_3 interferes with the other signals creating Y, it is observed interference-free by the second decoder creating Z. The cognition of the P2P signal may model the fact that the same user produced a P2P message to another point, and hence naturally it is cognizant of the message W_3. This channel model generalizes several previously studied setups: without Encoder 3, the system reduces to a MAC with a cribbing encoder as in [26]. Replacing the signal X_3 with a state process and ignoring the structure of X_3, we get a MAC with states available at a cribbing encoder as in [24,27]. Removing Encoder 2, the problem reduces to the standard Z-Interference channel, and removing Encoder 1, we get the Cognitive Z-Interference channel, as in [36]. The Z-Gaussian Cognitive Interference channel was further studied in [37]. The model of a cooperative state-dependent MAC which is considered in [29] is very closely related to a special case of the MA-CZIC which is obtained by replacing the interfering signal X_k of the MA-CZIC with an i.i.d. state sequence S_k which is known non-causally to the cognitive transmitter. Some of the results which appear in this paper were presented in part in [38,39].

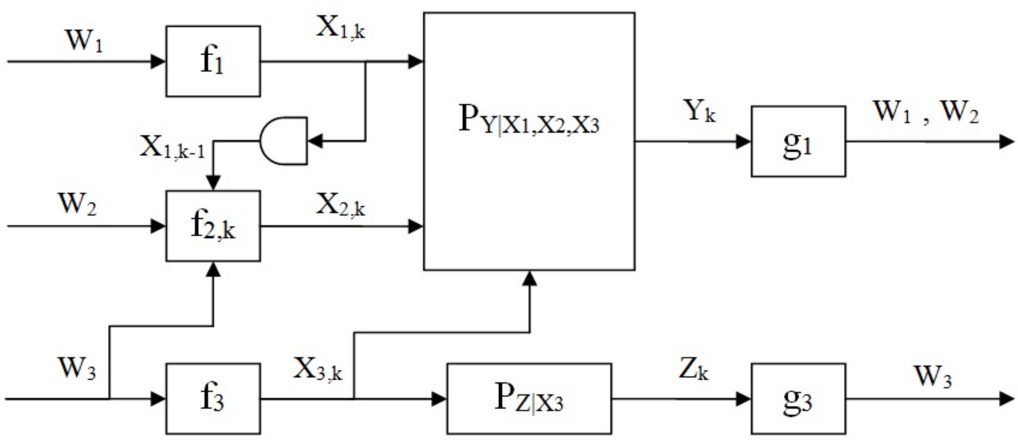

Figure 1. Multiple-Access Cognitive Z-Interference Channel (MA-CZIC).

The rest of the paper is organized as follows. In Section 2 we formally define the memoryless MA-CZIC with *causal* and *strictly causal* cribbing encoder. In Section 3 we proceed to derive inner and outer bounds on the capacity region of the channel with *causal* and *strictly causal* cribbing encoders including a special case of the channel where the bounds coincide and the capacity region is established. Section 4 is devoted to the case of full unidirectional cooperation from Encoder 1 to Encoder 2 (a common message setup). Section 5 deals with the case of partial cribbing. Finally, concluding remarks are given in Section 6.

2. Channel Model and Preliminaries

Throughout this work, we will use uppercase letters (e.g., X) to denote random variables (RVs) and lowercase letters (e.g., x) to show their realization. Boldface letters are used for denoting n-vectors, e.g., $\mathbf{x} = x^n = (x_1, ..., x_n)$. For a set of RVs $S = \{X_1, ..., X_k\}$, $A_\epsilon^n(S)$ denotes the set of ϵ-strongly, jointly typical n-sequences of S as defined in ([40], Chapter 13). We may omit the index n from $A_\epsilon^n(S)$ when it is clear from the context.

A more formal definition of the problem is as follows: A discrete memoryless multiple-access Z-interference channel (MA-CZIC) is defined by the input alphabets $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3)$ and output alphabets

$(\mathcal{Y}, \mathcal{Z})$ and by the transition probabilities $P_{Y|X_1,X_2,X_3}$ and $P_{Z|X_3}$, that is, the channel outputs are generated in the following manner:

$$\Pr\left(y^n, z^n | x_1^n, x_2^n, x_3^n\right) = \prod_{t=1}^{n} p(y_t | x_{1,t}, x_{2,t}, x_{3,t}) p(z_t | x_{3,t}). \tag{1}$$

Encoder i, $i \in \{1, 2, 3\}$, sends a message W_i which is drawn uniformly over the set $M_i \triangleq \{1, \ldots, 2^{nR_i}\}$ to its destined receiver. It is further assumed that Encoder 2 "cribs" causally and observes the sequence of channel inputs emitted by Encoder 1 during all past transmissions before generating its next channel input. The model is depicted in Figure 1.

An $(2^{nR_1}, 2^{nR_2}, 2^{nR_3}, n)$ code for the MA-CZIC with *strictly causal* Encoder 2 consists of:

1. Encoder 1 defined by a deterministic mapping
$$f_1 : M_1 \to \mathcal{X}_1^n \tag{2}$$
which maps the message W_1 to a channel input codeword.
2. Encoder 2 which observes X_1^{i-1} and W_3 prior to transmitting $X_{2,i}$, is defined by the mappings
$$f_{2,k}^{(sc)} : M_2 \times M_3 \times \mathcal{X}_1^{k-1} \to \mathcal{X}_2 \qquad k = 1, \ldots, n. \tag{3}$$
3. Encoder 3 is defined by a deterministic mapping
$$f_3 : M_3 \to \mathcal{X}_3^n. \tag{4}$$
4. The primary (main) decoder is defined by a mapping
$$g_1 : \mathcal{Y}^n \to M_1 \times M_2. \tag{5}$$
5. The secondary decoder is defined by a mapping
$$g_3 : \mathcal{Z}^n \to M_3. \tag{6}$$

An $(2^{nR_1}, 2^{nR_2}, 2^{nR_3}, n)$ code for the MA-CZIC with *causal* Encoder 2 differs only in the fact that Encoder 2 observes X_1^i (including the current symbol, $X_{1,i}$) before transmitting $X_{2,i}$, and is defined by a the mappings
$$f_{2,k}^{(c)} : M_2 \times M_3 \times \mathcal{X}_1^{k} \to \mathcal{X}_2 \qquad k = 1, 2, \ldots, n. \tag{7}$$

For a given code, the block average error probability is

$$P_e^{(n)} = \frac{1}{2^{n(R_1+R_2+R_3)}} \sum_{w_1=1}^{2^{nR_1}} \sum_{w_2=1}^{2^{nR_2}} \sum_{w_3=1}^{2^{nR_3}} \Pr\{g_1(Y^n) \neq (w_1, w_2) \cup g_3(Z^n) \neq w_3 | W_i = w_i,\ i = 1, 2, 3\}. \tag{8}$$

A rate-triple (R_1, R_2, R_3) is said to be achievable for the MA-CZIC if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, 2^{nR_3}, n)$ codes with $\lim_{n\to\infty} P_e^{(n)} = 0$. The capacity region of the MA-CZIC with a cribbing encoder is the closure of the set of achievable rate-triples.

3. Main Results

In this section, we provide inner and outer bounds to the capacity region of the discrete memoryless MA-CZIC.

3.1. Inner Bound

We next present achievable regions for the strictly causal and the causal MA-CZICs.

Definition 1. *Let $\mathcal{R}_{sc}$ be the region defined by the closure of the convex hull of the set of all rate-triples (R_1, R_2, R_3) satisfying*

$$R_1 \leq H(X_1|V) \tag{9a}$$
$$R_2 \leq I(U;Y|VLX_1) - I(U;X_3|VL) \tag{9b}$$
$$R_1 + R_2 \leq I(VUX_1;Y|L) - I(U;X_3|VL) \tag{9c}$$
$$R_3 \leq I(X_3;Z|L) + \min\{I(L;Y), I(L;Z)\} \tag{9d}$$

for some probability distribution of the form

$$P_{VLUX_1X_2X_3} = P_V P_L P_{X_3|L} P_{X_1|V} P_{UX_2|VLX_3}. \tag{10}$$

Theorem 1. *The region $\mathcal{R}_{sc}$ is achievable for the MA-CZIC with a strictly causal cribbing encoder.*

The proof appears in Appendix A.

Definition 2. *Let $\mathcal{R}_c$ be the region defined by the closure of the convex hull of the set of all rate-triples (R_1, R_2, R_3) satisfying* (9a)–(9d)*, for some probability distribution of the form*

$$P_{VLUX_1X_2X_3} = P_V P_L P_{X_3|L} P_{X_1|V} P_{U|VLX_3} P_{X_2|UVLX_3X_1}. \tag{11}$$

Theorem 2. *The region $\mathcal{R}_c$ is achievable for the MA-CZIC with a causal cribbing encoder.*

The outline of the proof appears in Appendix B.

A few comments regarding the achievability region (9a)–(9d) are in order. In the coding scheme, Encoder 1 and Encoder 2 use Block–Markov superposition encoding, while the primary decoder uses backward decoding [27]. In this scheme, the RV V represents the "resolution information" [26]; i.e., the current block information used for encoding the proceeding block. Encoder 3 uses rate-splitting, where the RV L represents the part of W_3 that can be decoded by both the primary and secondary decoders as can be observed by the term $\min\{I(L;Y), I(L;Z)\}$ which appears in (9d). The complementary part of W_3, while fully decoded by the secondary decoder, serves as a channel state for the primary channel in the form of X_3. To reduce interference the cognitive encoder (Encoder 2) additionally uses Gel'fand–Pinsker binning [41] of U against X_3, assuming an already successful decoding of V and L at the primary decoder, as can be seen in (9b).

It is important to note that the achievable region $\mathcal{R}_{sc}$ is consistent with previously studied special cases: By setting $R_1 = 0$ and $X_1 = 0$ we can also set $V = \emptyset$. The equations then reduce to

$$R_2 \leq I(U;Y|L) - I(U;X_3|L) \tag{12}$$
$$R_3 \leq I(X_3;Z|L) + \min\{I(L;Y), I(L;Z)\}, \tag{13}$$

and this results in the region achievable for the cognitive Z-Interference channel, studied in [36], with user 2 and user 3 as the cognitive and non-cognitive users, respectively.

Removing Encoder 2 by setting $R_2 = 0$ and $X_2 = 0$ we can also set $U = \emptyset$ and we get the classical Z-Interference channel with the 2 users, Encoder 1 and Encoder 3. In this case Y is dependent of V only through X_1, since $V \to X_1 \to Y$, and $I(VUX_1;Y|L) = I(X_1;Y|L)$. We get

$$R_1 \le I(X_1;Y|L) \tag{14}$$
$$R_3 \le I(X_3;Z|L) + \min\{I(L;Y), I(L;Z)\}. \tag{15}$$

An interesting setup arises by setting $W_2 = 0$ without removing Encoder 2. This models a relay channel, where Encoder 2 is a relay which has no message of its own and learns the information of the transmitter by cribbing (modeling excellent SNR conditions on this path). This model relates to [42], if the structure of the primary user (X_3) is not accounted for, thus assuming i.i.d. state symbols known a-causally at the relay, as in [42].

Removing Encoder 3 by setting $R_3 = 0$ and $X_3 = 0$ we can also set $L = \emptyset$, and the expression reduces to

$$R_1 \le H(X_1|V) \tag{16}$$
$$R_2 \le I(U;Y|VX_1) \tag{17}$$
$$R_1 + R_2 \le I(VUX_1;Y). \tag{18}$$

By setting $U = X_2$ we get the achievable region of the MAC with Encoder 2 as the cribbing encoder [26].

By setting $L = 0$, removing inequality (9d) and replacing X_3 by S, whose given probability distribution is not to be optimized, the region reduces to the one in [27].

It is worth noting that in the case of a Gaussian channel, Encoder 2 can become fully cognitive of the message W_1 from a single sample of X_1. This special case can be made non-trivial by adding a noisy channel or some deterministic function (quantizer for instance) between X_1 and Encoder 2.

Finally, we examine the case where Encoder 1's output may be viewed as two parts $X_1 = (X_{1a}, X_{1b})$ where only the first part of the input affects the channel; i.e., $P_{Y|X_1X_2X_3} = P_{Y|X_{1a}X_2X_3}$. In this case, if the second part X_{1b} is rich enough (e.g. continuous alphabet) Encoder 1 is able to transfer to Encoder 2 infinite amount of data, specifically the entire message W_1. This is equivalent to the case of full cooperation from Encoder 1 to Encoder 2; i.e., the case where Encoder 2 has full knowledge of Encoder 1's data W_1. Hence, the cooperative state-dependent MAC where the state is known non-causally at the cognitive encoder [29] may also be considered as a special case of the MA-CZIC, when X_3 is replaced with an i.i.d. state S.

3.2. Outer Bound

In this section, we present an outer bound on the achievable region of the strictly causal and causal MA-CZIC.

Theorem 3. *Achievable rate-triples (R_1, R_2, R_3) for the MA-CZIC with a strictly causal cribbing encoder belong to the closure of the convex hull of all rate triples that satisfy*

$$R_1 \le H(X_1|V) \tag{19a}$$
$$R_2 \le I(U;Y|VLX_1) - I(U;Z|VL) \tag{19b}$$
$$R_1 + R_2 \le I(VUX_1;Y|L) - I(VU;Z|L) \tag{19c}$$
$$R_3 \le I(X_3;Z|L) + \min\{I(L;Y), I(L;Z)\} \tag{19d}$$

for some probability distribution of the form

$$P_{VLUX_1X_2X_3} = P_{X_3}P_V P_{L|X_3V} P_{X_1|V} P_{UX_2|VLX_3}. \tag{20}$$

The proof is provided in Appendix C, it is based on Fano's Inequality [40] and from the Csiszár and Körner's identity ([43], Lemma 7).

Theorem 4. *Achievable rate-triples (R_1, R_2, R_3) for the MA-CZIC with a causal cribbing encoder belong to the closure of the convex hull of all rate regions given by (19a)–(19d) for some probability distribution of the form*

$$P_{VLUX_1X_2X_3} = P_{X_3}P_V P_{L|X_3V}P_{X_1|V}P_{UX_2|VLX_3X_1}. \tag{21}$$

The outline of the proof is provided in Appendix D.

As for the alphabet cardinalities: using standard applications of Carathéodory's Theorem we obtain that it is sufficient to consider the alphabet cardinalities which are bounded as follows:

$$\begin{aligned} |\mathcal{L}| &\leq |\mathcal{X}_1||\mathcal{X}_2||\mathcal{X}_3| + 4 && (22)\\ |\mathcal{V}| &\leq |\mathcal{X}_1||\mathcal{X}_2||\mathcal{X}_3||\mathcal{L}| + 3 && (23)\\ |\mathcal{U}| &\leq |\mathcal{X}_1||\mathcal{X}_2||\mathcal{X}_3||\mathcal{L}||\mathcal{V}| + 2. && (24) \end{aligned}$$

The details are omitted for the sake of brevity.

3.3. Special Cases

For the special case of a *more-capable* MA-CZIC channel we can actually establish the capacity region of the channel, both in the causal and the strictly causal cases.

Definition 3. *We say that the strictly-causal MA-CZIC is more-capable if $I(X_3;Y) \geq I(X_3;Z)$ for all probability distributions of the form $P_V P_{X_1|V}P_{X_3}\ P_{X_2|VX_3}P_{Y|X_1X_2X_3}P_{Z|X_3}$.*

Theorem 5. *The capacity region of the more-capable strictly-causal MA-CZIC channel is the closure of the convex hull of the set of all rate-triples (R_1, R_2, R_3) satisfying*

$$\begin{aligned} R_1 &\leq H(X_1|V) && (25a)\\ R_2 &\leq I(X_2;Y|VX_1X_3) && (25b)\\ R_1 + R_2 &\leq I(X_1X_2;Y|X_3) && (25c)\\ R_3 &\leq I(X_3;Z) && (25d) \end{aligned}$$

for some probability distribution of the form

$$P_{VX_1X_2X_3} = P_V P_{X_1|V}P_{X_3}P_{X_2|VX_3}. \tag{26}$$

The proof of Theorem 5 is provided in Appendix E.

Theorem 6. *The capacity region of the more-capable causal MA-CZIC channel is the closure of the convex hull of the set of all rate-triples (R_1, R_2, R_3) satisfying (25a)–(25d), for some probability distribution of the form*

$$P_{VX_1X_2X_3} = P_V P_{X_1|V}P_{X_3}P_{X_2|VX_1X_3}. \tag{27}$$

The proof of Theorem 6; i.e., the *causal* case, follows in the same manner as that of Theorem 5 and thus omitted.

Unfortunately, the requirement that the MA-CZIC is *more-capable* implies that the receiver Y has a better reception of the signal X_3 than its designated receiver Z, which is somewhat optimistic.

We next consider the cases where either one of the transmitters wishes to achieve its maximal possible rate; i.e., the vertex point of the capacity region.

- **Maximal rate at Transmitter 1**:

Transmitter 2 may help Transmitter 1's transmission and by doing so increase its rate. Therefore, we that assume Transmitter 2 dedicates its transmission to help transmitting W_1. Transmitter 3 should minimize its interference at the Y Receiver. Setting $L = X_3$, the entire interference caused by transmitter 3 at receiver 1 may be reduced via successive cancellation decoding. With no interference caused by transmitter 3, transmitter 2 may drop the Gelfand–Pinsker scheme, setting $U = X_2$ to maximize the rates. Thus, from (9a)–(9d) we get

$$\begin{aligned} R_1 &\le \min\{H(X_1|V), I(VX_1X_2;Y|X_3)\} && \text{(28a)}\\ R_2 &\le \min\{I(X_2;Y|VX_1X_3),\\ &\qquad I(VX_1X_2;Y|X_3) - R_1\} && \text{(28b)}\\ R_3 &\le \min\{I(X_3;Y), I(X_3;Z)\}. && \text{(28c)} \end{aligned}$$

From the Markov chain $V - X_1X_2X_3 - Y$ we get $I(VX_1X_2;Y|X_3) = I(X_1X_2;Y|X_3)$. Therefore, we can rewrite (28a)–(28c) as

$$\begin{aligned} R_1 &\le \min\{H(X_1|V), I(X_1X_2;Y|X_3)\} && \text{(29a)}\\ R_2 &\le \min\{I(X_2;Y|VX_1X_3),\\ &\qquad I(X_1X_2;Y|X_3) - R_1\} && \text{(29b)}\\ R_3 &\le \min\{I(X_3;Y), I(X_3;Z)\}. && \text{(29c)} \end{aligned}$$

where in the strictly-causal MA-CZIC case, the union is over all probability distributions of the form

$$P_{VX_1X_2X_3} = P_V P_{X_3} P_{X_1|V} P_{X_2|V}. \tag{30}$$

- **Maximal rate at Transmitter 2**:

Both Transmitter 1 and 3 are not cognitive and have no knowledge of the message W_2, thus they cannot help convey W_2 to Y and should only reduce their interference to a minimal level. Setting $L = X_3$ and $U = X_2$ follows as in maximizing R_1 in (28a)–(29c).

Therefore, we get (28a)–(28c) where the maximization is on R_2 instead of R_1, i.e.

$$\begin{aligned} R_1 &\le \min\{H(X_1|V), I(X_1X_2;Y|X_3) - R_2\} && \text{(31a)}\\ R_2 &\le I(X_2;Y|VX_1X_3) && \text{(31b)}\\ R_3 &\le \min\{I(X_3;Y), I(X_3;Z)\}. && \text{(31c)} \end{aligned}$$

- **Maximal rate at Transmitter 3**:

Looking at (9d) and (19d) we see that the lower and upper bounds on R_3 coincide. Since transmitter 3 is not affected by the transmission of both transmitters 1 and 2, we can treat the transmitter 3—receiver 3 pair as a single user channel and thus achieve the Shannon capacity; i.e.,

$$R_3 \le I(X_3;Z). \tag{32}$$

In the general case, the maximum rate at transmitter 3 is achieved by setting $L = 0$. In this case, the higher rate at transmitter 3 comes at the expense of the other transmitters, since L was used for conveying part of the interference X_3 to Y. Thus, (9a)–(9d) become

$$
\begin{aligned}
R_1 &\leq H(X_1|V) && (33a)\\
R_2 &\leq I(U;Y|VX_1) - I(U;X_3|V) && (33b)\\
R_1 + R_2 &\leq I(VUX_1;Y) - I(U;X_3|V) && (33c)\\
R_3 &\leq I(X_3;Z). && (33d)
\end{aligned}
$$

Examining (9d), we see that maximum rate at transmitter 3 may also be achieved without affecting R_1 and R_2. This is true when the receiver Y is *less-noisy* than receiver Z in the sense that $I(L;Y) \geq I(L;Z)$ for all probability distributions of the form (20). In this case (9d) becomes

$$
\begin{aligned}
R_3 &\leq I(X_3;Z|L) + \min\{I(L;Y), I(L;Z)\}\\
&= I(X_3;Z|L) + I(L;Z)\\
&= I(X_3;Z) \qquad (34)
\end{aligned}
$$

Actually, it suffices to require that the channel will be *more-capable*; i.e., $I(X_3;Y) \geq I(X_3;Z)$ for all probability distributions of the form (20), for achieving maximum rate at R_3.

4. Cooperative Encoding

Let us now consider the case of full unidirectional cooperation from Encoder 1 to Encoder 2. This becomes a setup in which Encoders 1 and 2 share a common message, and Encoder 2 transmits a separate additional private message. Thus we have an interference cognitive channel (cognition in terms of W_3) with a common message, as depicted in Figure 2. Hence, Encoder 2 is given by the mapping

$$f_2 : M_1 \times M_2 \times M_3 \to \mathcal{X}_2^n \qquad k = 1, \ldots, n. \qquad (35)$$

For this channel setup, a simpler outer bound on the capacity region can be derived providing some insights on the original problem. A special case of this channel, in which the secondary channel is removed and X_3 is replaced with an i.i.d. state S for the main channel; i.e., $P_{Y|X_1X_2S}$, was studied in [29] and the single-letter characterization of the capacity region was established for that channel.

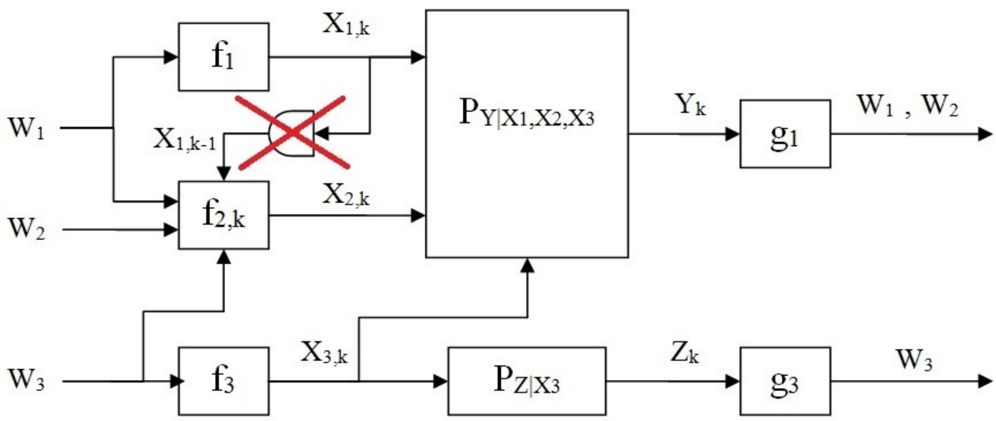

Figure 2. MA-CZIC with full unidirectional cooperation from Encoder 1 to Encoder 2.

The following theorem provides a single-letter expression for an achievable region of the MA-CZIC with full unidirectional cooperation (common message).

Theorem 7. *The closure of the convex hull of the set of all rate-triples (R_1, R_2, R_3) satisfying*

$$R_1 + R_2 \leq I(X_1U; Y|L) - I(U; X_3|LX_1) \tag{36a}$$
$$R_2 \leq I(U; Y|LX_1) - I(U; X_3|LX_1) \tag{36b}$$
$$R_3 \leq I(X_3; Z|L) + \min\{I(L; Y), I(L; Z)\} \tag{36c}$$

for some probability distribution of the form

$$P_{LUX_1X_2X_3} = P_{X_1}P_LP_{X_3|L}P_{UX_2|X_1LX_3} \tag{37}$$

is achievable for the MA-CZIC with with full unidirectional cooperation.

The outline of the proof for Theorem 7 appears in Appendix F.

The following theorem provides a single-letter expression for an outer bound on the capacity region of the MA-CZIC with full unidirectional cooperation.

Theorem 8. *Achievable rate-triples (R_1, R_2, R_3) for the MA-CZIC with full unidirectional cooperation belong to the closure of the convex hull of rate-regions given by*

$$R_1 + R_2 \leq I(VUX_1; Y|L) - I(VU; Z|L) \tag{38a}$$
$$R_2 \leq I(U; Y|LVX_1) - I(U; Z|LV) \tag{38b}$$
$$R_3 \leq I(X_3; Z|L) + \min\{I(L; Y), I(L; Z)\} \tag{38c}$$

for some probability distribution of the form

$$P_{LVUX_1X_2X_3} = P_VP_{X_3}P_{L|VX_3}1_{\{X_1=f(V)\}}P_{UX_2|VLX_3} \tag{39}$$

The outline of the proof of Theorem 8 is provided in Appendix G.

Notice that inequalities (38a)–(38c) are identical to (19b)–(19d), where the probability distribution form (39) is a special case of (20). Thus, the outer bound established for R_2, R_3 and the sum-rate $R_1 + R_2$ in the *strictly-causal* MA-CZIC, also holds for the case of full unidirectional cooperation. However, we would expect the outer bounds on R_2 and the sum-rate $R_1 + R_2$ to be smaller for the channel with the cribbing encoder, thus implying that the outer bound for the MA-CZIC is generally not tight.

5. Partial Cribbing

Next we consider the case of partial cribbing, where Encoder 2 views X_1 through a deterministic function

$$h : \mathcal{X}_1 \to \mathcal{Y}_2 \tag{40}$$

instead of obtaining X_1 directly. This cribbing scheme is motivated by continuous input alphabet MA-CZIC, since perfect cribbing results in the degenerated case of full cooperation between the encoders and requires an infinite capacity link.

We define the *strictly-causal* MA-CZIC with partial cribbing as in (2)–(6) with the exception that Encoder 2 is defined by the mapping

$$f_{2,k}^{(sc)} : M_2 \times M_3 \times \mathcal{Y}_2^{k-1} \to \mathcal{X}_2 \qquad k = 1, \ldots, n \tag{41}$$

where $y_{2,k} = h(x_{1,k})$ for $k = 1, \ldots, n$. The *causal* MA-CZIC with partial cribbing differs by setting

$$f_{2,k}^{(sc)} : M_2 \times M_3 \times \mathcal{Y}_2^{k} \to \mathcal{X}_2 \qquad k = 1, \ldots, n. \tag{42}$$

It is worth noticing that the state-dependent MAC with state information known non-causally at one encoder [44] is a special case of the MA-CZIC with partial cribbing. This case is derived by setting $h(X_1) \equiv 0$ and replacing X_3 with an i.i.d. state S. The capacity region of this simpler case remains an open problem. Therefore, it is hard to expect that capacity region would be established for the MA-CZIC with partial cribbing. Next, we establish inner and outer bounds for the MA-CZIC with partial cribbing.

Theorem 9. *The closure of the convex hull of the set of rate-triples* (R_1, R_2, R_3) *satisfying*

$$R_1 \le H(Y_2|V) + I(X_1;Y|VY_2UL) \tag{43}$$
$$R_2 \le I(U;Y|VLX_1) - I(U;X_3|VL) \tag{44}$$
$$R_1 + R_2 \le I(VUX_1;Y|L) - I(U;X_3|VL) \tag{45}$$
$$R_1 + R_2 \le I(UX_1;Y|VLY_2) + H(Y_2|V) - I(U;X_3|VL) \tag{46}$$
$$R_3 \le I(X_3;Z|L) + \min\{I(L;Y), I(L;Z)\} \tag{47}$$

for some probability distribution of the form

$$P_{VLUX_1Y_2X_2X_3} = P_V P_L P_{X_3|L} P_{X_1Y_2|V} P_{UX_2|VLX_3}. \tag{48}$$

is achievable for the strictly-causal MA-CZIC with partial cribbing.

The outline proof of Theorem 9 appears in Appendix H.

Theorem 10. *The closure of the convex hull of the set of rate-triples* (R_1, R_2, R_3) *satisfying* (43)–(47)*, for some probability distribution of the form*

$$P_{VLUX_1Y_2X_2X_3} = P_V P_L P_{X_3|L} P_{X_1Y_2|V} P_{U|VLX_3} P_{X_2|VY_2ULX_3}. \tag{49}$$

is achievable for the causal MA-CZIC with partial cribbing.

The proof of Theorem 10 is similar to that of Theorem 9 and thus is omitted.

Comparing this result to the achievability region found for the MA-CZIC we can see that inequalities (44), (45) and (47) are identical to (9b)–(9d), while inequality (43) differs and inequality (46) was added. In correspondence, the coding scheme for the MA-CZIC with partial cribbing differs from Theorem 1 mainly in Encoder 1. Encoder 1 now needs to transmit data in a lossy manner to both Y receiver and Encoder 2. To do so, Encoder 1 employs the rate-splitting technique. It splits its message W_1 into two parts (W_{1a}, W_{1b}) with rates R_{1a}, R_{1b} accordingly, such that $R_1 = R_{1a} + R_{1b}$. The rate R_{1a} represents the rate of transmission to Encoder 2. Combining the rate-splitting with the superposition block Markov encoding (SBME) at Encoder 1 results in another codebook $\{y_2^n\}$, in addition to the two codebooks $\{v^n\}$ and $\{x_1^n\}$. The codebooks are created in an i.i.d. manner as follows: First, $2^{nR_{1a}}$ codewords $\{v^n\}$ are created using P_V. Then, for each codeword v^n, $2^{nR_{1a}}$ codewords $\{y_2^n\}$ are drawn i.i.d. $\sim P_{Y_2|V}$ given v^n. Finally, for each pair (v^n, y^n), $2^{R_{1b}}$ codewords $\{x^n\}$ are drawn i.i.d. $\sim P_{X_1|Y_2V}$ given (v^n, y_2^n). Next, as in the scheme which corresponds to Theorem 1 SBME coding scheme, the index of the codeword y_2^n in time i becomes the index of v^n in time $i+1$. For successful decoding at Encoder 2, we must require $R_{1a} \le H(Y_2|V)$. The rate R_{1a} is therefore that of the information jointly transmitted to Y by both encoders. The remaining quantity in (43), that is, $I(X_1;Y|VY_2UL)$ represents the rate R_{1b} super-imposed by X_1 and decoded by Y via successive decoding. One may notice that in inequalities (44)–(45), the pair (X_1, Y_2) can replace X_1, however since Y_2 is a deterministic function of X_1 it can be dropped.

This result is based on [28], where a capacity region was established for the case of the two-user MAC with cribbing through a deterministic function at both encoders.

Theorem 11. *Achievable rate-triples (R_1, R_2, R_3) for the strictly-causal MA-CZIC with partial cribbing belong to a closure of the convex hull of the set of rate-regions given by*

$$R_1 \leq H(Y_2|V) + I(X_1; Y|VY_2UL) \tag{50}$$
$$R_2 \leq I(U; Y|TLX_1) - I(U; Z|TL) \tag{51}$$
$$R_1 + R_2 \leq I(TUX_1; Y|L) - I(TU; Z|L) \tag{52}$$
$$R_1 + R_2 \leq I(TUX_1; Y|VLY_2) + H(Y_2|V) - I(TU; Z|VL) \tag{53}$$
$$R_3 \leq I(X_3; Z|L) + \min\{I(L; Y), I(L; Z)\} \tag{54}$$

for some probability distribution of the form

$$P_{VLUTX_1Y_2X_2X_3} = P_T 1_{\{V=h(T)\}} P_{X_3} P_{L|TX_3} P_{X_1|T} 1_{\{Y_2=h(X_1)\}} P_{UX_2|VLX_3}. \tag{55}$$

The proof of Theorem 11 as well as the following Theorem Theorem 12 appear in Appendix I.

Theorem 12. *Achievable rate-triples (R_1, R_2, R_3) for the causal MA-CZIC with partial cribbing belong to the closure of the convex hull of the set of rate-regions given by* (50)–(54)*, for some probability distribution of the form*

$$P_{VLUTX_1Y_2X_2X_3} = P_T 1_{\{V=h(T)\}} P_{X_3} P_{L|TX_3} P_{X_1|T} 1_{\{Y_2=h(X_1)\}} P_{UX_2|VLX_3X_1}. \tag{56}$$

It is easy to see that setting $h(\cdot)$ to be the identity function, $T = V$ and $Y_2 = X_1$, the region of Theorem 11 degenerates to the outer bound of the MA-CZIC with noiseless cribbing (Theorem 3).

A related problem is the Cognitive State-Dependent MAC with Partial Cribbing. This setup is obtained by removing user 3 and replacing X_3 with an i.i.d. state S known non-causally at Encoder 2. From the inner bound (Theorem 9) and outer bound (Theorem 11) for the MA-CZIC with partial cribbing derived in previous sections it is immediate to derive inner and outer bounds for the channel by setting $L = 0$ and $X_3 = Z = S$. Doing so yields the following inner and outer bounds.

Theorem 13. *The closure of the convex hull of the set of rate-pairs (R_1, R_2) satisfying*

$$R_1 \leq H(Y_2|V) + I(X_1; Y|VY_2U) \tag{57}$$
$$R_2 \leq I(U; Y|VX_1) - I(U; S|V) \tag{58}$$
$$R_1 + R_2 \leq I(VUX_1; Y) - I(U; S|V) \tag{59}$$
$$R_1 + R_2 \leq I(UX_1; Y|VY_2) + H(Y_2|V) - I(U; S|V) \tag{60}$$

for some probability distribution of the form

$$P_{VLUX_1Y_2X_2S} = P_V P_S P_{X_1Y_2|V} P_{UX_2|VS} \tag{61}$$

is achievable for the state-dependent cognitive MAC with partial (strictly-causal) cribbing.

Theorem 14. *Achievable rate-pairs (R_1, R_2) for the state-dependent cognitive MAC with partial (strictly-causal) cribbing belong to the closure of the convex hull of rate-regions given by*

$$R_1 \leq H(Y_2|V) + I(X_1; Y|VY_2U) \tag{62}$$
$$R_2 \leq I(U; Y|TX_1) - I(U; S|T) \tag{63}$$
$$R_1 + R_2 \leq I(TUX_1; Y) - I(TU; S) \tag{64}$$
$$R_1 + R_2 \leq I(TUX_1; Y|VY_2) + H(Y_2|V) - I(TU; S|V) \tag{65}$$

for some probability distribution of the form

$$P_{VUTX_1Y_2X_2S} = P_T 1_{\{V=h(T)\}} P_S P_{X_1|T} 1_{\{Y_2=h(X_1)\}} P_{UX_2|VS}. \tag{66}$$

6. Discussion and Future Work

The use of cognitive radio holds tremendous promise in better exploiting the available spectrum. Sensing its environment, a cognitive radio can use it as network side information resulting in better performances for all users. The cognitive transmitter may use this information to reduce interference at its end, reduce interference for the other users or help relaying information. However, obtaining this side-information is not always practical in actual scenarios. The assumption of a-priori knowledge of the other user's information may only be applied to certain situations where the transmitters share information through a separate channel. The assumption of causally sensing the environment is more realistic in many cases of distinct transmitters. Nevertheless, the cognitive transmitter will most likely acquire a noisy version of the information limiting its ability to cooperate. In addition, sensing the environment involves complicated implementations of the transmitter as well as power consumption for which the cognition improvement is weighed against. Nevertheless, the improved transmission rates achieved via cognitive schemes motivate their integration into various wireless systems such as Wi-Fi and Cellular networks. We note that cribbing requires parallel receiver/transmit technology (duplex operation), which is useful and usually available, as in the 5G systems. Although receiving much attention recently ([15,16]), many of the fundamental problems of cognitive multi-terminal networks remain unsolved.

In this paper we investigated some cognitive aspects of multi-terminal communication networks. We introduced the MA-CZIC as generalization of a compound cognitive multi-terminal network. The MA-CZIC incorporates various multi-terminal communication channels—MAC, Z-IFC—as well as several cognition aspects—cooperation and cribbing. For the MA-CZIC we have drawn inner and outer bounds on its capacity region. In an effort to better characterize the capacity region, we studied the extreme points of the achievability region, and were able to find the capacity region in the case the channel is more-capable. Furthermore, we investigated some variations of the channel regarding the nature of cooperation between the cognitive encoder—Encoder 2—and the non-cognitive encoder sharing its receiver—Encoder 1. The case in which Encoder 2 has better cognition abilities and obtains full knowledge of Encoder 1's message was investigated. Furthermore, the case where Encoder 2 has worse cognition abilities and cribs from Encoder 1 via a deterministic function, such as quantizer, was studied.

As for possible future work, several directions can be considered. First, it would be interesting to identify some concrete non-trivial channel specification for which the MA-CZIC inner and outer bounds coincide, at least in partial regions. Finding such a channel may help us get insight about the capacity region as well as the margins given by the inner and outer bounds. Moreover, the characterization of the capacity region may be further improved by examining different interference regimes. Determining the exact capacity region for the MA-CZIC will subsequently result in the capacity region for the cognitive Z-IFC [36] as a special case. We believe that the opposite derivation also applies; i.e., the capacity region of the MA-CZIC will follow from the capacity region of the cognitive Z-IFC. Our model assumed that the cognitive transmitter—transmitter 2—has full non-causal knowledge of the interference signal X_3. While modeling the interference signal as a transmitter is very realistic in many scenarios, the assumption of non-causal knowledge of the signal, may not hold in practice in case the cognitive transmitter has sensing capabilities but not shared information. Therefore, the model where transmitter 2 cribs from transmitter 3 is very much in place, and it would be very interesting to see if it is possible to determine the capacity region for the channel. Possible iprovement of the achievable bounds may incorporate the fact that X_3 is associated with a coding scheme, and hence the interference can be mitigated by partial/full decoding, with possible aid of the cognizant transmitter 2.

Acknowledgments: This work was supported by the Heron consortium via the Israel ministry of economy and science.

Author Contributions: All three authors cooperated in the theoretical research leading to the reported results and in composing this submitted paper.

Conflicts of Interest: The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

MAC	Multiple-Access Channel
IFC	Interference Channel
OFDMA	Orthogonal Frequency-Division Multiple Access
HK	Han-Kobayashi
QoS	Quality of Service
DF	Decode and Forward
AF	Amplify and Forward
CIFC	Cognitive Interference Channel
MA-CZIC	Multiple-Access Cognitive Z-Interference Channel
P2P	point-to-point
RV	Random Variable
AEP	Asymptotic Equipartition Property

Appendix A

Proof of Theorem 1. Below is a description of the random coding scheme we use to prove achievability of rate-triples in $\mathcal{R}_{sc}$, the analysis of the average probability of error is omitted.

We propose the following coding scheme, which includes Block–Markov superposition coding, backward decoding, rate splitting and Gelfand–Pinsker coding [41]. The coding scheme combines the coding techniques of [36] with that of [27], which, in turn, is based on the coding technique of [26,29].

For a fixed distribution $P_V P_L P_{X_3|L} P_{X_1|V} P_{UX_2|VLX_3}$ the coding schemes are as follows:

Appendix A.1. Encoder 3 and Decoder 3 Coding Scheme

Appendix A.1.1. Encoder 3 Codebook generation

Generate independently $2^{n\gamma}$ codewords $\mathbf{l} = (l_1, \ldots, l_n)$, each with probability $\Pr(\mathbf{l}) = \prod_{i=1}^{n} p_L(l_i)$. These codewords constitute the inner codebook of Transmitter 3. Denote them as $\mathbf{l}(k)$ where $k \in \{1, \ldots, 2^{n\gamma}\}$. For each codeword $\mathbf{l}(k)$, generate $2^{n(R_3-\gamma)}$ codewords $\mathbf{x}_3 = (x_{3,1}, \ldots, x_{3,n})$, each with probability $\Pr(\mathbf{x}_3|\mathbf{l}(k)) = \prod_{i=1}^{n} P_{X_3|L}(x_{3,i}|l_i(k))$. Denote them as $\mathbf{x}_3(j,k)$, $j = 1, \ldots, 2^{n(R_3-\gamma)}$. The codewords $\{\mathbf{x}_3(j,k)\}_{j=1}^{2^{n(R_3-\gamma)}}$ constitute the outer codebook of Transmitter 3 associated with the codeword $\mathbf{l}(k)$.

Appendix A.1.2. Encoding Scheme of Encoder 3

Encoder 3 splits its message W_3 into two independent parts $W_3 = (W_{3a}, W_{3b})$, with rates γ and $R_3 - \gamma$ respectively. For $W_{3a} = w_{3a}$ and $W_{3b} = w_{3b}$ it transmits $\mathbf{x}_3(w_{3a}, w_{3b})$.

Appendix A.1.3. Receiver 3 Decoding

Receiver 3 looks for $\hat{w}_3 = (\hat{w}_{3a}, \hat{w}_{3b})$ such that

$$\big(\mathbf{l}(\hat{w}_{3a}), \mathbf{x}_3(\hat{w}_{3a}, \hat{w}_{3b}), \mathbf{z}\big) \in A_\epsilon(L, X_3, Z).$$

If no such $\hat{w}_3$ exists an error is declared, and if there exists more than one $\hat{w}_3$ that satisfies the condition, the decoder chooses $\hat{w}_3$ at random among them.

Appendix A.2. Encoder 1, Encoder 2 and Main Decoder Coding Scheme

We consider B Blocks, each of n symbols. A sequence of $B-1$ message pairs $(W_1^{(b)}, W_2^{(b)})$ for $b = 1, ..., B-1$, will be transmitted during B transmission blocks. As $B \to \infty$, for a fixed n, the rate pair of the message (W_1, W_2), $(\tilde{R}_1, \tilde{R}_2) = (R_1(B-1)/B,\ R_2(B-1)/B)$ converges to (R_1, R_2).

Appendix A.2.1. Encoder 1 Codebook generation

Generate 2^{nR_1} codewords $\mathbf{v} = (v_1, ..., v_n)$, each with probability $\Pr(\mathbf{v}) = \prod_{i=1}^n P_V(v_i)$. These codewords constitute the inner codebook of Transmitter 1. Denote them as $\mathbf{v}(w_0)$, where $w_0 \in \{1, ..., 2^{nR_1}\}$. For each codeword $\mathbf{v}(w_0)$ generate 2^{nR_1} codewords $\mathbf{x}_1$, each with probability $\Pr(\mathbf{x}_1|\mathbf{v}(w_0)) = \prod_{i=1}^n P_{X_1|V}(x_{1,i}|v_i(w_0))$. These codewords, $\{\mathbf{x}_1\}$, constitute the outer codebook of Transmitter 1 associated with $\mathbf{v}(w_0)$. Denote them as $\mathbf{x}_1(w_1, w_0)$ where w_0 is as before, representing the index of the codeword $\mathbf{v}(w_0)$ in the inner codebook and $w_1 \in \{1, ..., 2^{nR_1}\}$ the index of the codeword $\mathbf{x}_1$ in the associated outer codebook.

Appendix A.2.2. Encoding Scheme of Encoder 1

Given $W_1^{(b)} = w_1^{(b)} \in \{1, ..., 2^{nR_1}\}$ for $b = 1, 2, ..., B$, we define $w_0^{(b+1)} = w_1^{(b)}$ for $b = 1, 2, ..., B-1$. In block 1 Encoder 1 sends

$$\mathbf{x}_1^{(1)} = \mathbf{x}_1(w_1^{(1)}, 1),$$

in block $b = 2, 3, ..., B-1$ Encoder 1 sends

$$\mathbf{x}_1^{(b)} = \mathbf{x}_1(w_1^{(b)}, w_0^{(b)})$$

and in block B Encoder 1 sends

$$\mathbf{x}_1^{(B)} = \mathbf{x}_1(1, w_0^{(B)})$$

Appendix A.2.3. Encoder 2 Codebook Generation

This encoder's codebook is based on both Encoder 1 and Encoder 3 inner codebooks. For each two codewords $\mathbf{v}(w_0)$ and $\mathbf{l}(k)$ generate $2^{n(R_2+\beta)}$ codewords $\mathbf{u} = (u_1, ..., u_n)$, each with probability $\Pr(\mathbf{u}|\mathbf{v}(w_0)\mathbf{l}(k)) = \prod_{i=1}^n p_{U|VL}(u_i|v_i(w_0)l_i(k))$. These codewords constitute Encoder 2's codebook associated with $\mathbf{v}(w_0)$ and $\mathbf{l}(k)$. Randomly partition each of the codebooks into 2^{nR_2} bins, each consisting of $2^{n\beta}$ codewords. Now label the codewords by $\mathbf{u}(w_0, w_2, w_{3a}, t)$, where the codebook is chosen according to $\mathbf{v}(w_0)$ and $\mathbf{l}(w_{3a})$, $w_2 \in \{1, ..., 2^{nR_2}\}$ defines the bin according to Encoder 2's message, and $t \in \{1, ..., 2^{n\beta}\}$ is the index within the bin.

Appendix A.2.4. Encoding Scheme of Encoder 2

Given $(w_0^{(b)}, w_1^{(b)}, w_3^{(b)})$ as before and $W_2^{(b)} = w_2^{(b)} \in \{1, ..., 2^{nR_2}\}$, search for the lowest $t \in \{1, ..., 2^{n\beta}\}$ such that $\mathbf{u}^{(b)} = \mathbf{u}(w_0^{(b)}, w_2^{(b)}, w_{3a}^{(b)}, t)$ is jointly typical with the triplet $\big(\mathbf{v}(w_0^{(b)}), \mathbf{l}(w_{3a}^{(b)}), \mathbf{x}_3(w_3^{(b)})\big)$, denoting that t as $t(w_0^{(b)}, w_2^{(b)}, w_3^{(b)})$. If such a t is not found or if the triplet $\big(\mathbf{v}(w_0^{(b)}), \mathbf{l}(w_{3a}^{(b)}), \mathbf{x}_3(w_3^{(b)})\big)$ is not jointly typical, an error is declared and $t(w_0^{(b)}, w_2^{(b)}, w_3^{(b)}) = 1$. Now, create the codeword $\mathbf{x}_2^{(b)} = \mathbf{x}_2(w_0^{(b)}, w_2^{(b)}, w_3^{(b)})$ by drawing its components i.i.d. conditionally on the quadruple $\big(\mathbf{v}(w_0^{(b)}), \mathbf{l}(w_{3a}^{(b)}), \mathbf{u}(w_0^{(b)}, w_2^{(b)}, w_{3a}^{(b)}, t), \mathbf{x}_3(w_3^{(b)})\big)$, where the conditional law is induced by (10).

In block 1 Encoder 2 sends

$$\mathbf{x}_2(1, w_2^{(1)}, w_3^{(1)}).$$

As a result of cribbing from Encoder 1, before the beginning of block $b = 2, 3, ..., B$, Encoder 2 has an estimate $\hat{\hat{w}}_0^{(b)}$ for $w_0^{(b)}$. Then, for $b = 2, 3, ..., B-1$, Encoder 2 sends

$$\mathbf{x}_2(\hat{\hat{w}}_0^{(b)}, w_2^{(b)}, w_3^{(b)})$$

and in block B Encoder 2 sends

$$\mathbf{x}_2(\hat{\hat{w}}_0^{(B)}, 1, w_3^{(B)}).$$

Schematic description of the encoding appears in Figure A1.

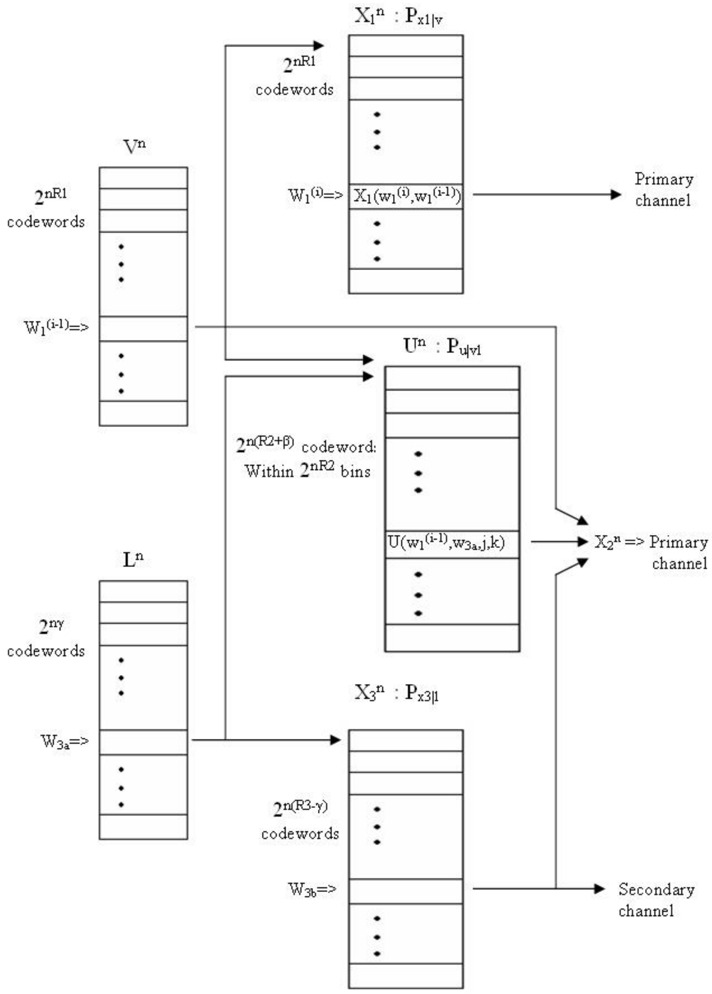

Figure A1. A schematic description of the codebooks hierarchy and encoding procedure at the three encoders.

Appendix A.2.5. Decoding at the Primary Receiver (g_1)

After receiving B blocks the decoder uses backward decoding starting from decoding block B moving on downward to block 1. In block B the receiver looks for $\hat{w}_0^{(B)} = \hat{w}_1^{(B-1)}$ such that

$$\left(\mathbf{v}(\hat{w}_1^{(B-1)}), \mathbf{x}_1(1, \hat{w}_1^{(B-1)}), \mathbf{u}(\hat{w}_1^{(B-1)}, 1, w_{3a}^{(B)}, t), \mathbf{l}(w_{3a}^{(B)}), \mathbf{y}^{(B)}\right) \in A_\epsilon(V, X_1, U, L, Y)$$

for some $w_{3a}^{(B)}$, where $t = t(\hat{w}_1^{(B-1)}, 1, w_3^{(b)})$.

At block $b = 1, 2, ..., B-1$, assuming that a decoding was done backward down to (and including) block $b+1$, the receiver decoded $\hat{w}_1^{(B-1)}, (\hat{w}_2^{(B-1)}, \hat{w}_1^{(B-2)}), ..., (\hat{w}_2^{(b+1)}, \hat{w}_1^{(b)})$. Then, to decode block b, the receiver looks for $(\hat{w}_2^{(b)}, \hat{w}_1^{(b-1)})$ such that

$$\left(\mathbf{v}(\hat{w}_1^{(b-1)}), \mathbf{x}_1(\hat{w}_1^{(b)}, \hat{w}_1^{(b-1)}), \mathbf{u}(\hat{w}_1^{(b-1)}, \hat{w}_2^{(b)}, w_{3a}^{(b)}, t), \mathbf{l}(w_{3a}^{(b)}), \mathbf{y}^{(b)}\right) \in A_\epsilon(V, X_1, U, L, Y)$$

for some $w_{3a}^{(b)}$, where $t = t(\hat{w}_1^{(b-1)}, \hat{w}_2^{(b)}, w_3^{(b)})$.

Appendix A.2.6. Decoding at Encoder 2

To obtain cooperation, after block $b = 1, 2, ..., B-1$, Encoder 2 chooses $\tilde{w}_1^{(b)}$ such that

$$\left(\mathbf{v}(\tilde{w}_0^{(b)}), \mathbf{x}_1(\tilde{w}_1^{(b)}, \tilde{w}_0^{(b)}), \mathbf{x}_1^{(b)}\right) \in A_\epsilon(V, X_1, X_1)$$

where $\tilde{w}_0{}^{(b)} = \tilde{w}_1^{(b-1)}$ was determined at the end of block $b-1$ and $\tilde{w}_0^{(1)} = 1$.

At each of the decoders, if a decoding step either fails to recover a unique index (or index pair) which satisfies the decoding rule, or there is more than one index (or index pair), then an index (or index pair) is chosen at random among the indices which satisfies the decoding rule.

It can be shown that the error probability will be arbitrarily small if (9a)–(9d) hold.

Appendix A.3. Bounding the Probability of Error

We define the error events $E_0^{(b)} - E_7^{(b)}$ as follows:

- $E_0^{(b)}$: Codebook error, the codewords $\mathbf{v}$,$\mathbf{x}_1$,$\mathbf{l}$,$\mathbf{x}_3$ are not jointly typical. That is

$$\left(\mathbf{v}(w_0^{(b)}), \mathbf{x}_1(w_1^{(b)}, w_0^{(b)})\right) \notin A_\epsilon(V, X_1) \cup \left(\mathbf{l}(w_{3a}^{(b)}), \mathbf{x}_3(w_{3a}^{(b)}, w_{3b}^{(b)})\right) \notin A_\epsilon(L, X_3)$$

- $E_1^{(b)}$: Error decoding $w_1^{(b)}$ at Encoder 2, that is, there exists $\tilde{w}_1^{(b)} \neq w_1^{(b)}$ such that

$$\left(\mathbf{v}(w_0^{(b)}), \mathbf{x}_1(\tilde{w}_1^{(b)}, w_0^{(b)}), \mathbf{x}_1^{(b)}\right) \in A_\epsilon(V, X_1, X_1) \tag{A1}$$

- $E_2^{(b)}$: Encoding error at Encoder 2, no suitable encoding index t. That is, there is no $t \in \{1, ..., 2^{n\beta}\}$ such that

$$\left(\mathbf{v}(w_0^{(b)}), \mathbf{l}(w_{3a}^{(b)}), \mathbf{u}(w_0^{(b)}, w_2^{(b)}, w_{3a}^{(b)}, t), \mathbf{x}_3(w_{3a}^{(b)}, w_{3b}^{(b)})\right) \in A_\epsilon(V, L, U, X_3)$$

- $E_3^{(b)}$: Channel error, one or more of the input signals is not jointly typical with the outputs $\mathbf{y}$ and $\mathbf{z}$. That is

$$\begin{aligned}&\left(\mathbf{v}(w_0^{(b)}), \mathbf{u}(w_0^{(b)}, w_2^{(b)}, w_{3a}^{(b)}, t), \mathbf{x}_1(w_1^{(b)}, w_0^{(b)}), \mathbf{l}(w_{3a}^{(b)}), \mathbf{x}_3(w_{3a}^{(b)}, w_{3b}^{(b)}), \mathbf{y}^{(b)}, \mathbf{z}^{(b)}\right)\\ &\notin A_\epsilon(V, U, X_1, L, X_3, Y, Z)\end{aligned} \tag{A2}$$

- $E_4^{(b)}$: Codebook error in decoding $w_{3a}^{(b)}$ at either one of the decoders, a false message was detected. That is, there exists $\tilde{w}_{3a}^{(b)} \neq w_{3a}^{(b)}$ such that

$$\left(\mathbf{l}(\tilde{w}_{3a}^{(b)}), \mathbf{y}^{(b)}\right) \in A_\epsilon(L, Y) \cup \left(\mathbf{l}(\tilde{w}_{3a}^{(b)}), \mathbf{z}^{(b)}\right) \in A_\epsilon(L, Z)$$

- $E_5^{(b)}$: Codebook error in decoding $w_0^{(b)}$. There exists $\tilde{w}_0^{(b)} \neq w_0^{(b)}$ such that

$$\left(\mathbf{v}(\tilde{w}_0^{(b)}), \mathbf{u}(\tilde{w}_0^{(b)}, j, w_{3a}^{(b)}, t), \mathbf{x}_1(w_1^{(b)}, \tilde{w}_0^{(b)}), \mathbf{l}(w_{3a}^{(b)}), \mathbf{y}^{(b)}\right) \in A_\epsilon(V, U, X_1, L, Y)$$

for some pair (j,t) $j \in W_2$, $t \in \{1,...,2^{n\beta}\}$.

- $E_6^{(b)}$: Codebook error in decoding $w_2^{(b)}$. There exists a different bin $\tilde{w}_2^{(b)} \neq w_2^{(b)}$, such that

$$\left(\mathbf{v}(w_0^{(b)}), \mathbf{u}(w_0^{(b)}, \tilde{w}_2^{(b)}, w_{3a}^{(b)}, t), \mathbf{x}_1(w_1^{(b)}, w_0^{(b)}), \mathbf{l}(w_{3a}^{(b)}), \mathbf{y}^{(b)}\right) \in A_\epsilon(V,U,X_1,L,Y)$$

for some $t \in \{1,...,2^{n\beta}\}$.

- $E_7^{(b)}$: Codebook error in decoding $w_{3b}^{(b)}$. There exists $\tilde{w}_{3b}^{(b)} \neq w_{3b}^{(b)}$ such that

$$\left(\mathbf{l}(w_{3a}^{(b)}), \mathbf{x}_3(w_{3a}^{(b)}, \tilde{w}_{3b}^{(b)}), \mathbf{z}^{(b)}\right) \in A_\epsilon(L,X_3,Z)$$

Notice that when Encoder 2 observes $\mathbf{x}_1$ error-free, as in this setup, the error event $E_1^{(b)}$ (A1) can be replaced with the explicit case of having two identical codewords in $\{\mathbf{x}_1\}$ codebook, i.e, there exists $\tilde{w}_1^{(b)} \neq w_1^{(b)}$ such that $\mathbf{x}_1(\tilde{w}_1^{(b)}, w_0^{(b)}) = \mathbf{x}_1^{(b)}$.

We now define the events $F_i,\ i = 0\ldots5$ as follows:

- $F_0 \triangleq \bigcup_{b=1}^B E_0^{(b)}$
- $F_1 \triangleq \bigcup_{b=1}^B (E_0^{(b)} \cup E_1^{(b)})$
- $F_2 \triangleq \bigcup_{b=1}^B (E_0^{(b)} \cup E_1^{(b)} \cup E_2^{(b)})$
- $F_3 \triangleq \bigcup_{b=1}^B (E_0^{(b)} \cup E_1^{(b)} \cup E_2^{(b)} \cup E_3^{(b)})$
- $F_4 \triangleq \bigcup_{b=1}^B (E_0^{(b)} \cup E_1^{(b)} \cup E_2^{(b)} \cup E_3^{(b)} \cup E_4^{(b)})$
- $F_5^{(b)} \triangleq \bigcup_{i=5}^7 E_i^{(b)}, \quad b = 1,...,B$

We upper bound the average probability of error $\bar{P}_e$ averaged over all codebooks and all random partitions, as in [27], by

$$\begin{aligned}\bar{P}_e \leq &\sum_{b=1}^B \{\Pr[E_0^{(b)}] + \Pr[E_1^{(b)}|F_0^c, E_1^{(1\ldots b-1)^C}]\} + \sum_{b=1}^B \{\Pr[E_2^{(b)}|F_1^c] + \Pr[E_3^{(b)}|F_2^c, E_3^{(1\ldots b-1)^C}]\} \\ &+ \sum_{b=1}^B \{\Pr[E_4^{(b)}|F_3^c] + \Pr[F_5^{(b)}|F_4^c, F_5^{(b+1\ldots B)^C}]\}\end{aligned} \tag{A3}$$

where $F^{(1\ldots b-1)^C}$ denotes the complement of the event $\bigcup_{i=1}^{b-1} F^{(i)}$.

Furthermore, we can upper bound each of the summands in the last component of (A3) by the union bound as

$$\begin{aligned}&\Pr[F_5^{(b)}|F_4^c, F_5^{(b+1\ldots B)^C}] \\ &= \Pr\left(\bigcup_{i=5}^7 E_i^{(b)}|F_4^c, F_5^{(b+1\ldots B)^C}\right) \\ &\leq \sum_{i=5}^7 \Pr(E_i^{(b)}|F_4^c, F_5^{(b+1\ldots B)^C}).\end{aligned} \tag{A4}$$

Now, we can separately examine and upper bound each of the summands in (A3):

- By the Asymptotic Equipartition Property (AEP) [40], $\Pr[E_0^{(b)}] \to 0$ as $n \to \infty$.
- In the second summand, $\Pr[E_1^{(b)}|F_0^c, E_1^{(1\ldots b-1)^C}]$, the conditioning on $F_0^c, E_1^{(1\ldots b-1)^C}$ insures that $(\mathbf{x}_1^{(b)}, \mathbf{v}^{(b)})$ are jointly typical, and that Encoder 2 decoded correctly all the previous messages $w_1^{(1)}, ..., w_1^{(b-1)}$ and specifically $w_0^{(b)}$ $(= w_1^{(b-1)})$. Since each codeword $\mathbf{x}_1(\cdot, w_0^{(b)})$ is drawn i.i.d. given $w_0^{(b)}$, from the strong typicality Lemma we get

$$Pr[E_{1,j}^{(b)}|F_0^c, E_1^{(1\ldots b-1)^C}] \leq 2^{-n[H(X_1|V)-\epsilon]}, \quad \forall j \neq w_1^{(b)} \tag{A5}$$

where $E_{1,j}^{(b)}$ is the event

$$\left(\mathbf{v}(w_0^{(b)}), \mathbf{x}_1(j, w_0^{(b)}), \mathbf{x}_1^{(b)}\right) \in A_\epsilon(V, X_1, X_1). \tag{A6}$$

Assuming, without loss of generality, that $w_1^{(b)} = 1$, we get by using the union bound

$$\Pr[E_1^{(b)}|F_0^c, E_1^{(1\ldots b-1)^C}] \leq \sum_{j=2}^{2^{nR_1}} Pr[E_{1,j}^{(b)}|F_0^c, E_1^{(1\ldots b-1)^C}] \leq (2^{nR_1} - 1) \times 2^{-n[H(X_1|V)-\epsilon]}.$$

Hence for

$$R_1 \leq H(X_1|V) - \epsilon \tag{A7}$$

we get $\Pr[E_1^{(b)}|F_0^c, E_1^{(1\ldots b-1)^C}] \to 0$ as $n \to \infty$.

- Since the codewords $\{\mathbf{u}\}$ are generated in an i.i.d. manner we have

$$E_2^{(b)} = \bigcap_{t=1}^{2^{n\beta}} E_{2,t}^{(b)} \tag{A8}$$

where $E_{2,t}^{(b)}$ is the event

$$\left(\mathbf{v}(w_0^{(b)}), \mathbf{l}(w_{3a}^{(b)}), \mathbf{u}(w_0^{(b)}, w_2^{(b)}, w_{3a}^{(b)}, t), \mathbf{x}_3(w_3^{(b)})\right) \notin A_\epsilon(V, L, U, X_3)$$

for a specific index t. Hence, we have

$$\Pr[E_2^{(b)}|F_1^c] = \prod_{t=1}^{2^{n\beta}} \Pr[E_{2,t}^{(b)}|F_1^c]. \tag{A9}$$

Conditioning on V and L in ([41], Lemma 3) we get

$$\Pr[E_{2,t}^{(b)}|F_1^c] \leq 1 - 2^{-n[I(U;X_3|VL)+\epsilon_1]} \tag{A10}$$

for all $t \in \{1, ..., 2^{n\beta}\}$, where $\epsilon_1 \to 0$ as $\epsilon \to 0$. Hence

$$\Pr[E_2^{(b)}|F_1^c] \leq (1 - 2^{-n[I(U;X_3|VL)+\epsilon_1]})^{2^{n\beta}}. \tag{A11}$$

The expression converges to 0 as $n \to \infty$ for

$$\beta > I(U; X_3|VL) + \epsilon_1. \tag{A12}$$

- By the AEP $\Pr[E_3^{(b)}|F_2^c, E_3^{(1\ldots b-1)^C}] \to 0$ as $n \to \infty$.
- If

$$\gamma \leq \min\{I(L;Y), I(L;Z)\} \tag{A13}$$

then, from joint typicality decoding, $\Pr[E_4^{(b)}|F_3^c] \to 0$ as $n \to \infty$.

- We state that

$$\Pr[E_5^{(b)}|F_4^c, F_5^{(b+1\ldots B)^C}] \leq \sum_{i=1}^{2^{nR_1}} \sum_{j=1}^{2^{nR_2}} \sum_{t=1}^{2^{n\beta}} \Pr[E_{5,ijt}^{(b)}|F_4^c, F_5^{(b+1\ldots B)^C}]$$

where $E_{5,ijt}^{(b)}$ stands for the event

$$\left(\mathbf{v}(i), \mathbf{u}(i, j, w_{3a}^{(b)}, t), \mathbf{x}_1(w_1^{(b)}, i), \mathbf{l}(w_{3a}^{(b)}), \mathbf{y}^{(b)}\right) \in A_\epsilon(V, U, X_1, L, Y).$$

By using the strong typicality Lemma we bound each of the summands above

$$\Pr[E_{5,ijt}^{(b)}|F_4^c, F_5^{(b+1\dots B)^C}] \le \frac{2^{n[H(VUX_1LY)+\epsilon]}}{2^{n[H(VUX_1L)-\epsilon]}2^{n[H(Y|L)-\epsilon]}}$$
$$= 2^{-n[I(UVX_1;Y|L)-3\epsilon]}.$$

Summing over all codewords we get

$$\Pr[E_5^{(b)}|F_4^c, F_5^{(b+1\dots B)^C}] \le 2^{-n[I(UVX_1;Y|L)+3\epsilon]}2^{n(R_1+R_2+\beta)}.$$

Therefore, if

$$R_1 + R_2 + \beta \le I(VUX_1; Y|L) + 3\epsilon \tag{A14}$$

then $\Pr[E_5^{(b)}|F_4^c, F_5^{(b+1\dots B)^C}] \to 0$ as $n \to \infty$.

- Similarly, using the same technique as the previous step, if

$$R_2 + \beta \le I(U; Y|VLX_1) + 3\epsilon \tag{A15}$$

then $\Pr[E_6^{(b)}|F_4^c, F_5^{(b+1\dots B)^C}] \to 0$ as $n \to \infty$.

- Finally, if

$$R_3 - \gamma \le I(X_3; Z|L) \tag{A16}$$

then $\Pr[E_7^{(b)}|F_4^c, F_5^{(b+1\dots B)^C}] \to 0$ as $n \to \infty$.

From (A7)–(A16) we get (9a)–(9d), thus concluding the proof. □

Appendix B

Outline of the Proof of Theorem 2:

The achievability part follows similarly to that of Theorem 1, the only difference being in the way the codeword $\mathbf{x}_2(\hat{\hat{w}}_0^{(b)}, w_2^{(b)}, w_3^{(b)})$ is generated. Here, the second encoder generates the codeword $\mathbf{x}_2(\hat{\hat{w}}_0^{(b)}, w_2^{(b)}, w_3^{(b)})$ by drawing its components i.i.d. conditionally on the quintuple $\big(\mathbf{v}(w_0^{(b)}), \mathbf{l}(w_{3a}^{(b)}), \mathbf{u}^{(b)}, \mathbf{x}_3^{(b)}, x_1^{(b)}\big)$, where the conditional law is induced by (11).

Appendix C

Proof of Theorem 3—Strictly Causal MA-CZIC Outer Bound. Consider an $(2^{nR_1}, 2^{nR_2}, 2^{nR_3}, n)$ code with average block error probability $P_e^{(n)} = n\epsilon_n$, and a probability distribution on $\mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{W}_3 \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \mathcal{Y} \times \mathcal{Z}$ given by

$$p_{W_1W_2W_3X_1^nX_2^nX_3^nY^nZ^n} = p_{W_1}p_{W_2}p_{W_3}1_{\{X_1^n=f_1(W_1)\}}1_{\{X_3^n=f_3(W_3)\}} \cdot \prod_{i=1}^{n} p_{X_{2i}|W_2W_3X_1^{i-1}}p_{Y_i|X_{1i}X_{2i}X_{3i}}p_{Z_i|X_{3i}}. \tag{A17}$$

For $i \in \{1, 2, ..., n\}$, let V_i, L_i and U_i be the random variables defined by

$$V_i \triangleq X_1^{i-1}, \quad L_i \triangleq (Y^{i-1}, Z_{i+1}^n), \quad U_i \triangleq W_2. \tag{A18}$$

and let U be the random variable defined by further defining Q to be an auxiliary (time-sharing) random variable that is distributed uniformly on the set $\{1, 2, ..., n\}$, and let

$$V \triangleq (V_Q, Q), \quad X_1 \triangleq X_{1Q}, \quad L \triangleq (L_Q, Q),$$
$$X_3 \triangleq X_{3Q}, \quad Y \triangleq Y_Q, \quad Z \triangleq Z_Q, \quad U \triangleq U_Q. \tag{A19}$$

We start with an upper bound on R_1

$$\begin{aligned}
nR_1 &= H(W_1|W_2) \\
&= I(W_1;Y^n|W_2) + H(W_1|W_2Y^n) \\
&\leq I(W_1;Y^n|W_2) + n\epsilon_n \\
&\overset{(a)}{=} I(X_1^n;Y^n|W_2) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(X_{1i};Y^n|W_2X_1^{i-1}) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} H(X_{1i}|X_1^{i-1}) + n\epsilon_n \\
&= \sum_{i=1}^{n} H(X_{1i}|V_i) + n\epsilon_n \\
&= H(X_{1Q}|V_QQ) + n\epsilon_n \\
&= H(X_1|V) + n\epsilon_n,
\end{aligned} \tag{A20}$$

where (a) follows from the encoding relation in (2).

Next, consider R_3

$$\begin{aligned}
nR_3 &= H(W_3) \\
&= I(W_3;Z^n) + H(W_3|Z^n) \\
&\leq I(W_3;Z^n) + n\epsilon_n \\
&= H(Z^n) - H(Z^n|W_3) + n\epsilon_n \\
&\overset{(b)}{\leq} H(Z^n) - H(Z^n|W_3X_3^n) + n\epsilon_n \\
&\overset{(c)}{=} H(Z^n) - H(Z^n|X_3^n) + n\epsilon_n \\
&\overset{(d)}{=} H(Z^n) - \sum_{i=1}^{n} H(Z_i|X_{3i}) + n\epsilon_n,
\end{aligned} \tag{A21}$$

where (b) follows from the fact that conditioning decreases entropy, (c) follows from the Markov chain $W_3 - X_3^n - Z^n$ and (d) follows since the channel $P_{Z|X_3}$ is memoryless.

Using the Csiszár-Körner's identity ([43], Lemma 7) we obtain

$$\begin{aligned}
H(Y^n) - H(Z^n) &= \sum_{i=1}^{n} [H(Y_i|Y^{i-1}Z_{i+1}^n) \\
&\qquad - H(Z_i|Y^{i-1}Z_{i+1}^n)] \\
&= \sum_{i=1}^{n} [H(Y_i|L_i) - H(Z_i|L_i)],
\end{aligned} \tag{A22}$$

where the last equality follows from (A18). Substituting (A19) into (A22) we get

$$\frac{1}{n}(H(Y^n) - H(Z^n)) = H(Y|L) - H(Z|L). \tag{A23}$$

Notice that (A23) implies that there exists a number γ where

$$\gamma = \frac{1}{n}H(Y^n) - H(Y|L) = \frac{1}{n}H(Z^n) - H(Z|L) \tag{A24}$$

$$0 \leq \gamma \leq \min\{I(L;Y), I(L;Z)\} \tag{A25}$$

where the right inequality of (A25) follows since $H(Y^n) \leq nH(Y)$ and $H(Z^n) \leq nH(Z)$, and the left inequality follows since

$$H(Y^n) = \sum_{i=1}^{n} H(Y_i|Y^{i-1}) \geq \sum_{i=1}^{n} H(Y_i|Y^{i-1}Z_{i+1}^n) = nH(Y|L). \tag{A26}$$

Following from (A21) we have

$$\begin{aligned}
R_3 &\leq \frac{1}{n}H(Z^n) - \frac{1}{n}\sum_{i=1}^{n} H(Z_i|X_{3i}) + \epsilon_n \\
&\overset{(e)}{=} H(Z|L) + \gamma - H(Z|X_3Q) + \epsilon_n \\
&\overset{(f)}{=} H(Z|L) + \gamma - H(Z|X_3) + \epsilon_n \\
&\leq H(Z|L) - H(Z|X_3) + \min\{I(L;Y), I(L;Z)\} + \epsilon_n \\
&\overset{(g)}{=} H(Z|L) - H(Z|X_3L) + \min\{I(L;Y), I(L;Z)\} + \epsilon_n \\
&= I(X_3;Z|L) + \min\{I(L;Y), I(L;Z)\} + \epsilon_n,
\end{aligned} \tag{A27}$$

where (e) follows from (A24) and the definitions of random variables in (A19); (f) follows since the channel $P_{Z|X_3}$ is memoryless, and (g) follows from the Markov chain $L - X_3 - Z$.

Next, consider R_2

$$\begin{aligned}
R_2 &= \frac{1}{n}H(W_2|W_1) \\
&\leq \frac{1}{n}I(W_2;Y^n|W_1) + \epsilon_n \\
&= \frac{1}{n}H(Y^n|W_1) - \frac{1}{n}H(Y^n|W_1W_2) + \epsilon_n.
\end{aligned} \tag{A28}$$

By conditioning (A22) on W_1 we get

$$\begin{aligned}
H(Y^n|W_1) - H(Z^n|W_1) &= \sum_{i=1}^{n}[H(Y_i|Y^{i-1}Z_{i+1}^nW_1) - H(Z_i|Y^{i-1}Z_{i+1}^nW_1)] \\
&\overset{(h)}{=} \sum_{i=1}^{n}[H(Y_i|Y^{i-1}Z_{i+1}^nW_1X_{1i}X_1^{i-1}) - H(Z_i|Y^{i-1}Z_{i+1}^nW_1X_1^{i-1})] \\
&\overset{(i)}{=} \sum_{i=1}^{n}[H(Y_i|Y^{i-1}Z_{i+1}^nX_{1i}X_1^{i-1}) - H(Z_i|Y^{i-1}Z_{i+1}^nX_1^{i-1})] \\
&= \sum_{i=1}^{n}[H(Y_i|L_iX_{1i}V_i) - H(Z_i|L_iV_i)],
\end{aligned} \tag{A29}$$

and hence

$$\frac{1}{n}H(Y^n|W_1) = \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_iX_{1i}V_i) - H(Z_i|L_iV_i)] + \frac{1}{n}H(Z^n|W_1), \tag{A30}$$

where (h) follows from the encoding relation in (2) and (i) follows since $W_1 - (X_{1i}, X_1^{i-1}, Y^{i-1}) - Y_i$ and $W_1 - (X_1^{i-1}, Y^{i-1}, Z_{i+1}^n) - Z_i$ are Markov chains.

In the same manner, conditioning (A22) on (W_1, W_2) yields

$$\frac{1}{n}H(Y^n|W_1W_2) = \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_iX_{1i}V_iW_2) - H(Z_i|L_iV_iW_2)] + \frac{1}{n}H(Z^n|W_1W_2). \tag{A31}$$

Substituting (A30) and (A31) into (A28) we get

$$\begin{aligned}
R_2 &\le \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_iX_{1i}V_i) - H(Z_i|L_iV_i)] + \frac{1}{n}H(Z^n|W_1) - \frac{1}{n}H(Z^n|W_1W_2) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_iX_{1i}V_iW_2) - H(Z_i|L_iV_iW_2)] + \epsilon_n \\
&\overset{(j)}{=} \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_iX_{1i}V_i) - H(Z_i|L_iV_i)] - \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_iX_{1i}V_iW_2) - H(Z_i|L_iV_iW_2)] + \epsilon_n \\
&\overset{(k)}{=} H(Y_Q|L_QX_{1Q}V_QQ) - H(Z_Q|L_QV_QQ) - H(Y_Q|L_QX_{1Q}V_QU_QQ) + H(Z_Q|L_QV_QU_QQ) + \epsilon_n \\
&= I(U_Q;Y_Q|V_QX_{1Q}L_QQ) - I(U_Q;Z_Q|V_QL_QQ) + \epsilon_n \\
&\overset{(l)}{=} I(U;Y|LX_1V) - I(U;Z|LV) + \epsilon_n,
\end{aligned} \tag{A32}$$

where (j) follows since Z^n is independent of (W_1, W_2) and therefore $H(Z^n|W_1) = H(Z^n|W_1W_2) = H(Z^n)$, (k) follows from the definitions of random variables in (A19), and (l) follows since the channel $P_{Y|X_1X_2X_3}P_{Z|X_3}$ is memoryless.

Finally, we consider the sum-rate $R_1 + R_2$

$$\begin{aligned}
R_1 + R_2 &= \frac{1}{n}H(W_1W_2) \\
&\le \frac{1}{n}I(W_1W_2;Y^n) + \epsilon_n \\
&= \frac{1}{n}H(Y^n) - \frac{1}{n}H(Y^n|W_1W_2) + \epsilon_n \\
&\overset{(m)}{=} \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_i) - H(Z_i|L_i)] + \epsilon_n + \frac{1}{n}H(Z^n) - \frac{1}{n}H(Z^n|W_1W_2) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_iX_{1i}V_iW_2) - H(Z_i|L_iV_iW_2)] \\
&\overset{(n)}{=} \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_i) - H(Z_i|L_i)] + \epsilon_n - \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_iX_{1i}V_iW_2) - H(Z_i|L_iV_iW_2)] \\
&\overset{(o)}{=} H(Y_Q|L_QQ) - H(Z_Q|L_QQ) + \epsilon_n - H(Y_Q|L_QX_{1Q}V_QU_QQ) + H(Z_Q|L_QV_QU_QQ) \\
&= I(V_QU_QX_{1Q};Y_Q|L_QQ) - I(V_QU_Q;Z_Q|L_QQ) + \epsilon_n \\
&\overset{(p)}{=} I(VUX_1;Y|L) - I(VU;Z|L) + \epsilon_n,
\end{aligned} \tag{A33}$$

where (m) follows from (A22) and (A31), (n) follows since Z^n is independent of W_1 and W_2 and therefore $H(Z^n|W_1) = H(Z^n|W_1W_2) = H(Z^n)$, (o) follows from the definitions of random variables in (A19), and (p) follows since the channel $P_{Y|X_1X_2X_3}P_{Z|X_3}$ is memoryless.

It remains to show that the joint law of the auxiliary random variables satisfy (20); i.e., we wish to show that the RVs V_i, U_i and L_i as chosen in (A18) satisfy

$$p_{U_kV_kL_kX_{1,k}X_{2,k}X_{3,k}} = p_{V_k}p_{X_{1,k}|V_k}p_{X_{3,k}}p_{L_k|V_kX_{3,k}}p_{U_k|L_kV_kX_{3,k}}p_{X_{2,k}|U_kV_kX_{3,k}L_k}. \tag{A34}$$

From (A17) and the encoding rules (2)–(4) we may write

$$\begin{aligned}
&p_{W_1W_2W_3X_1^kX_2^kX_3^nY^{k-1}Z^n} \\
&= p_{W_1}p_{X_1^{k-1}|W_1}p_{X_{1,k}|W_1X_1^{k-1}}p_{W_3}p_{X_3^n|W_3}\cdot p_{Z^n|X_3^n}p_{W_2}p_{X_2^k|W_2X_1^{k-1}W_3}p_{Y^{k-1}|X_1^{k-1}X_2^{k-1}X_3^{k-1}}.
\end{aligned} \tag{A35}$$

Since $p_{W_3}p_{X_3^n|W_3} = p_{X_3^n}p_{W_3|X_3^n}$, summing this joint law over w_1, w_3 and all possible sub-sequences z^k we obtain

$$
\begin{aligned}
& p_{W_2X_1^kX_2^kX_3^nY^{k-1}Z_{k+1}^n} \\
= & \sum_{w_1,w_3,z^k} p_{W_1W_2W_3X_1^kX_2^kX_3^nY^{k-1}Z^n} \\
= & p_{X_1^{k-1}}p_{X_{1,k}|X_1^{k-1}}p_{X_3^n}p_{Z_{k+1}^n|X_{3,k+1}^n} \cdot p_{W_2}p_{X_2^k|W_2X_1^{k-1}X_3^n}p_{Y^{k-1}|X_1^{k-1}X_2^{k-1}X_3^{k-1}} \\
= & p_{X_1^{k-1}}p_{X_{1,k}|X_1^{k-1}}p_{X_{3,k}}p_{X_3^{k-1}|X_{3,k}}p_{X_{3,k+1}^n|X_3^{k-1}X_{3,k}} \\
& \cdot p_{Z_{k+1}^n|X_{3,k+1}^n}p_{W_2}p_{X_2^k|W_2X_1^{k-1}X_3^{k-1}X_{3,k}X_{3,k+1}^n} \cdot p_{Y^{k-1}|X_1^{k-1}X_2^{k-1}X_3^{k-1}}.
\end{aligned} \tag{A36}
$$

From the memorylessness of the channel we get

$$
\begin{aligned}
& p_{X_{3,k}}p_{X_3^{k-1}|X_{3,k}}p_{X_{3,k+1}^n|X_3^{k-1}X_{3,k}}p_{Z_{k+1}^n|X_{3,k+1}^n} \\
= & p_{X_3^{k-1}X_{3,k}X_{3,k+1}^nZ_{k+1}^n} \\
= & p_{X_{3,k}}p_{Z_{k+1}^n|X_{3,k}}p_{X_3^{k-1}|X_{3,k}Z_{k+1}^n}p_{X_{3,k+1}^n|X_3^{k-1}X_{3,k}Z_{k+1}^n}.
\end{aligned} \tag{A37}
$$

From (A37), summing the joint law in (A36) over all possible sub-sequences $x_{3,k+1}^n$ we obtain

$$
\begin{aligned}
& p_{W_2X_1^kX_2^kX_3^kY^{k-1}Z_{k+1}^n} \\
= & \sum_{x_{3,k+1}^n} p_{W_2X_1^kX_2^kX_3^nY^{k-1}Z_{k+1}^n} \\
= & p_{X_1^{k-1}}p_{X_{1,k}|X_1^{k-1}}p_{X_{3,k}}p_{Z_{k+1}^n|X_{3,k}}p_{X_3^{k-1}|Z_{k+1}^nX_{3,k}}p_{W_2}p_{X_2^k|W_2X_1^{k-1}X_3^{k-1}X_{3,k}Z_{k+1}^n}p_{Y^{k-1}|X_1^{k-1}X_2^{k-1}X_3^{k-1}}.
\end{aligned} \tag{A38}
$$

From the memoryless property of the channel we may write

$$
\begin{aligned}
& p_{X_2^k|W_2X_1^{k-1}X_3^{k-1}X_{3,k}Z_{k+1}^n}p_{Y^{k-1}|X_1^{k-1}X_2^{k-1}X_3^{k-1}} \\
= & p_{X_{2,k}|W_2X_1^{k-1}X_2^{k-1}X_3^{k-1}X_{3,k}Z_{k+1}^n} \cdot p_{X_2^{k-1}|W_2X_1^{k-1}X_3^{k-1}X_{3,k}Z_{k+1}^n} \cdot p_{Y^{k-1}|X_1^{k-1}X_2^{k-1}X_3^{k-1}W_2X_{3,k}Z_{k+1}^n} \\
= & p_{X_{2,k}Y^{k-1}|W_2X_1^{k-1}X_2^{k-1}X_3^{k-1}X_{3,k}Z_{k+1}^n} \cdot p_{X_2^{k-1}|W_2X_1^{k-1}X_3^{k-1}X_{3,k}Z_{k+1}^n}.
\end{aligned} \tag{A39}
$$

Summing (A38) over all possible sub-sequences (x_2^{k-1}, x_3^{k-1}) and using (A39) we obtain

$$
\begin{aligned}
& p_{W_2X_1^kX_{2,k}X_{3,k}Y^{k-1}Z_{k+1}^n} \\
= & \sum_{(x_2^{k-1},x_3^{k-1})} p_{W_2X_1^kX_2^kX_3^kY^{k-1}Z_{k+1}^n} \\
= & p_{X_1^{k-1}}p_{X_{1,k}|X_1^{k-1}}p_{X_{3,k}}p_{Z_{k+1}^n|X_{3,k}}p_{W_2}p_{X_{2,k}Y^{k-1}|W_2X_1^{k-1}X_{3,k}Z_{k+1}^n} \\
= & p_{X_1^{k-1}}p_{X_{1,k}|X_1^{k-1}}p_{X_{3,k}}p_{Z_{k+1}^n|X_{3,k}}p_{W_2}p_{X_{2,k}|W_2X_1^{k-1}X_{3,k}Y^{k-1}Z_{k+1}^n}p_{Y^{k-1}|W_2X_1^{k-1}X_{3,k}Z_{k+1}^n}.
\end{aligned} \tag{A40}
$$

From the encoding rules (2)–(4) it holds that $p_{Z_{k+1}^n|X_{3,k}} = p_{Z_{k+1}^n|W_2X_1^{k-1}X_{3,k}}$, hence we can write (A40) as

$$
\begin{aligned}
& p_{W_2X_1^kX_{2,k}X_{3,k}Y^{k-1}Z_{k+1}^n} \\
= & p_{W_2X_1^{k-1}X_{1,k}X_{2,k}X_{3,k}Y^{k-1}Z_{k+1}^n} \\
= & p_{X_1^{k-1}}p_{X_{1,k}|X_1^{k-1}}p_{X_{3,k}}p_{W_2} \cdot p_{X_{2,k}|W_2X_1^{k-1}X_{3,k}Y^{k-1}Z_{k+1}^n}p_{Y^{k-1}Z_{k+1}^n|W_2X_1^{k-1}X_{3,k}}.
\end{aligned} \tag{A41}
$$

From (A41) and (A18) we get

$$p_{U_kV_kL_kX_{1,k}X_{2,k}X_{3,k}} = p_{V_k}p_{X_{1,k}|V_k}p_{X_{3,k}}p_{U_k}p_{X_{2,k}|U_kV_kX_{3,k}L_k}p_{L_k|U_kV_kX_{3,k}}. \tag{A42}$$

Using the identity

$$\begin{aligned} p_{L_k|U_kV_kX_{3,k}} &= \frac{p_{U_k|L_kV_kX_{3,k}}p_{L_kV_kX_{3,k}}}{p_{U_kV_kX_{3,k}}} \\ &= \frac{p_{U_k|L_kV_kX_{3,k}}p_{L_k|V_kX_{3,k}}p_{V_k}p_{X_{3,k}}}{p_{U_k}p_{V_k}p_{X_{3,k}}} \\ &= \frac{p_{U_k|L_kV_kX_{3,k}}p_{L_k|V_kX_{3,k}}}{p_{U_k}} \end{aligned} \tag{A43}$$

in (A42) we get (20), thus establishing the desired form of the probability function. □

Appendix D

Outline of the proof of Theorem 4—Causal MA-CZIC Outer Bound:

The outer-bound for the *causal* MA-CZIC follows similarly to that of Theorem 2. Consider an $(2^{nR_1}, 2^{nR_2}, 2^{nR_3}, n)$ code with average block error probability $P_e^{(n)} = n\epsilon_n$, and a probability distribution on $\mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{W}_3 \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \times \mathcal{Y} \times \mathcal{Z}$ given by

$$p_{W_1W_2W_3X_1^nX_2^nX_3^nY^nZ^n} = p_{W_1}p_{W_2}p_{W_3}1_{\{X_1^n=f_1(W_1)\}}1_{\{X_3^n=f_3(W_3)\}} \cdot \prod_{i=1}^{n} p_{X_{2i}|W_2W_3X_1^i}p_{Y_i|X_{1i}X_{2i}X_{3i}}p_{Z_i|X_{3i}}. \tag{A44}$$

The inequalities for the causal cribbing case are identical to (A20), (A27), (A32) and (A33). It remains to obtain the joint law of the random variables.

It can be seen that in this case (A41) becomes

$$\begin{aligned} &p_{W_2X_1^kX_{2,k}X_{3,k}Y^{k-1}Z_{k+1}^n} \\ &= p_{W_2X_1^{k-1}X_{1,k}X_{2,k}X_{3,k}Y^{k-1}Z_{k+1}^n} \\ &= p_{X_1^{k-1}}p_{X_{1,k}|X_1^{k-1}}p_{X_{3,k}}p_{W_2} \cdot p_{X_{2,k}|W_2X_1^{k-1}X_{1,k}X_{3,k}Y^{k-1}Z_{k+1}^n}p_{Y^{k-1}Z_{k+1}^n|W_2X_1^{k-1}X_{3,k}} \end{aligned} \tag{A45}$$

and using (A43) we get from (A45)

$$p_{U_kV_kL_kX_{1,k}X_{2,k}X_{3,k}} = p_{V_k}p_{X_{1,k}|V_k}p_{X_{3,k}}p_{L_k|V_kX_{3,k}}p_{X_{2,k}|V_kX_{1,k}X_{3,k}L_k}p_{U_k|L_kV_kX_{3,k}}. \tag{A46}$$

Appendix E

Proof of Theorem 5—More Capable Channel—Strictly Causal Case. *Proof of the Converse Part:* For $i \in \{1, 2, ..., n\}$, let V_i be defined as in (A18). Define Q to be an auxiliary random variable that is distributed uniformly on the set $\{1, 2, ..., n\}$, and let V, X_1, X_2, X_3 be defined as in (A19).

We bound R_1 as in (A20), we get

$$R_1 \leq H(X_1|V). \tag{A47}$$

Next, consider R_2

$$
\begin{aligned}
R_2 &= \frac{1}{n} H(W_2|W_1W_3) \\
&\leq \frac{1}{n} I(W_2; Y^n|W_1W_3) + \epsilon_n \\
&= \frac{1}{n} H(Y^n|W_1W_3) - \frac{1}{n} H(Y^n|W_1W_2W_3) + \epsilon_n \\
&= \frac{1}{n} \sum_{i=1}^{n} [H(Y_i|Y^{i-1}W_1W_3) - H(Y_i|Y^{i-1}W_1W_2W_3)] + \epsilon_n \\
&\overset{(a)}{=} \frac{1}{n} \sum_{i=1}^{n} [H(Y_i|Y^{i-1}W_1X_1^{i-1}X_{1i}W_3X_{3i}) - H(Y_i|Y^{i-1}W_1X_1^{i-1}X_{1i}W_2X_{2i}W_3X_{3i})] + \epsilon_n \\
&\overset{(b)}{\leq} \frac{1}{n} \sum_{i=1}^{n} [H(Y_i|X_1^{i-1}X_{1i}X_{3i}) - H(Y_i|Y^{i-1}W_1X_1^{i-1}X_{1i}W_2X_{2i}W_3X_{3i})] + \epsilon_n \\
&\overset{(c)}{=} \frac{1}{n} \sum_{i=1}^{n} [H(Y_i|X_1^{i-1}X_{1i}X_{3i}) - H(Y_i|X_1^{i-1}X_{1i}X_{2i}X_{3i})] + \epsilon_n \\
&= \frac{1}{n} \sum_{i=1}^{n} I(X_{2i}; Y_i|X_1^{i-1}X_{1i}X_{3i}) + \epsilon_n \\
&= I(X_{2Q}; Y_Q|V_QX_{1Q}X_{3Q}Q) + \epsilon_n \\
&= I(X_2; Y|VX_1X_3) + \epsilon_n,
\end{aligned} \tag{A48}
$$

where

(a) follows from the encoding relation in (2)–(4),
(b) follows since conditioning reduces entropy, and
(c) follows since $(W_1, W_2, W_3, X_1^{i-1}, Y^{i-1}) - (X_{1i}, X_{2i}, X_{3i}) - Y_i$ is a Markov chain.

Next, consider the sum-rate $R_1 + R_2$

$$
\begin{aligned}
R_1 + R_2 &= \frac{1}{n} H(W_1W_2|W_3) \\
&\leq \frac{1}{n} I(W_1W_2; Y^n|W_3) + \epsilon_n \\
&= \frac{1}{n} H(Y^n|W_3) - \frac{1}{n} H(Y^n|W_1W_2W_3) + \epsilon_n \\
&= \frac{1}{n} \sum_{i=1}^{n} [H(Y_i|Y^{i-1}W_3) - H(Y_i|Y^{i-1}W_1W_2W_3)] + \epsilon_n \\
&\overset{(a)}{=} \frac{1}{n} \sum_{i=1}^{n} [H(Y_i|Y^{i-1}W_3X_{3i}) - H(Y_i|Y^{i-1}W_1X_1^{i-1}X_{1i}W_2X_{2i}W_3X_{3i})] + \epsilon_n \\
&\overset{(b)}{\leq} \frac{1}{n} \sum_{i=1}^{n} [H(Y_i|X_{3i}) - H(Y_i|Y^{i-1}W_1X_1^{i-1}X_{1i}W_2X_{2i}W_3X_{3i})] + \epsilon_n \\
&\overset{(c)}{=} \frac{1}{n} \sum_{i=1}^{n} [H(Y_i|X_{3i}) - H(Y_i|X_{1i}X_{2i}X_{3i})] + \epsilon_n \\
&= \frac{1}{n} \sum_{i=1}^{n} I(X_{1i}X_{2i}; Y_i|X_{3i}) + \epsilon_n \\
&= I(X_{1Q}X_{2Q}; Y_Q|X_{3Q}Q) + \epsilon_n \\
&= I(X_1X_2; Y|X_3) + \epsilon_n,
\end{aligned}
$$

where the reasoning for steps (a)–(c) is as in (A48).

Finally, clearly

$$R_3 \le I(X_3;Z). \tag{A49}$$

Since in this case, the only auxiliary random variable used is V, defined the same as in (A19), and (27) is a special case of (10), it follows that V satisfies (27).

Proof of the Direct Part: It is easy to verify that the region in (9a)–(9c) contains the region in (25a)–(25d). To realize this, set $X_2 = U$ and $L = X_3$. Hence in (9b), (9c) we get $I(U;X_3|L = X_3) = 0$ and $I(U;Z|L = X_3) = 0$ and both equations coincide with (25b), (25c). The inequality (9a) remains as it is and (9d) becomes $R_3 \le I(X_3;Z)$ since $\min\{I(L;Y), I(L;Z)\} = I(X_3;Z)$ for $L = X_3$ in the *more-capable* case. Hence, since the p.m.f, in (27) is a special case of the probability mass function in (10), the region (25a)–(25d) is achievable thus concluding the proof of Theorem 5. □

Appendix F

Outline of the Proof of Theorem 7:

The achievability theorem is based on random coding scheme in addition to superposition coding, rate-splitting, and Gel'fand-Pinsker binning. However, since there is no cribbing involved, and Encoder 2 has full knowledge of W_1, there is no need of Block–Markov coding and Backward-Decoding, and the coding scheme becomes simpler than the one used in proving Theorem 1. In what follows, we sketch the main elements of the encoding and decoding procedures and provide an intuitive explanation for the proposed choices.

For a distribution $P_{LUX_1X_2X_3}$ satisfying (37), User 3 uses the same rate-splitting coding technique as in Appendix A. It encodes one part of W_3 by an inner codebook represented by L and the second part by an outer codebook represented by X_3, where the inner codebook can be decoded by both decoders. Now, User 1 may transmit at rate $R_1' = I(X_1;Y|L)$. The cognitive user, User 2, relying on the fact that L and X_1 were decoded by the main decoder, bins U against X_3, hence transmitting at rate $R_2' = I(U;Y|LX_1) - I(U;X_3|LX_1)$. Now the information sent by Encoder 2 at rate R_2' may be shared between the private message W_2 and the common message W_1 in such a manner that

$$\begin{aligned} R_2 &\le R_2' \\ R_1 + R_2 &\le R_1' + R_2' \end{aligned}$$

thus, establishing (36a)–(36c).

Appendix G

Outline of the proof of Theorem 8 :

Let the RVs $X_1, X_2, X_3, L, V, U, Q$ be defined as in (A18)–(A19) with the exception of $V_i \triangleq W_1$. We start by bounding R_3 as in (A27), thus getting

$$R_3 \le I(X_3;Z|L) + \min\{I(L;Y), I(L;Z)\}. \tag{A50}$$

Next, consider R_2

$$\begin{aligned} R_2 &= \frac{1}{n} H(W_2|W_1) \\ &\le \frac{1}{n} I(W_2;Y^n|W_1) + \epsilon_n \\ &= \frac{1}{n} H(Y^n|W_1) - \frac{1}{n} H(Y^n|W_1W_2) + \epsilon_n. \end{aligned}$$

Using the identities (A22) and (A31) we get

$$\begin{aligned}
R_2 &\le \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_iW_1) - H(Z_i|L_iW_1)] + \frac{1}{n}H(Z^n|W_1) - \frac{1}{n}H(Z^n|W_1W_2) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_iW_1W_2) - H(Z_i|L_iW_1W_2)] + \epsilon_n \\
&\overset{(a)}{=} \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_iW_1) - H(Z_i|L_iW_1)] - \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_iW_1W_2) - H(Z_i|L_iW_1W_2)] + \epsilon_n \\
&\overset{(b)}{=} \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_iW_1X_{1i}) - H(Y_i|L_iW_1X_{1i}W_2)] - \frac{1}{n}\sum_{i=1}^{n}[H(Z_i|L_iW_1) - H(Z_i|L_iW_1W_2)] + \epsilon_n \\
&= \frac{1}{n}\sum_{i=1}^{n}[I(W_2;Y_i|L_iW_1X_{1i}) - I(W_2;Z_i|L_iW_1)] + \epsilon_n \\
&= I(W_2;Y_Q|L_QW_1X_{1Q}Q) - I(W_2;Z_Q|L_QW_1Q) + \epsilon_n \\
&= I(U_Q;Y_Q|L_QV_QX_{1Q}Q) - I(U_Q;Z_Q|L_QV_QQ) + \epsilon_n \\
&= I(U;Y|LVX_1) - I(U;Z|LV) + \epsilon_n.
\end{aligned} \tag{A51}$$

Now, consider the sum-rate $R_1 + R_2$

$$\begin{aligned}
R_1 + R_2 &= \frac{1}{n}H(W_1W_2) \\
&\le \frac{1}{n}I(W_1W_2;Y^n) + \epsilon_n \\
&= \frac{1}{n}H(Y^n) - \frac{1}{n}H(Y^n|W_1W_2) + \epsilon_n.
\end{aligned}$$

Again, using the identities (A22) and (A31) we get

$$\begin{aligned}
R_1 + R_2 &\le \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_i) - H(Z_i|L_i)] + \epsilon_n + \frac{1}{n}H(Z^n) - \frac{1}{n}H(Z^n|W_1W_2) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_iW_1W_2) - H(Z_i|L_iW_1W_2)] \\
&\overset{(a)}{=} \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_i) - H(Z_i|L_i)] + \epsilon_n - \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_iW_1W_2) - H(Z_i|L_iW_1W_2)] \\
&\overset{(b)}{=} \frac{1}{n}\sum_{i=1}^{n}[H(Y_i|L_i) - H(Y_i|L_iW_1X_{1i}W_2)] - \frac{1}{n}\sum_{i=1}^{n}[H(Z_i|L_i) - H(Z_i|L_iW_1W_2)] + \epsilon_n \\
&= \frac{1}{n}\sum_{i=1}^{n}[I(W_1W_2X_{1i};Y_i|L_i) - I(W_1W_2;Z_i|L_i)] + \epsilon_n \\
&= I(W_1W_2X_{1Q};Y_Q|L_QQ) - I(W_1W_2;Z_Q|L_QQ) + \epsilon_n \\
&= I(V_QU_QX_{1Q};Y_Q|L_QQ) - I(V_QU_Q;Z_Q|L_QQ) + \epsilon_n \\
&= I(VUX_1;Y|L) - I(VU;Z|L) + \epsilon_n,
\end{aligned} \tag{A52}$$

where (a) follows since Z^n is independent of (W_1, W_2), (b) follows from the encoding relation in (2). Similarly to Theorem 3, it can easily be seen that auxiliary random variables defined here satisfy (37).

Appendix H

Proof of Theorem 9—Partial Cribbing MA-CZIC Inner Bound. We introduce the following coding scheme, based on the coding scheme of Appendix A. The difference from Appendix A is that Encoder 1 now uses rate-splitting in addition to block Markov superposition coding. Encoders 2 and 3 use the same coding scheme as in Appendix A. Since the analysis of the average probability of error is very similar to that of Appendix A, for the sake of brevity we omit it as well as similar parts to Appendix A which will not be repeated here. For a fixed distribution $P_V P_L P_{X_3|L} P_{X_1 Y_2|V} P_{UX_2|VLX_3}$ the coding schemes are as follows:

Encoder 3 and Decoder 3 Coding Scheme: Same as in Appendix A.

Encoder 1 Coding Scheme: We consider B Blocks, each of n symbols. A sequence of $B-1$ message pairs $(W_1^{(b)}, W_2^{(b)})$ for $b = 1, ..., B-1$, will be transmitted during B transmission blocks.

Encoder 1 Codebook generation: Encoder 1 splits its message W_1 into two independent parts $W_1 = (W_{1a}, W_{1b})$, with rates R_{1a} and R_{1b} accordingly. Generate $2^{nR_{1a}}$ codewords $\mathbf{v} = (v_1, ..., v_n)$, each with probability $\Pr(\mathbf{v}) = \prod_{i=1}^n P_V(v_i)$. These codewords constitute the inner codebook of Transmitter 1. Denote them as $\mathbf{v}(w_{0a})$, where $w_{0a} \in \{1, ..., 2^{nR_{1a}}\}$. For each codeword $\mathbf{v}(w_{0a})$ generate $2^{nR_{1a}}$ codewords $\mathbf{y}_2$, each with probability $\Pr(\mathbf{y}_2|\mathbf{v}(w_{0a})) = \prod_{i=1}^n P_{Y_2|V}(y_{2,i}|v_i(w_{0a}))$. These codewords, $\{\mathbf{x}_1\}$, constitute the outer codebook of Transmitter 1 associated with $\mathbf{v}(w_{0a})$. Denote them as $\mathbf{y}_2(w_{1a}, w_{0a})$ where w_{0a} is as before, representing the index of the codeword $\mathbf{v}(w_{0a})$ in the inner codebook and $w_{1a} \in \{1, ..., 2^{nR_{1a}}\}$ the index of the codeword $\mathbf{y}_2$ in the associated outer codebook. Finally, for each pair $(\mathbf{v}(w_{0a}), \mathbf{y}_2(w_{1a}, w_{0a}))$, generate $2^{R_{1b}}$ codewords $\mathbf{x}_1$ each with probability $\Pr(\mathbf{x}_1|\mathbf{v}(w_{0a}), \mathbf{y}_2(w_{1a}, w_{0a})) = \prod_{i=1}^n P_{X_1|VY_2}(x_{1,i}|v_i(w_{0a}), y_{2i}(w_{1a}, w_{0a}))$. Denote them as $\mathbf{x}_1(w_{1b}, w_{1a}, w_{0a})$.

Encoding Scheme of Encoder 1: Given $W_{1i}^{(b)} = w_1^{(b)} \in \{1, ..., 2^{nR_{1i}}\}$, where $i \in \{a, b\}$, for $b = 1, 2, ..., B$, we define $w_{0a}^{(b+1)} = w_{1a}^{(b)}$ for $b = 1, 2, ..., B-1$.

In block 1 Encoder 1 sends

$$\mathbf{x}_1^{(1)} = \mathbf{x}_1(w_{1b}^{(1)}, w_{1a}^{(1)}, 1),$$

in block $b = 2, 3, ..., B-1$ Encoder 1 sends

$$\mathbf{x}_1^{(b)} = \mathbf{x}_1(w_{1b}^{(b)}, w_{1a}^{(b)}, w_{0a}^{(b)})$$

and in block B Encoder 1 sends

$$\mathbf{x}_1^{(B)} = \mathbf{x}_1(1, 1, w_{0a}^{(B)})$$

Encoder 2 Coding Scheme: Same as in Appendix A, where the index w_0 is now replaced with w_{0a}, and x_1, unknown at Encoder 2, is replaced with y_2.

Decoding at the primary receiver (g_1): After receiving B blocks the decoder uses backward decoding starting from decoding block B moving on downward to block 1. In block B the receiver looks for $\hat{w}_{0a}^{(B)} = \hat{w}_{1a}^{(B-1)}$ such that

$$\left(\mathbf{v}(\hat{w}_{1a}^{(B-1)}), \mathbf{y}_2(1, \hat{w}_{1a}^{(B-1)}), \mathbf{x}_1(1, 1, \hat{w}_{1a}^{(B-1)}), \mathbf{u}(\hat{w}_{1a}^{(B-1)}, 1, w_{3a}^{(B)}, t), \mathbf{l}(w_{3a}^{(B)}), \mathbf{y}^{(B)}\right) \quad \in A_\epsilon(V, Y_2, X_1, U, L, Y)$$

for some $w_{3a}^{(B)}$, where $t = t(\hat{w}_1^{(B-1)}, 1, w_3^{(b)})$.

In block $b = 1, 2, ..., B-1$, assuming that a decoding was done backward down to (and including) block $b+1$, the receiver decoded $\hat{w}_{1a}^{(B-1)}, (\hat{w}_2^{(B-1)}, \hat{w}_{1b}^{(B-2)}, \hat{w}_{1a}^{(B-2)}), \ldots, (\hat{w}_2^{(b+1)}, \hat{w}_{1b}^{(b+1)}, \hat{w}_{1a}^{(b)})$. Then, to decode block b, the receiver looks for $(\hat{w}_2^{(b)}, \hat{w}_{1b}^{(b)}, \hat{w}_{1a}^{(b-1)})$ such that

$$\begin{aligned} &\left(\mathbf{v}(\hat{w}_{1a}^{(b-1)}), \mathbf{y}_2(\hat{w}_{1a}^{(b)}, \hat{w}_{1a}^{(b-1)}), \mathbf{x}_1(\hat{w}_{1b}^{(b)}, \hat{w}_{1a}^{(b)}, \hat{w}_{1a}^{(b-1)}), \mathbf{u}(\hat{w}_{1a}^{(b-1)}, \hat{w}_2^{(b)}, w_{3a}^{(b)}, t), \mathbf{l}(w_{3a}^{(b)}), \mathbf{y}^{(b)}\right) \\ &\in A_\epsilon(V, Y_2, X_1, U, L, Y) \end{aligned} \tag{A53}$$

for some $w_{3a}^{(b)}$, where $t = t(\hat{w}_{1a}^{(b-1)}, \hat{w}_2^{(b)}, w_3^{(b)})$.

Decoding at Encoder 2: To obtain cooperation, after block $b = 1, 2, ..., B-1$, Encoder 2 chooses $\tilde{w}_{1a}^{(b)}$ such that

$$\left(\mathbf{v}(\tilde{w}_{0a}^{(b)}), \mathbf{y}_2(\tilde{w}_{1a}^{(b)}, \tilde{w}_{0a}^{(b)}), \mathbf{y}_2^{(b)}\right) \in A_\epsilon(V, Y_2, Y_2)$$

where $\tilde{w_{0a}}^{(b)} = \tilde{w}_{1a}^{(b-1)}$ was determined at the end of block $b-1$ and $\tilde{w}_{0a}^{(1)} = 1$.

At each of the decoders, if a decoding step either fails to recover a unique index (or index pair) which satisfies the decoding rule, or there is more than one index (or index pair), then an index (or index pair) is chosen at random among the indices which satisfies the decoding rule. □

Appendix I

Proof of Theorem 11—Partial Cribbing MA-CZIC Outer Bound. Let the RVs X_1, X_2, X_3, L, U, Q be defined as in (A18)–(A19). In addition define V_i and T_i to be

$$V_i \triangleq Y_2^{i-1}, \quad T_i \triangleq X_1^{i-1} \tag{A54}$$

accordingly, define V, T, Y_2 as follows

$$V \triangleq (V_Q, Q), \quad Y_2 \triangleq (Y_{2Q}, Q), \quad T \triangleq (T_Q, Q). \tag{A55}$$

We start with an upper bound on R_1

$$\begin{aligned}
nR_1 &= H(W_1|W_2) \\
&\overset{(a)}{=} H(W_1 Y_2^n|W_2) \\
&= H(Y_2^n|W_2) + H(W_1|Y_2^n W_2) \\
&= H(Y_2^n|W_2) + I(W_1; Y^n|Y_2^n W_2) + n\epsilon_n \\
&= H(Y_2^n|W_2) + H(Y^n|Y_2^n W_2) + H(Y^n|Y_2^n W_1 W_2) + n\epsilon_n \\
&= \sum_{i=1}^n H(Y_{2i}|Y_2^{i-1} W_2) + H(Y^n|Y_2^n W_2) + H(Y^n|Y_2^n W_1 W_2) + n\epsilon_n \\
&\overset{(b)}{=} \sum_{i=1}^n [H(Y_{2i}|Y_2^{i-1} W_2) + H(Y_i|Y_2^n W_2 L_i) - H(Z_i|Y_2^n W_2 L_i) + H(Y_i|Y_2^n W_1 W_2 L_i) - H(Z_i|Y_2^n W_1 W_2 L_i)] \\
&\quad + H(Z^n|W_2 Y_2^n) - H(Z^n|W_1 W_2 Y_2^n) + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^n [H(Y_{2i}|Y_2^{i-1}) + H(Y_i|Y_2^n W_2 L_i) - H(Z_i|Y_2^n W_2 L_i) + H(Y_i|Y_2^n W_1 W_2 L_i) - H(Z_i|Y_2^n W_1 W_2 L_i)] + n\epsilon_n \\
&= \sum_{i=1}^n [H(Y_{2i}|Y_2^{i-1}) + I(W_1; Y_i|Y_2^n W_2 L_i) - I(W_1; Z_i|Y_2^n W_2 L_i)] + n\epsilon_n \\
&\overset{(d)}{=} \sum_{i=1}^n [H(Y_{2i}|Y_2^{i-1}) + I(X_{1i}; Y_i|Y_2^{i-1} Y_{2i} W_2 L_i) - I(X_1^{i-1}; Z_i|Y_2^{i-1} W_2 L_i)] + n\epsilon_n \\
&\leq \sum_{i=1}^n [H(Y_{2i}|Y_2^{i-1}) + I(X_{1i}; Y_i|Y_2^{i-1} Y_{2i} W_2 L_i)] + n\epsilon_n.
\end{aligned} \tag{A56}$$

where (a) follows from the encoding relation and the fact that Y_2 is a deterministic function of X_1. Step (b) follows from the identity (A22). Step (c) follows from the fact that conditioning decreases entropy and that Z^n is independent of the triplet (W_1, W_2, Y_2^n). Step (d) follows from the Markov chains $W_1 - (X_{1i}, Y_2^{i-1}, Y_2^i, W_2, L_i) - Y_i$ and $W_1 - (X_1^{i-1}, Y_2^{i-1}, W_2, L_i) - Z_i$.

Next, we bound R_3 as in Appendix C. We get

$$R_3 \leq I(X_3; Z|L) + \min\{I(L; Y), I(L; Z)\} + \epsilon_n. \tag{A57}$$

We continue to bound R_2 as in Appendix C. It is easy to see that the bound for the MA-CZIC with full strictly-causal cribbing must also bound the MA-CZIC with partial cribbing. Hence, we get

$$R_2 \leq \frac{1}{n}\sum_{i=1}^{n}[I(W_2;Y_i|L_iX_{1i}X_1^{i-1}) - I(W_2;Z_i|L_iX_1^{i-1})] + \epsilon_n. \tag{A58}$$

Finally, we consider the sum-rate $R_1 + R_2$. As in Appendix C, the bound for the MA-CZIC with full strictly-causal cribbing must also bound the MA-CZIC with partial cribbing. we get

$$R_1 + R_2 \leq \frac{1}{n}\sum_{i=1}^{n}[I(X_{1i}X_1^{i-1}W_2;Y_i|L_i) - I(X_1^{i-1}W_2;Z_i|L_i)] + \epsilon_n. \tag{A59}$$

A second bound on the sum-rate is obtained as follows

$$\begin{aligned}
&n(R_1+R_2)\\
&= H(W_1W_2)\\
&\overset{(e)}{=} H(W_1W_2Y_2^n)\\
&= H(Y_2^n) + H(W_1W_2|Y_2^n)\\
&= H(Y_2^n) + I(W_1W_2;Y^n|Y_2^n) + n\epsilon_n\\
&= H(Y_2^n) + H(Y^n|Y_2^n) + H(Y^n|Y_2^nW_1W_2) + n\epsilon_n\\
&= \sum_{i=1}^{n} H(Y_{2i}|Y_2^{i-1}) + H(Y^n|Y_2^n) + H(Y^n|Y_2^nW_1W_2) + n\epsilon_n\\
&\overset{(f)}{=} \sum_{i=1}^{n}[H(Y_{2i}|Y_2^{i-1}) + H(Y_i|Y_2^nL_i) - H(Z_i|Y_2^nL_i) + H(Y_i|Y_2^nW_1W_2L_i) - H(Z_i|Y_2^nW_1W_2L_i)]\\
&\quad + H(Z^n|Y_2^n) - H(Z^n|W_1W_2Y_2^n) + n\epsilon_n\\
&\overset{(g)}{=} \sum_{i=1}^{n}[H(Y_{2i}|Y_2^{i-1}) + H(Y_i|Y_2^nL_i) - H(Z_i|Y_2^nL_i) + H(Y_i|Y_2^nW_1W_2L_i) - H(Z_i|Y_2^nW_1W_2L_i)] + n\epsilon_n\\
&= \sum_{i=1}^{n}[H(Y_{2i}|Y_2^{i-1}) + I(W_1W_2;Y_i|Y_2^nL_i) - I(W_1W_2;Z_i|Y_2^nL_i)] + n\epsilon_n\\
&\overset{(h)}{=} \sum_{i=1}^{n}[H(Y_{2i}|Y_2^{i-1}) + I(X_{1i}X_1^{i-1}W_2;Y_i|Y_2^{i-1}Y_2^iL_i) - I(X_1^{i-1}W_2;Z_i|Y_2^{i-1}L_i)] + n\epsilon_n,
\end{aligned} \tag{A60}$$

where (e) follows from the encoding relation and the fact that Y_2 is a deterministic function of X_1. Step (f) follows from the identity (A22). Step (g) follows from the fact that Z^n is independent of the triplet (W_1, W_2, Y_2^n). Step (h) follows from the fact that conditioning reduces entropy and from the Markov chains

$$W_1 - (X_{1i}, X_1^{i-1}, W_2, Y_2^{i-1}, Y_2^i, L_i) - Y_i$$

$$W_1 - (X_1^{i-1}, W_2, Y_2^{i-1}, L_i) - Z_i$$

Now, similarly to Appendix C, we use (A56)–(A60) and the time-sharing RV Q to derive the outer bound. □

References

1. Ahlswede, R. Multi-way communication channels. In Proceedings of the Second International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, 2–8 September 1971; pp. 23–52.
2. Liao, H. Multiple Access Channels. Ph.D. Thesis, Department of Electrical Engineering, University of Hawaii, Honolulu, HI, USA, 1972.
3. Shannon, C.E. Two-way communication channels. In Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Statistical Laboratory of the University of California, Berkeley, CA, USA, 20 June–30 July 1960; University of California Press: Berkeley, CA, USA, 1961; Volume 1, pp. 611–644.

4. Ahlswede, R. The capacity region of a channel with two senders and two receivers. *Ann. Probab.* **1974**, *2*, 805–814.
5. Han, T.; Kobayashi, K. A new achievable rate region for the interference channel. *IEEE Trans. Inf. Theory* **1981**, *27*, 49–60.
6. Chong, H.F.; Motani, M.; Garg, H.; El Gamal, H. On the Han-Kobayashi Region for the Interference Channel. *IEEE Trans. Inf. Theory* **2008**, *54*, 3188–3195.
7. Carleial, A.B. Interference channels. *IEEE Trans. Inf. Theory* **1978**, *IT-24*, 60–70.
8. Sason, I. On achievable rate regions for the Gaussian interference channel. *IEEE Trans. Inf. Theory* **2004**, *50*, 1345–1356.
9. Kramer, G. Review of rate regions for interference channels. In Proceedings of the International Zurich Seminar on Communications, Zurich, Switzerland, 2–4 March 2006; pp. 152–165.
10. Etkin, R.; Tse, D.; Wang, H. Gaussian Interference Channel Capacity to within One Bit. *IEEE Trans. Inf. Theory* **2008**, *54*, 5534–5562.
11. Telatar, E.; Tse, D. Bounds on the capacity region of a class of interference channels. In Proceedings of the IEEE International Symposium Information Theory (ISIT), Nice, France, 24–29 June 2007.
12. Maric, I.; Goldsmith, A.; Kramer, G.; Shamai (Shitz), S. On the capacity of interference channels with one cooperating transmitter. *Eur. Trans. Telecommun.* **2008**, *19*, 329–495.
13. El Gamal, A.; Kim, Y.H. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2012, ISBN 1-107-00873-1.
14. Gamal, A.E.; Costa, M. The capacity region of a class of deterministic interference channels (Corresp.). *IEEE Trans. Inf. Theory* **1982**, *28*, 343–346.
15. Goldsmith, A.; Jafar, S.A.; Maric, I.; Srinivasa, S. Breaking Spectrum Gridlock with Cognitive Radios: An Information Theoretic Perspective. *Proc. IEEE* **2009**, *97*, 894–914.
16. Jovicic, A.; Viswanath, P. Cognitive Radio: An Information-Theoretic Perspective. *IEEE Trans. Inf. Theory* **2009**, *55*, 3945–3958.
17. Wang, B.; Liu, K.J.R. Advances in cognitive radio networks: A survey. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 5–23.
18. Zhao, Q.; Sadler, B.M. A Survey of Dynamic Spectrum Access. *IEEE Signal Process. Mag.* **2007**, *3*, 79–89.
19. Guzzon, E.; Benedetto, F.; Giunta, G. Performance Improvements of OFDM Signals Spectrum Sensing in Cognitive Radio. In Proceedings of the 2012 IEEE Vehicular Technology Conference (VTC Fall), Quebec City, QC, Canada, 3–6 September 2012; Volume 5, pp. 1–5.
20. Devroye, N.; Mitran, P.; Tarokh, V. Achievable rates in cognitive radio channels. *IEEE Trans. Inf. Theory* **2006**, *52*, 1813–1827.
21. Rini, S.; Kurniawan, E.; Goldsmith, A. Primary Rate-Splitting Achieves Capacity for the Gaussian Cognitive Interference Channel. *arXiv* **2012**, arXiv:1204.2083.
22. Rini, S.; Kurniawan, E.; Goldsmith, A. Combining superposition coding and binning achieves capacity for the Gaussian cognitive interference channel. In Proceedings of the 2012 IEEE Information Theory Workshop (ITW), Lausanne, Switzerland, 3–7 September 2012; pp. 227–231.
23. Wu, Z.; Vu, M. Partial Decode-Forward Binning for Full-Duplex Causal Cognitive Interference Channels. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Cambridge, MA, USA, 1–6 July 2012; pp. 1331–1335.
24. Duan, R.; Liang, Y. Bounds and Capacity Theorems for Cognitive Interference Channels with State. *IEEE Trans. Inf. Theory* **2015**, *61*, 280–304.
25. Rini, S.; Huppert, C. On the Capacity of the Cognitive Interference Channel with a Common Cognitive Message. *IEEE Trans. Inf. Theory* **2015**, *26*, 432–447.
26. Willems, F.; van der Meulen, E. The discrete memoryless multiple-access channel with cribbing encoders. *IEEE Trans. Inf. Theory* **1985**, *IT-31*, 313–327.
27. Bross, S.; Lapidoth, A. The state-dependent multiple-access channel with states available at a cribbing encoder. In Proceedings of the 2010 IEEE 26th Convention of Electrical and Electronics Engineers in Israel (IEEEI), Eilat, Israel, 17–20 November 2010; pp. 665–669.
28. Permuter, H.H.; Asnani, H. Multiple Access Channel with Partial and Controlled Cribbing Encoders. *IEEE Trans. Inf. Theory* **2013**, *59*, 2252–2266.

29. Somekh-Baruch, A.; Shamai (Shitz), S.; Verdú, S. Cooperative Multiple-Access Encoding With States Available at One Transmitter. *IEEE Trans. Inf. Theory* **2008**, *54*, 4448–4469.
30. Kopetz, T.; Permuter, H.H.; Shamai (Shitz), S. Multiple Access Channels With Combined Cooperation and Partial Cribbing. *IEEE Trans. Inf. Theory* **2016**, *62*, 825–848.
31. Kolte, R.; Özgür, A.; Permuter, H. Cooperative Binning for Semideterministic Channels. *IEEE Trans. Inf. Theory* **2016**, *62*, 1231–1249.
32. Permuter, H.H.; Shamai (Shitz), S.; Somekh-Baruch, A. Message and State Cooperation in Multiple Access Channels. *IEEE Trans. Inf. Theory* **2011**, *57*, 6379–6396.
33. Mokari, N.; Saeedi, H.; Navaie, K. Channel Coding Increases the Achievable Rate of the Cognitive Networks. *IEEE Commun. Lett.* **2013**, *17*, 495–498.
34. Passiatore, C.; Camarda, P. A P2P Resource Sharing Algorithm (P2P-RSA) for 802.22b Networks. In Proceedings of the 3rd International Conference on Context-Aware Systems and Applications, Dubai, UAE, 15–16 October 2014.
35. Costa, M.H.M. Writing on dirty paper. *IEEE Trans. Inf. Theory* **1983**, *29*, 439–441.
36. Liu, N.; Maric, I.; Goldsmith, A.; Shamai (Shitz), S. Capacity Bounds and Exact Results for the Cognitive Z-Interference Channel. *IEEE Trans. Inf. Theory* **2013**, *59*, 886–893.
37. Rini, S.; Tuninetti, D.; Devroye, N. Inner and Outer bounds for the Gaussian cognitive interference channel and new capacity results. *IEEE Trans. Inf. Theory* **2012**, *58*, 820–848.
38. Shimonovich, J.; Somekh-Baruch, A.; Shamai (Shitz), S. Cognitive cooperative communications on the Multiple Access Channel. In Proceedings of the 2013 IEEE Information Theory Workshop (ITW), Sevilla, Spain, 9–13 September 2013; pp. 1–5.
39. Shimonovich, J.; Somekh-Baruch, A.; Shamai (Shitz), S. Cognitive aspects in a Multiple Access Channel. In Proceedings of the 2012 IEEE 27th Convention of Electrical and Electronics Engineers in Eilat, Israel, 14–17 November 2012; pp. 1–3.
40. Cover, T.; Thomas, J. *Elements of Information Theory*, 1st ed.; Wiley: Hoboken, NJ, USA, 1991.
41. Gel'fand, S.I.; Pinsker, M.S. Coding for Channels with Random Parameters. *Probl. Contr. Inf. Theory* **1980**, *9*, 19–31.
42. Zaidi, A.; Kotagiri, S.; Laneman, J.; Vandendorpe, L. Cooperative Relaying with State Available Noncausally at the Relay. *IEEE Trans. Inf. Theory* **2010**, *56*, 2272–2298.
43. Csiszár, I.; Körner, J. Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory* **1978**, *24*, 339–348.
44. Kotagiri, S.; Laneman, J.N. Multiple Access Channels with State Information Known at Some Encoders. *EURASIP J. Wireless Commun. Netw.* **2008**, *2008*, doi:10.1155/2008/450680.