New Inner and Outer Bounds for the Memoryless Cognitive Interference Channel and Some New Capacity Results

Stefano Rini, Daniela Tuninetti, and Natasha Devroye

***Abstract*—The *cognitive interference channel* is a two-user interference channel in which one transmitter is non-causally provided with the message of the other transmitter. This channel model has been extensively studied in the past years and capacity results have been proved for certain classes of channels. This paper presents new inner and outer bounds for the capacity region of the cognitive interference channel, as well as new capacity results. Previously proposed outer bounds are expressed in terms of auxiliary random variables for which no cardinality constraint of their alphabet is known. Consequently, it is not possible to evaluate such outer bounds explicitly for a given channel. The outer bound derived in this work is based on an idea originally devised by Sato for channels without receiver cooperation and results in an outer bound that does not contain auxiliary random variables, thus allowing it to be more easily evaluated. The inner bound presented in this work—which includes rate splitting, superposition coding, a broadcast channel-like binning scheme and Gel'fand Pinsker coding—is the largest known to date and is explicitly shown to include all previously proposed achievable rate regions. The novel inner and outer bounds are shown to coincide in certain cases. In particular, capacity is proved for a class of channels in the so-called "*better cognitive decoding*" *regime*, which includes the regimes in which capacity was known. Finally, the capacity region of the *semi-deterministic* cognitive interference channel, in which the signal at the cognitive receiver is an arbitrary deterministic function of the channel inputs, is established.**

***Index Terms*—Achievable region, better cognitive decoding regime, capacity, cognitive channel, cognitive interference channel, inner bound, interference channel with degraded message sets, outer bound, semi-deterministic channel.**

I. Introduction

PRESENTLY, the frequency spectrum is allocated to different entities by dividing it into licensed lots. Licensed users have exclusive access to their licensed frequency lot or band and cannot interfere with the users in neighboring lots. The constant increase of wireless services has led to a situation where new services have a difficult time obtaining spectrum licenses and thus cannot be accommodated without discontinuing, or revoking, the licenses of others. This situation has been termed "spectrum gridlock" [4] and is viewed as one of the factors in preventing the emergence of new services and technologies by entities not already owning significant spectrum licenses.

In recent years, several strategies for overcoming this spectrum gridlock have been proposed [4]. In particular, *collaboration* among devices and adaptive transmission strategies are envisioned to overcome this spectrum gridlock. That is, smart devices may cooperate to *share* frequency, time and resources to communicate more efficiently and effectively. The role of information theory in this scenario is to determine the ultimate performance limits of such a collaborative network. Given the complexity of this task in its fullest generality, researchers have focussed on simple models with few idealized assumptions.

One of the most well studied and simplest collaborative models is the *cognitive interference channel*. This channel is similar to the classical two-user interference channel: two senders wish to send information to two receivers. Each transmitter has one intended receiver forming two transmitter-receiver pairs termed the *primary* and the *secondary/cognitive* pairs/users. Concurrent transmission creates undesired interference at the receivers. This channel model differs from the classical interference channel in the assumptions made about the ability of the transmitters to collaborate: collaboration among transmitters is modeled by the idealized assumption that the secondary/cognitive transmitter has full a-priori/non-causal knowledge of the primary message.[1]

A. Past Work

The cognitive interference channel was firstly posed from an information theoretic perspective in [7], where the channel was formally defined and the first achievable rate region was obtained, demonstrating that a cognitive interference channel, employing a form of asymmetric transmitter cooperation, could achieve larger rate regions than the classical interference channel. In [7], an outer bound for the Gaussian channel based on the broadcast channel was also presented. Another outer bound was derived in [6], together with the first capacity result for a class of channels with "very weak interference" in which (in Gaussian noise) treating interference at the primary receiver

Manuscript received March 23, 2010; revised October 06, 2010; accepted November 18, 2010. Date of current version June 22, 2011. Parts of this work were presented in [1]–[3]. The work of D. Tuninetti and S. Rini was supported in part by the National Science Foundation under Award 0643954. The work of N. Devroye was supported in part by the National Science Foundation under Awards 1017436 and 1053933. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the NSF.

The authors are with the Electrical and Computer Engineering Department, University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: srini2@uic.edu; danielat@uic.edu; devroye@uic.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Communicated by M. Franceschetti, Associate Editor for Communication Networks.

Digital Object Identifier 10.1109/TIT.2011.2146310

[1]This channel model has also been referred to as "unidirectional cooperation" [5] and transmission with "degraded message sets" [6].

as noise is optimal. The same achievable rate region of [6] was simultaneously derived in [8], where the authors further characterized the maximum rate the cognitive user can achieve under the constraint that the primary user's rate and mode of operation is the same as when the cognitive user is not present. Another capacity result was proved in [5] for channels with "very strong interference," where, without loss of optimality, both receivers can decode both messages and the cognitive channel reduces to a compound multiple access channel. The capacity is also known for the case where the cognitive user is required to decode both messages [9], both with and without secrecy/confidentiality constraints.

For the general memoryless cognitive interference channel the capacity region remains unknown. Tools such as rate-splitting, binning, and superposition coding have been used to derive different achievable rate regions. The authors of [10] proposed an achievable rate region that encompasses all the previously proposed inner bounds and derived a new outer bound using an argument originally devised for the broadcast channel in [11]. A further improvement of the inner bound of [10] is provided in [12] where the authors include a new feature in the transmission scheme allowing the cognitive transmitter to broadcast part of the message of the primary pair. This broadcast strategy is also encountered in the scheme derived in [13] for the broadcast channel with cognitive relays, which contains the cognitive interference channel as special case.

B. Main Contributions and Paper Organization

In this paper, we establish a series of new results for the general memoryless cognitive interference channel. Section II introduces the basic definitions and notation and summarizes known results including general inner bounds, outer bounds and capacity in the "very weak interference" [6], [8] and "very strong interference" [14] regimes.[2] Our contributions start in Section III and may be summarized as follows.

1) A new outer bound for the capacity region is presented in Section III: this outer bound is looser than some previously derived outer bounds but it does not include auxiliary random variables and thus it can be easily evaluated.
2) In Section IV, we present a new inner bound.
3) We show that the newly derived inner bound region encompasses all previously presented achievable rate regions in Section V.
4) We derive the capacity region of the cognitive interference channel in the "*better cognitive decoding*" regime in Section VI. This regime includes the "very weak interference" [6], [8] and the "very strong interference" [14] regimes and is thus the largest set of general memoryless channels for which capacity is known.
5) Section VII focuses on the *semi-deterministic* cognitive interference channel in which the output at the cognitive receiver is an arbitrary deterministic function of the channel inputs. We determine capacity for this channel by showing the achievability of the outer bound first derived in [6].
6) In Section VIII, we consider the *deterministic* cognitive interference channel: in this case both channel outputs are arbitrary deterministic functions of the inputs. This channel is a subclass of the semi-deterministic channel. For this channel we show the achievability of the outer bound proposed in Section III, thus showing that our outer bound can be tight.
7) The paper concludes with a couple of examples in Section IX which provide insight on the role of cognition. We consider two deterministic cognitive interference channels and show the achievability of the outer bound of Section III with zero-error transmission strategies over one channel use (i.e., in which case the capacity region coincides with the zero-error capacity region). The capacity achieving scheme in these channel models has the interesting feature that the non-causal message knowledge at the cognitive transmitter allows the primary user to achieve a rate that is higher than in the absence of the cognitive user, thus showing that cognition can benefit both the cognitive pair and the primary pair.

Section X concludes the paper. Some of the proofs are collected in the Appendix.

[2]We note here that we are not entirely consistent with past uses of the terms "strong/weak" interference. Our convention is to use "strong/weak" interference to denote regimes inspired by similar results for the classical interference channel. In particular, we denote by "strong/weak" interference a regime where an outer bound derived for a general memoryless channel can be simplified and/or tightened, and denote by "very strong/very weak" interference a regime in which additional conditions are imposed on top of the "strong/weak" interference conditions to show capacity. Therefore, the "very strong/very weak" interference regimes form subsets of the "strong/weak" interference regimes.

II. Channel Model and Known Results

A. Channel Model

A two-user InterFerence Channel (IFC) is a multi-terminal network with two senders and two receivers. Each transmitter i wishes to communicate a message W_i to receiver i, $i \in [1:2]$. In the classical IFC the two transmitters operate independently and have no knowledge of each others' message. Here we consider a variation of this set up assuming that transmitter 1 (the cognitive transmitter), in addition to its own message W_1, also knows the message W_2 of transmitter 2 (the primary transmitter). We refer to transmitter/receiver 1 as the *cognitive pair* and to transmitter/receiver 2 as the *primary pair*. This model, shown in Fig. 1, is termed the Cognitive InterFerence Channel (CIFC) and is an idealized model for unilateral transmitter cooperation. The Discrete Memoryless CIFC (DM-CIFC) is a CIFC with finite cardinality input and output alphabets and a memoryless channel described by the transition probability $p_{Y_1,Y_2|X_1,X_2}$. Achievable rate regions will be derived for DM-CIFC; these regions may be extended to continuous alphabets by standard arguments [15].

Transmitter i, $i \in [1:2]$, wishes to communicate a message W_i, uniformly distributed on $[1:2^{NR_i}]$, to receiver i in $N \in \mathbb{N}$ channel uses at rate $R_i \in \mathbb{R}_+$. The two messages are independent. A rate pair (R_1, R_2) is said to be achievable if there exists a sequence of encoding functions

$$X_1^N = X_1^N(W_1, W_2), \quad X_2^N = X_2^N(W_2)$$

and a sequence of decoding functions

$$\hat{W}_i = \hat{W}_i(Y_i^N), \quad i \in [1:2]$$

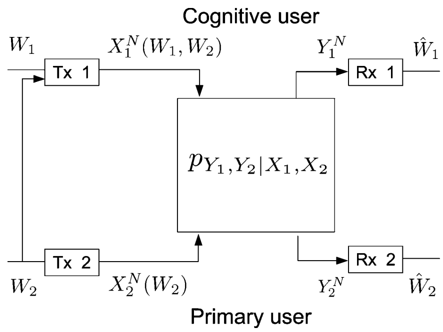

Fig. 1. General two-user Cognitive InterFerence Channel (CIFC) considered in this work.

such that

$$\lim_{N\longrightarrow\infty}\max_{i\in[1:2]}\Pr\left[\hat{W}_i\neq W_i\right]=0.$$

The capacity region is the convex closure of the region of all achievable (R_1,R_2)-pairs [15]. Since the receivers do not cooperate, the capacity of the CIFC only depends on the marginal conditional distributions $p_{Y_1|X_1,X_2}=\sum_{Y_2}p_{Y_1,Y_2|X_1,X_2}$ and $p_{Y_2|X_1,X_2}=\sum_{Y_1}p_{Y_1,Y_2|X_1,X_2}$.

We next summarize existing inner bounds, outer bounds and capacity results available for the general CIFC.

B. Known Inner Bounds

The rate regions in [12, Th. 2], [16, Th. 1], and [13, Th. 4.1] are achievable but it is not known whether any of them contains all the others. We will propose a new achievable rate region that provably includes them all. Note that the region in [12, Th. 2] is known to contain those in [10, Th. 1] and [17].

C. Known Outer Bounds

The tightest known outer bound for the capacity region of the general CIFC is given in [10, Th. 4]. This outer bound is derived using an argument originally devised in [11] for the "more capable" Broadcast Channel (BC) and contains three auxiliary Random Variables (RVs). Since we will not be using the outer bound in [10, Th. 4] in this work, we do not report it for sake of space.

The first outer bound for the general CIFC was obtained in [6, Th. 3.2]; the proof was also inspired by the converse of the BC [18] and contains one auxiliary RV.

Theorem 1. Outer Bound of [6, Th. 3.2]: If (R_1,R_2) lies in the capacity region of the CIFC then

$$R_1\leq I(X_1;Y_1|X_2), \tag{1a}$$
$$R_2\leq I(X_2,U;Y_2), \tag{1b}$$
$$R_1+R_2\leq I(X_2,U;Y_2)+I(X_1;Y_1|X_2,U) \tag{1c}$$

for some input distribution p_{U,X_1,X_2}.

The outer bound in Theorem 1 can be simplified in two instances called "weak interference" and "strong interference".

Corollary 2. "Weak Interference" Outer Bound of [6, Prop. 3.4]: When the following condition is satisfied:

$$I(U;Y_2|X_2)\leq I(U;Y_1|X_2)\quad\forall p_{U,X_1,X_2} \tag{2}$$

the outer bound in Theorem 1 can be expressed as

$$R_1\leq I(Y_1;X_1|U,X_2), \tag{3a}$$
$$R_2\leq I(U,X_2;Y_2) \tag{3b}$$

for some input distribution p_{U,X_1,X_2}.

In this work, the condition in (2) is referred to as the "weak interference" condition.

Corollary 3. "Strong Interference" Outer Bound of [14, Th. 5]: When the following condition is satisfied:

$$I(X_1;Y_1|X_2)\leq I(X_1;Y_2|X_2)\quad\forall p_{X_1,X_2} \tag{4}$$

the outer bound in Theorem 1 can be expressed as

$$R_1\leq I(Y_1;X_1|X_2), \tag{5a}$$
$$R_1+R_2\leq I(Y_2;X_1,X_2) \tag{5b}$$

for some input distribution p_{U,X_1,X_2}.

In this work, we refer to the condition in (4) as the "strong interference" condition.

D. Known Capacity Results

The outer bound of Theorem 1 may be shown to be achievable in a subset of the "weak interference" (2) and of the "strong interference" (4) regimes.

Theorem 4."Very Weak Interference" Capacity of [6, Th. 3.4] and [8, Th. 4.1] : The outer bound of Theorem 1, expressed as in Corollary 2, is the capacity region if for all p_{U,X_1,X_2}

$$I(U;Y_2|X_2)\leq I(U;Y_1|X_2), \tag{6a}$$
$$I(X_2;Y_2)\leq I(X_2;Y_1). \tag{6b}$$

In this work, we refer to the pair of conditions in (6) as "very weak interference". In this regime capacity is achieved by having the primary encoder (user 2) transmit as in a point-to-point channel and the secondary encoder (user 1) perform Gel'fand–Pinsker binning against the interference created by primary encoder (user 2).

Theorem 5."Very Strong Interference" Capacity of [14, Th. 5]: The outer bound of Theorem 1, expressed as in Corollary 3, is the capacity region if for all p_{X_1,X_2}

$$I(X_1;Y_1|X_2)\leq I(X_1;Y_2|X_2), \tag{7a}$$
$$I(Y_2;X_1,X_2)\leq I(Y_1;X_1,X_2). \tag{7b}$$

In this work, we refer to the pair of conditions in (7) as "very strong interference." In this regime, capacity is achieved by having both receivers decode both messages as in a compound Multiple Access Channel (MAC).

III. A New Outer Bound

The outer bound in Theorem 1 cannot be evaluated in general since it includes an auxiliary RV whose cardinality has not yet been bounded. In the following we thus propose a new outer bound, looser in general than Theorem 1, but without auxiliary RVs. This new bound can be easily evaluated and it is tight for some channels, as we shall show in the following sections.

Theorem 6. New Outer Bound: If (R_1, R_2) lies in the capacity region of the general CIFC then

$$R_1 \le I(Y_1; X_1|X_2), \quad (8a)$$
$$R_2 \le I(X_1, X_2; Y_2), \quad (8b)$$
$$R_1 + R_2 \le I(X_1, X_2; Y_2) + I(Y_1; X_1|Y_2', X_2) \quad (8c)$$

for some distribution p_{X_1,X_2} and where the joint conditional distribution $p_{Y_1,Y_2'|X_1,X_2}$ can be chosen so as to tighten the sum-rate bound as long as Y_2' has the same conditional marginal distribution as Y_2, i.e., $p_{Y_2'|X_1,X_2} = p_{Y_2|X_1,X_2}$.

Proof: The proof may be found in Appendix A. The idea behind this outer bound is to exploit the fact that the capacity region only depends on the conditional marginal distributions because the receivers do not cooperate [19]. ■

Remark 1: The outer bound in Theorem 6 contains the outer bound in Theorem 1. Indeed, for a fixed distribution p_{X_1,X_2}, the bounds on R_1 are the same ((1a) = (8a)). For the bound on R_2 we have $I(X_2, U; Y_2) \le I(X_2, X_1, U; Y_2) = I(X_1, X_2; Y_2)$ (which implies (1b) $\le$ (8b)) because of the Markov chain $U - X_1, X_2 - Y_1, Y_2$. For the sum-rate

$$\begin{aligned}(1c) &= I(X_2, U; Y_2) + I(X_1; Y_1|X_2, U)\\ &\overset{(a)}{\le} I(X_2, U; Y_2) + I(X_1; Y_1, Y_2'|X_2, U)\\ &= H(Y_2) - H(Y_2'|X_2, U, X_1)\\ &\quad + H(Y_1|X_2, U, Y_2') - H(Y_1|X_2, U, X_1, Y_2')\\ &\overset{(b)}{\le} I(Y_2; X_2, X_1) + I(Y_1; X_1|X_2, Y_2') = (8c).\end{aligned}$$

where (a) holds with equality if and only if $I(X_1; Y_2'|X_2, U, Y_1) = 0$ and (b) holds with equality if and only if $I(Y_1; U|X_2, Y_2') = 0$. We currently cannot relate these equality conditions to any specific class of CIFC.

Remark 2: The outer bound in Theorem 6 reduces to the "strong interference" outer bound in Corollary 3 when the condition in (4) holds; in fact the condition in (4) implies the condition in (5) as follows:

$$0 \le I(Y_1; X_1|Y_2', X_2) \le I(Y_2; X_1|Y_2', X_2)\ \forall p_{X_1,X_2,Y_2'}.$$

Now with $Y_2' = Y_2$ the above inequality implies $I(Y_1; X_1|Y_2, X_2) = 0$ thus yielding (8c) = (8b). Hence, with (8b) being redundant, the region in (8) coincides with the region in (5).

IV. A New Inner Bound

As the CIFC encompasses classical interference and broadcast channels, we expect to see a combination of their achievability proving techniques surface in any unified scheme for the CIFC. Our achievability scheme employs the following classical techniques.

- **Rate-splitting**: As in the Han and Kobayashi's scheme for the classical IFC [20], also employed in [7], [10], [17]. While rate-splitting may be useful in general, is not necessary in the "very weak" [6] and "very strong interference" [5] regimes of (6) and (7), respectively.
- **Superposition-coding**: Capacity achieving for "more capable" BC [18], in the CIFC the superposition of primary messages on top of cognitive ones, as in [10], [17], is known to be capacity achieving in "very strong interference."
- **Binning:** Gel'fand-Pinsker coding [21], often simply referred to as binning, allows a transmitter to "pre-cancel" (portions of) the interference known to be experienced at a receiver. Binning is also used by Marton [22] in deriving the largest known achievable rate region for the general memoryless BC.
- **Simultaneous decoding:** Useful in MACs, BCs, classical IFCs and used in all known achievable rate regions for the CIFC, a receiver jointly decodes its intended private and common messages and the common message from the interfering user.

We now present a new achievable rate region for the CIFC which generalizes all the known achievable rate regions presented in [6], [10], [12], [16], [17] and [13]. In Section V we will show that this achievable rate region, despite being built upon similar encoding schemes, generalizes and includes all other known achievable rate regions. The intuitive reason behind this inclusion lies in the structure of our encoder consisting of joint binning (rather than sequential as in some of the other regions), the full generality of our input distributions (lacking in some of the other known regions) and the presence of a broadcast channel like scheme at the cognitive transmitter (also noted in the region of [12]) and a slightly different rate-split than previous work. We note however that we do not claim strict containment of any of the previously proposed rate regions.

Theorem 7. New Inner Bound (Region $\mathcal{R}_{\rm RTD}$): A non-negative rate pair (R_1, R_2) such that

$$R_1 = R_{1c} + R_{1pb}, \quad (9a)$$
$$R_2 = R_{2c} + R_{2pa} + R_{2pb} \quad (9b)$$

is achievable for the CIFC if

$$(R_{1c}', R_{1pb}', R_{2pb}', R_{1c}, R_{1pb}, R_{2c}, R_{2pa}, R_{2pb}) \in \mathbb{R}_+^8$$

satisfies the inequalities in (11) some input distribution

$$p_{U_{1c},U_{2c},U_{2pa},U_{1pb},U_{2pb},X_1,X_2}. \quad (10)$$

Moreover the following rate-bound can be dropped (see (11a)–(11k) at the bottom of the next page).

- Equation (11d) when $R_{2c} = R_{2pa} = R_{2pb} = R_{2pb}' = 0$.
- Equation (11e) when $R_{2pa} = R_{2pb} = R_{2pb}' = 0$.
- Equation (11g) when $R_{2pb} = R_{2pb}' = 0$.
- Equation (11i) when $R_{1c} = R_{1c}' = R_{1pb} = R_{1pb}' = 0$,

since they correspond to the event that a non-intended common message is incorrectly decoded when no other intended message is incorrectly decoded.

Proof: The meaning of the RVs in Theorem 7 is as follows. Both transmitters perform superposition of two codewords: a common one (to be decoded at both decoders) and a private one (to be decoded at the intended decoder only). In particular:

- Rate R_1 is split into R_{1c} and R_{1pb} and conveyed through the RVs U_{1c} and U_{1pb}, respectively.

- Rate R_2 is split into R_{2c}, R_{2pa} and R_{2pb} and conveyed through the RVs U_{2c}, X_2 and U_{2pb}, respectively.
- U_{2c} is the common message of transmitter 2 with rate R_{2c}. The subscript "c" stands for "common."
- X_2 is the private message of transmitter 2 to be sent by transmitter 2 superimposed to U_{2c} and with rate R_{2pa}. The subscript "p" stands for "private" and the subscript "a" stands for "alone."
- U_{1c} is the common message of transmitter 1. It is superimposed to U_{2c} and—conditioned on U_{2c}—is binned against X_2.
- U_{1pb} and U_{2pb} are the private messages of transmitter 1 and transmitter 2, respectively and are sent by transmitter 1 only. They are binned against one another conditioned on U_{2c}, as in Marton's achievable rate region for the broadcast channel [22]. The subscript "b" stands for "broadcast."
- X_1 is finally superimposed to all the previous RVs and transmitted over the channel.

A graphical representation of the encoding scheme of Theorem 7 can be found in Fig. 2. Each box in the figure represents either an auxiliary RV or an input RV, which convey their appropriate messages. Primary and cognitive RVs are in blue squares and green rhomboids, respectively. A solid/dashed line from a RV U_i to a RV U_j indicates that the RV U_i is superposed onto/binned against the RV U_j. Given the non-causal message knowledge at the cognitive transmitter, the cognitive RVs can be binned against primary RVs but not vice-versa. Furthermore, a RV may not be binned against a RV over which it is superposed. In the achievable scheme of Theorem 7, X_1 is obtained as a function of all other RVs and it is not indicated in Fig. 2.

Rate Splitting: Let w_1 and w_2 be two independent RVs uniformly distributed on $[1:2^{nR_1}]$ and $[1:2^{nR_2}]$, respectively. Consider splitting the messages, as follows:

$$W_1 = (W_{1c}, W_{1pb}), \quad W_2 = (W_{2c}, W_{2pb}, W_{2pa})$$

where the (sub)messages w_j are independent and uniformly distributed on $[1:2^{nR_j}]$, $j \in \{1c, 2c, 1pb, 2pb, 2pa\}$, so that the rates satisfy (9).

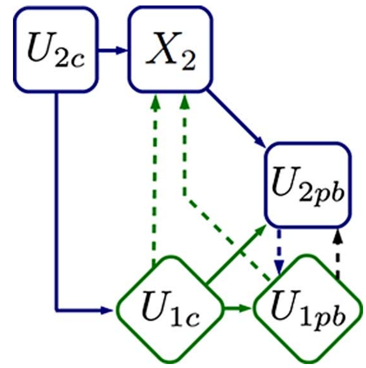

Fig. 2. Codebook generation for the encoding scheme in Th. 7. The RVs carrying a primary message are placed is blue squares while the RVs carrying a cognitive message are in green rhomboids. Lines connecting the different RVs specify encoding operations: solid lines indicate superposition coding while dashed lines indicate Gel'fand–Pinsker binning. The RVs carrying a private message, U_{1pb}, X_2, U_{2pb}, are superimposed onto the RVs carrying a common message, U_{1c}, U_{2c}. Similarly, the RVs carrying a cognitive message, U_{1c}, U_{1pb}, are superimposed onto the RV carrying primary common message U_{2c} and binned against the primary private messages X_2, U_{2pb}. Finally the cognitive private RV U_{1pb} and the primary private RV U_{2pb} are binned against each other as in the Marton's scheme for the broadcast channel.

Codebook Generation: Consider a distribution in (10). The codebooks are generated as follows:

- Select uniformly at random $2^{NR_{2c}}$ length-N sequences $U_{2c}^N(w_{2c})$, $w_{2c} \in [1:2^{NR_{2c}}]$, from the typical set $T_\epsilon^N(p_{U_{2c}})$.
- For every $w_{2c} \in [1:2^{NR_{2c}}]$, select uniformly at random $2^{NR_{2pa}}$ length-N sequences $X_2^N(w_{2c}, w_{2pa})$, $w_{2pa} \in [1:2^{NR_{2pa}}]$, from the typical set $T_\epsilon^N(p_{X_2,U_{2c}}|U_{2c}^N(w_{2c}))$.
- For every $w_{2c} \in [1:2^{NR_{2c}}]$, select uniformly at random $2^{N(R_{1c}+R'_{1c})}$ length-N sequences $U_{1c}^N(w_{2c}, w_{1c}, b_0)$, $w_{1c} \in [1:2^{NR_{1c}}]$ and $b_0 \in [1:2^{NR'_{1c}}]$, from the typical set $T_\epsilon^N(p_{U_{1c}U_{2c}}|U_{2c}^N(w_{2c}))$
- For every $w_{2c} \in [1:2^{NR_{2c}}]$, $w_{2pa} \in [1:2^{NR_{2pa}}]$, $w_{1c} \in [1:2^{NR_{1c}}]$ and $b_0 \in [1:2^{NR'_{1c}}]$, select uniformly at random $2^{N(R_{2pb}+R'_{2pb})}$ length-N sequences $U_{2pb}^N(w_{2c}, w_{2pa}, w_{1c}, b_0, w_{2pb}, b_2)$, $w_{2pb} \in$

$$
\begin{aligned}
R'_{1c} &= I(U_{1c}; X_2|U_{2c}) && \text{(11a)}\\
R'_{1c} + R'_{1pb} &\geq I(U_{1pb}; X_2|U_{1c}, U_{2c}) + I(U_{1c}; X_2|U_{2c}) && \text{(11b)}\\
R'_{1c} + R'_{1pb} + R'_{2pb} &\geq I(U_{1pb}; X_2, U_{2pb}|U_{1c}, U_{2c}) + I(U_{1c}; X_2|U_{2c}) && \text{(11c)}\\
R_{2c} + R_{2pa} + (R_{1c} + R'_{1c}) + (R_{2pb} + R'_{2pb}) &\leq I(Y_2; U_{2pb}, U_{1c}, X_2, U_{2c}) + I(U_{1c}; X_2|U_{2c}) && \text{(11d)}\\
R_{2pa} + (R_{1c} + R'_{1c}) + (R_{2pb} + R'_{2pb}) &\leq I(Y_2; U_{2pb}, U_{1c}, X_2|U_{2c}) + I(U_{1c}; X_2|U_{2c}) && \text{(11e)}\\
R_{2pa} + (R_{2pb} + R'_{2pb}) &\leq I(Y_2; U_{2pb}, X_2|U_{1c}, U_{2c}) + I(U_{1c}; X_2|U_{2c}) && \text{(11f)}\\
(R_{1c} + R'_{1c}) + (R_{2pb} + R'_{2pb}) &\leq I(Y_2; U_{2pb}, U_{1c}|X_2, U_{2c}) + I(U_{1c}; X_2|U_{2c}) && \text{(11g)}\\
(R_{2pb} + R'_{2pb}) &\leq I(Y_2; U_{2pb}|U_{1c}, X_2, U_{2c}) && \text{(11h)}\\
R_{2c} + (R_{1c} + R'_{1c}) + (R_{1pb} + R'_{1pb}) &\leq I(Y_1; U_{1pb}, U_{1c}, U_{2c}) && \text{(11i)}\\
(R_{1c} + R'_{1c}) + (R_{1pb} + R'_{1pb}) &\leq I(Y_1; U_{1pb}, U_{1c}|U_{2c}) && \text{(11j)}\\
(R_{1pb} + R'_{1pb}) &\leq I(Y_1; U_{1pb}|U_{1c}, U_{2c}). && \text{(11k)}
\end{aligned}
$$

$[1 : 2^{NR_{2pb}}]$ and $b_2 \in [1 : 2^{NR'_{2pb}}]$, from the typical set $T_\epsilon^N(p_{U_{2pb},U_{2c},U_{1c},X_2}|U_{2c}^N(w_{2c}), X_2^N(w_{2c}, w_{2pa}), U_{1c}^N(w_{2c}, w_{1c}, b_0))$.

- For every $w_{2c} \in [1 : 2^{NR_{2c}}]$, $w_{1c} \in [1 : 2^{NR_{1c}}]$ and $b_0 \in [1 : 2^{NR'_{1c}}]$, select uniformly at random $2^{N(R_{1pb}+R'_{1pb})}$ length-N sequences $U_{1pb}^N(w_{2c}, w_{1c}, b_0, w_{1pb}, b_1)$, $w_{1pb} \in [1 : 2^{NR_{1pb}}]$ and $b_1 \in [1 : 2^{NR'_{1pb}}]$, from the typical set $T_\epsilon^N(p_{U_{1pb},U_{2c},U_{1c}}|U_{2c}^N(w_{2c}), U_{1c}^N(w_{2c}, w_{1c}, b_0))$.
- For every $w_{2c} \in [1 : 2^{NR_{2c}}]$, $w_{2pa} \in [1 : 2^{NR_{2pa}}]$, $w_{1c} \in [1 : 2^{NR_{1c}}]$, $b_0 \in [1 : 2^{NR'_{1c}}]$, $w_{1pb} \in [1 : 2^{NR_{1pb}}]$, $b_1 \in [1 : 2^{NR'_{1pb}}]$, $w_{2pb} \in [1 : 2^{NR_{2pb}}]$, $b_2 \in [1 : 2^{NR'_{2pb}}]$, select a length-$N$ channel input $X_1^N(w_{2pa}, w_{2c}, w_{1c}, b_0, w_{1pb}, b_1, w_{2pb}, b_2)$ from the typical set $T_\epsilon^N(p_{X_1,U_{2c},U_{1c},X_2,U_{2pb},U_{1pb}}|U_{2c}^N(w_{2c}), X_2^N(w_{2c}, w_{2pa}), U_{1c}^N(w_{2c}, w_{1c}, b_0), U_{2pb}^N(w_{2c}, w_{2pa}, w_{1c}, b_0, w_{2pb}, b_2), U_{1pb}^N(w_{2c}, w_{1c}, b_0, w_{1pb}, b_1))$.

Encoding: Given $w_2 = (w_{2c}, w_{2pb}, w_{2pa})$, encoder 2 sends the codeword $X_2^N(w_{2c}, w_{2pa})$ (notice that encoder 2 does not send w_{2pb}).

Given $w_2 = (w_{2c}, w_{2pb}, w_{2pa})$ and $w_1 = (w_{1c}, w_{1pb})$, encoder 1 looks for a triplet (b_0, b_1, b_2) such that

$$\begin{aligned}&\big(U_{2c}^N(w_{2c}), X_2^N(w_{2c}, w_{2pa}), U_{1c}^N(w_{2c}, w_{1c}, b_0),\\ &U_{1pb}^N(w_{2c}, w_{1c}, b_0, w_{1pb}, b_1), U_{2pb}^N(w_{2c}, w_{1c}, b_0, w_{2pb}, b_2)\big)\\ &\in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}}).\end{aligned}$$

If more than one such triplet exists, it picks one uniformly at random from the found ones. If no such triplet exists, it sets $(b_0, b_1, b_2) = (1,1,1)$; in this case we say that an encoding error occurred. For the selected (b_0, b_1, b_2), encoder 1 sends $X_1^N(w_{2pa}, w_{2c}, w_{1c}, b_0, w_{1pb}, b_1, w_{2pb}, b_2)$.

Decoding: Decoder 2 looks for a unique tuple $(w_{2c}, w_{2pa}, w_{2pb})$ and some (w_{1c}, b_0, b_2) such that

$$\begin{aligned}&(U_{2c}^N(w_{2c}), X_2^N(w_{2c}, w_{2pa}), U_{1c}^N(w_{2c}, w_{1c}, b_0),\\ &U_{2pb}^N(w_{2c}, w_{1c}, b_0, w_{2pb}, b_2), Y_2^N) \in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{2pb},Y_2}).\end{aligned}$$

If none or more than one such triplet $(w_{2c}, w_{2pa}, w_{2pb})$ exist, decoder 2 sets $(w_{2c}, w_{2pa}, w_{2pb}) = (1,1,1)$; in this case we say that a decoding error occurred.

Decoder 1 looks for a unique pair (w_{1c}, w_{1pb}) and some (w_{2c}, b_0, b_1) such that

$$\begin{aligned}&(U_{2c}^N(w_{2c}), U_{1c}^N(w_{2c}, w_{1c}, b_0), U_{1pb}^N(w_{2c}, w_{1c}, b_0, w_{1pb}, b_1), Y_1^N)\\ &\in T_\epsilon^N(p_{U_{2c},U_{1c},U_{1pb},Y_1}).\end{aligned}$$

If none or more than one such pair (w_{1c}, w_{1pb}) exist, decoder 1 sets $(w_{1c}, w_{1pb}) = (1,1)$; in this case we say that a decoding error occurred.

Error Analysis: The detailed error analysis is found in Appendix B. In particular: the probability of encoding error goes to zero if conditions (11a)–(11c) hold; the probability of error at decoder 2 goes to zero if conditions (11d)–(11h) hold; and the probability of error at decoder 1 goes to zero if conditions (11i)–(11k) hold. ∎

Remark 3 (Two Step Binning): It is also possible to perform binning at encoder 1 in a sequential manner, similarly to [23], as follows. First, First, U_{1c} is binned against X_2 conditioned on U_{2c}; then, U_{1pb} and U_{2pb} are binned against each other conditioned on (U_{1c}, U_{2c}, X_2). With respect to the encoding operation of Theorem 7, this affects the achievable rate region as follows.

Given the message $w_2 = (w_{2c}, w_{2pb}, w_{2pa})$ and the message $w_1 = (w_{1c}, w_{1pb})$, encoder 1 looks for a b_0 such that

$$\begin{aligned}&(U_{2c}^N(w_{2c}), X_2^N(w_{2c}, w_{2pa}), U_{1c}^N(w_{2c}, w_{1c}, b_0))\\ &\in T_\epsilon^N(p_{U_{2c},X_2,U_{1c}}).\end{aligned}$$

If more than one such b_0 exists, it picks one uniformly at random. If no such b_0 exists, it sets $b_0 = 1$; in this case an error occurred. For the selected b_0, encoder 1 looks for a pair (b_1, b_2) such that

$$\begin{aligned}&(U_{2c}^N(w_{2c}), X_2^N(w_{2c}, w_{2pa}), U_{1c}^N(w_{2c}, w_{1c}, b_0),\\ &U_{1pb}^N(w_{2c}, w_{1c}, b_0, w_{1pb}, b_1), U_{2pb}^N(w_{2c}, w_{1c}, b_0, w_{2pb}, b_2))\\ &\in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}}).\end{aligned}$$

If more than one such (b_1, b_2) exists, it picks one uniformly at random from the found ones. If no such (b_1, b_2) exists, it sets $(b_1, b_2) = (1,1)$; in this case an error occurred. For the selected (b_0, b_1, b_2), encoder 1 sends $X_1^N(w_{2pa}, w_{2c}, w_{1c}, b_0, w_{1pb}, b_1, w_{2pb}, b_2)$.

Lemma 8: This two step encoding procedure is successful with high probability if

$$R'_{1c} \geq I(U_{1c}; X_2|U_{2c}), \tag{12a}$$
$$R'_{1pb} \geq I(U_{1pb}; X_2|U_{2c}, U_{1c}), \tag{12b}$$
$$R'_{1pb} + R'_{2pb} \geq I(U_{1pb}; X_2, U_{2pb}|U_{2c}, U_{1c}). \tag{12c}$$

Proof: The proof is found in Appendix C. ∎

From the Fourier–Motzkin elimination [24] of the region in (11), it is possible to conclude that the binning rate R_{1c} in (11a) may be taken to satisfy the constraint in (11a) with equality without loss of generality. This implies that the two step binning in lemma 8 has the same performance as joint binning in Theorem 7, i.e., by setting (12a) to hold with equality, which may be done without loss of generality, the joint and the two-step binning rate bounds are equivalent.

V. Comparison of $\mathcal{R}_{\mathrm{RTD}}$ With Existing Achievable Rate Regions

Theorem 9. The Region $\mathcal{R}_{\mathrm{RTD}}$ is the Largest Known Achievable Region: The region in Theorem 7 contains all known achievable rate regions for the CIFC. In particular, showing inclusion of the rate regions [12, Th. 2], [16, Th. 1] and [13, Th. 4.1] is sufficient to demonstrate the largest known CIFC region, since the region of [12, Th. 2] is shown to contain those of [10, Th. 1] and [17].

The proof of Theorem 9 is presented in the following subsections.

A. *Devroye et al.'s Region [16, Th. 1]*

In Appendix C we show that the region of [16, Th. 1], indicated as $\mathcal{R}_{\rm DMT}$, is contained in our new region $\mathcal{R}_{\rm RTD}$. In the proof:

- We make a correspondence between the random variables and corresponding rates of $\mathcal{R}_{\rm DMT}$ and $\mathcal{R}_{\rm RTD}$.
- We define new regions $\mathcal{R}_{\rm DMT} \subseteq \mathcal{R}_{\rm DMT}^{\rm out}$ and $\mathcal{R}_{\rm RTD}^{\rm in} \subseteq \mathcal{R}_{\rm RTD}$ which are easier to compare: they have identical input distribution decompositions and similar rate equations.
- For any fixed input distribution, we make an equation-by-equation comparison, which leads to $\mathcal{R}_{\rm DMT}^{\rm out} \subseteq \mathcal{R}_{\rm RTD}^{\rm in}$ and thus $\mathcal{R}_{\rm DMT} \subseteq \mathcal{R}_{\rm RTD}$.

B. *Cao and Chen's Region [12, Th. 2]*

The region in [12, Th. 2] uses a similar encoding structure as that of $\mathcal{R}_{\rm RTD}$ with two exceptions.

1) The binning in [12, Th. 2] is done sequentially rather than jointly as in $\mathcal{R}_{\rm RTD}$, leading to binning constraints [12, Th. 2, eq. (42)–(44)] as opposed to (11a)–(11c) in Theorem 7. Notable is that both schemes have adopted a Marton-like binning scheme at the cognitive transmitter, as first introduced in the context of the CIFC in [12].
2) The primary message is split into two parts in [12, Th. 2] (i.e., $R_1 = R_{11} + R_{10}$, note the reversal of indices), while we explicitly split the primary message into three parts (i.e., $R_2 = R_{2c} + R_{2pa} + R_{2pb}$).

In Appendix E we show that the region of [12, Th. 2], denoted as $\mathcal{R}_{\rm CC}$, satisfies $\mathcal{R}_{\rm CC} \subseteq \mathcal{R}_{\rm RTD}$ in two steps.

- We first show that we may without loss of generality set $U_{11} = \emptyset$ in [12, Th. 2].
- We next make a correspondence between a subset of our RVs and those of $\mathcal{R}_{\rm CC}$ showing that the region in [12, Th. 2] is a special case of our region in Theorem 7.

We also note that the region of [25], used to prove capacity for the cognitive Z-IFC when the interference-free component is noiseless, is a special case of the region in [12] and is thus also contained in our achievable region.

C. *Jiang et al.'s Region [13, Th. 4.1]*

The scheme in [13, Th. 4.1], originally designed for the general broadcast channel with cognitive relays (or interference channel with a cognitive relay [26]) that subsumes the CIFC, may give a achievable region for the CIFC by setting certain channel inputs to be empty sets. The scheme in [13, Th. 4.1] also incorporates a broadcasting strategy as in our achievable region through (U_{1pb}, U_{2pb}). However, the common codewords are created independently instead of having the common codeword of transmitter 1 superposed to the common codeword of transmitter 2. The former choice introduces more rate constraints than the latter and allows us to show inclusion in $\mathcal{R}_{\rm RTD}$ after equating random variables. The proof of the containment of the achievable rate region of [13, Th. 4.1] in $\mathcal{R}_{\rm RTD}$ is found in Appendix F.

VI. New Capacity Results for the CIFC

We now look at the expression of the outer bound in [6, Th. 3.2] (here in Theorem 1) to gain insight into potentially capacity achieving schemes. In particular, we look at the expression of the corner points of the outer bound region for a fixed distribution p_{U,X_1,X_2} and try to interpret the auxiliary RV as private or common messages to be matched to one of the RVs in the achievable scheme in Theorem 7. By doing so, we will show capacity for a class of channels in what we term the "*better cognitive decoding*" regime, which contains the "very strong" (see Theorem 5) and the "very weak" (see Theorem 4) interference regimes for which capacity was previously known. Thus, the "*better cognitive decoding*" regime corresponds to the largest class of general CIFC for which capacity is currently known.

The outer bound region of Theorem 1 [6, Th. 3.2] has at most two corner points where both the R_1-coordinate and the R_2-coordinate are nonzero:

$$\left(R_1^{\rm out(a)}, R_2^{\rm out(a)}\right) = \left(I(Y_1; X_1|U, X_2), I(Y_2; U, X_2)\right) \quad (13)$$

and

$$\begin{aligned}&\left(R_1^{\rm out(b)}, R_2^{\rm out(b)}\right)\\ &= \left(I(Y_1; X_1|U, X_2) + I(Y_2; U, X_2) - \Delta, \Delta\right)\end{aligned} \quad (14)$$

for $\Delta \triangleq [I(Y_2; U, X_2) - I(Y_1; U|X_2)]^+$, since the largest possible R_1 is

$$\begin{aligned}R_1^{\rm out(b)} &= \min\{I(Y_1; X_1|X_2),\\ &\qquad I(Y_2; U, X_2) + I(Y_1; X_1|U, X_2)\}\\ &= I(Y_1; X_1|U, X_2) + I(Y_2; U, X_2) - \Delta\end{aligned}$$

which results in an R_2:

$$\begin{aligned}R_2^{\rm out(b)} &= \min\{I(Y_2; U, X_2),\\ &\qquad I(Y_2; U, X_2) + I(Y_1; X_1|U, X_2) - R_1^{\rm out(b)}\}\\ &= \Delta.\end{aligned}$$

Proving the achievability of both these corner points for any p_{U,X_1,X_2} shows capacity by a simple time sharing argument. We can now look at the corner point expression and try to draw some intuition on the achievable schemes that can possibly achieve these rates.

- For the corner point $(R_1^{\rm out(a)}, R_2^{\rm out(a)})$ in (13) we can interpret (U, X_2) as a common message from encoder 2 striped out at decoder 1 before decoding the private message from encoder 1 in X_1.
- The corner point $(R_1^{\rm out(b)}, R_2^{\rm out(b)})$ in (14) has two possible expressions:
  1) If $I(Y_1; U|X_2) < I(Y_2; U, X_2)$, i.e., $\Delta > 0$:
  $$\begin{aligned}&(R_1^{\rm out(b)'}, R_2^{\rm out(b)'})\\ &= (I(Y_1; X_1, U|X_2), I(Y_2; U, X_2) - I(Y_1; U|X_2)),\end{aligned}$$
  which suggests that X_2 is the common primary message and U and X_1 are, respectively, the cognitive common and private message.
  2) If $I(Y_1; U|X_2) \geq I(Y_2; U, X_2)$, i.e., $\Delta = 0$:
  $$\begin{aligned}&(R_1^{\rm out(b)''}, R_2^{\rm out(b)''})\\ &= (I(Y_2; U, X_2) + I(Y_1; X_1|X_2, U), 0).\end{aligned}$$

In this case, since

$$I(Y_2;U,X_2)+I(Y_1;X_1|X_2,U)\leq I(Y_1;X_1|X_2),$$

the rate pint $(R_1^{\mathrm{out(b)}''}, R_2^{\mathrm{out(b)}''})$ is dominated by $(I(Y_1;X_1|X_2),0)$, which is always achievable. Hence, to show capacity we do not need to consider the case $\Delta=0$.

Guided by these observations, we consider a scheme that has only the components U_{2c}, U_{1c} and U_{1pb} in Theorem 7. That is, the primary message w_2 is common and the cognitive message w_1 is split into a private and a common message. Note that this proposed scheme coincides with that of [27], which achieves capacity if the cognitive receiver is required to decode both messages (with and without the secrecy constraint); for this reason we term the regime where the scheme with only U_{2c}, U_{1c} and U_{1pb} in Theorem 7 achieves capacity the "*better cognitive decoding*" regime. This also corresponds to the achievable schemes in [27] and in [9].

Theorem 10. New Capacity Result for the "Better Cognitive Decoding" Regime: The outer bound of Theorem 1 is the capacity region if for all $p_{X_1,X_2,U}$

$$I(Y_1;X_2,U)\geq I(Y_2;X_2,U). \tag{15}$$

Moreover, the "better cognitive decoding" condition in (15) includes the "very weak interference" condition in (6) and the "very strong interference" condition in (7).

Proof: Consider the achievable rate region in Theorem 7 with

$$X_1=U_{1pb}, U_{1c}=U, \quad X_2=U_{2c}=U_{2pb},$$

and $R_{2pa}=R_{2pb}=R'_{1c}=R'_{1pb}=R'_{2pb}=0$. In the resulting scheme, the message from transmitter 2 to receiver 2 is all common while the message from transmitter 1 to receiver 1 is split into common and private parts. The achievable rate region in Theorem 7 reduces to

$$R_2+R_{1c}\leq I(Y_2;U,X_2), \tag{16a}$$
$$R_2+R_{1c}+R_{1pb}\leq I(Y_1;U,X_2,X_1), \tag{16b}$$
$$R_{1c}+R_{1pb}\leq I(Y_1;U,X_1|X_2), \tag{16c}$$
$$R_{1pb}\leq I(Y_1;X_1|X_2,U). \tag{16d}$$

After Fourier-Motzkin elimination [24] the region in (16) becomes

$$R_1\leq I(Y_1;U,X_1|X_2), \tag{17a}$$
$$R_2\leq I(Y_2;U,X_2), \tag{17b}$$
$$R_1+R_2\leq I(Y_2;U,X_2)+I(Y_1;X_1|X_2,U), \tag{17c}$$
$$R_1+R_2\leq I(Y_1;X_2,U,X_1). \tag{17d}$$

We see that (1a)=(17a), (1b)=(17b), (1c)=(17c), and (17d) is redundant because of the "better cognitive decoding" condition in (15).

Moreover, the "better cognitive decoding" condition in (15) is looser than both the "very weak interference" and the "very strong interference" conditions in (6) and (7), respectively, because by summing the two equations in (6) we obtain

$$\begin{aligned}&I(U;Y_1|X_2)+I(X_2;Y_1)\geq I(U;Y_2|X_2)+I(X_2;Y_2)\\ &\iff I(Y_1;U,X_2)\geq I(Y_2;U,X_2)\iff \text{eq.}(15),\end{aligned}$$

and by summing the two equations in (7) we obtain

$$\begin{aligned}&I(Y_1;X_1,X_2)+I(X_1;Y_2|X_2)\geq I(Y_2;X_1,X_2)+I(X_1;Y_1|X_2)\\ &\iff I(Y_1;X_2)\geq I(Y_2;X_2)\iff \text{eq.}(15)\text{ for } U=\emptyset.\end{aligned}$$

Since both (6) and (7) imply (15), we conclude that (15) is more general than the previous two. ■

Remark 4: The scheme that achieves capacity in "very weak interference" is obtained by setting $R_{1c}=0$ in (17) so that the entire cognitive message is private and the primary message is common. The scheme that achieves capacity in "very strong interference" is obtained by setting $R_{1pb}=0$ in (17) so that both transmitters send only common messages. The scheme that we use to show the achievability in the "better cognitive decoding" regime mixes these two schemes by splitting the cognitive message into common and private messages. This relaxes the "very strong interference" achievability conditions as now the cognitive encoder needs to decode only part of the cognitive message. The scheme also relaxes the "very weak interference" achievability condition since it allows the cognitive encoder to decode part of the cognitive message and remove its undesirable effects.

VII. Capacity for the Semi-Deterministic CIFC

We next consider a class of *semi-deterministic* CIFC for which the signal at the cognitive receiver is an *arbitrary deterministic* function of the channel inputs, that is

$$Y_1=f_1(X_1,X_2) \tag{18}$$

for some function f_1. The class of channels in (18) was first introduced in [12], in the spirit of [28] and the capacity was derived under the additional conditions that (a) $I(Y_1;X_2)\geq I(Y_2;X_2)$ for all p_{X_1,X_2} and (b) f_1 is invertible given X_1. Here we extend the result in [12] by determining the capacity region in general—i.e., with no extra conditions besides the one in (18).

Theorem 11. New Capacity Result for the Semi-Deterministic Channel: The capacity region of the semi-deterministic CIFC in (18) consists of all non-negative pairs (R_1,R_2) such that

$$R_1\leq H(Y_1|X_2), \tag{19a}$$
$$R_2\leq I(Y_2;U,X_2), \tag{19b}$$
$$R_1+R_2\leq I(Y_2;U,X_2)+H(Y_1|U,X_2) \tag{19c}$$

taken over the union of all distributions p_{U,X_1,X_2}.

Proof: The converse follows by considering the outer bound of Theorem 1 with the additional deterministic assumption in (18) i.e., $H(Y_1|X_1,X_2)=0$.

For the achievability, consider the region in Theorem 7 for $U_{2c} = U_{1c} = \emptyset, U_1 = U_{1pb}$ and $U_2 = U_{2pb}$ and $R_{2c} = R_{1c} = R_{2pb} = 0$, that is

$$R'_{1pb} \geq I(U_1; X_2), \tag{20a}$$
$$R'_{1pb} + R'_{2pb} \geq I(U_1; U_2, X_2), \tag{20b}$$
$$R_2 + R'_{2pb} \leq I(Y_2; U_2, X_2), \tag{20c}$$
$$R_1 + R'_{1pb} \leq I(Y_1; U_1) \tag{20d}$$

for any $P_{U_1U_2X_1X_2}$. After Fourier–Motzkin elimination, the region in (20) may be rewritten as

$$R_1 \leq I(Y_1; U_1) - I(U_1; X_2), \tag{21a}$$
$$R_2 \leq I(Y_2; U_2, X_2), \tag{21b}$$
$$\begin{aligned} R_1 + R_2 &\leq I(Y_2; U_2, X_2) + I(Y_1; U_1) - I(U_1; U_2, X_2) \\ &= (21a) + (21b) - I(U_1; U_2|X_2). \end{aligned} \tag{21c}$$

Finally, by choosing $U_1 = Y_1$ (possible because Y_1 is a deterministic function of the inputs and both inputs are known at transmitter 1) and $U_2 = U$, the achievable rate region in (21) reduces to the outer bound in (19). ■

Remark 5: The achievability scheme in (20) cannot be obtained as a special case of any previously known achievability schemes except possibly the one proposed in [13] for the classical IFC with a cognitive relay. The RV U_{2pb}, which broadcasts the private primary message from transmitter 1, appears in [12] as well but it is not possible to obtain the scheme in (20) with a specific choice of the RVs. In the scheme of [12] the same primary private message is embedded in U_{2pb} and X_2, while in Theorem 11 U_{2pb} and X_2 carry two different primary private messages.

Remark 6: We used the achievability scheme for the semi-deterministic CIFC in (21) in [3], [29] to prove capacity to within 1 bit for the Gaussian CIFC. This supports the notion that results for (semi)-deterministic channels may carry over to noisy networks.

VIII. Capacity for the Deterministic CIFC

In the *deterministic* CIFC both outputs are arbitrary deterministic functions of the channel inputs, that is

$$Y_1 = f_1(X_1, X_2), \tag{22a}$$
$$Y_2 = f_2(X_1, X_2) \tag{22b}$$

for some functions f_1 and f_2. The class of channels in (22) is a subclass of the semi-deterministic CIFC in (18) for which Theorem 11 is the capacity. However, we rederive here the capacity region for the deterministic channel in (22) to show the achievability of the outer bound of Theorem 6 when letting $Y'_2 = Y_2$, instead of the outer bound of Theorem 1.

Theorem 12. New Capacity Result for the Deterministic Channel: The capacity region of the deterministic IFC in (22) consists of all non-negative pairs (R_1, R_2) such that

$$R_1 \leq H(Y_1|X_2), \tag{23a}$$
$$R_2 \leq H(Y_2), \tag{23b}$$
$$R_1 + R_2 \leq H(Y_2) + H(Y_1|Y_2, X_2) \tag{23c}$$

taken over the union of all distributions p_{X_1,X_2}.

Proof: The achievability follows immediately by choosing $U = Y_2$ in the capacity region in (19). Note that it is possible to set $U = Y_2$ because the codebook $U = U_{2pb}$ is generated at the cognitive transmitter that knows both inputs and thus knows Y_2 (because Y_2 is a deterministic function of the inputs by assumption). The choice $U = Y_2$ also maximizes the R_2-bound in (19b) since

$$I(Y_2; U, X_2) \leq I(Y_2; U, X_1, X_2) = H(Y_2).$$

However, it is not evident *a priori* that $U = Y_2$ also maximizes the sum-rate in (19c). To show that the sum-rate is indeed bounded by (23c), we use the sum-rate outer bound in (8c). Since we are dealing with deterministic channels, we can only choose $Y'_2 = Y_2$, from which the claim follows. ■

IX. Examples

The scheme that achieves capacity for the deterministic and semi-deterministic CIFC uses the RV U_{2pb} to perform Gel'fand–Pinsker binning to achieve the most general distribution among (X_2, U_{1pb}, U_{2pb}) with, quite interestingly, $R_{2pb} = 0$ and $R'_{2pb} \geq 0$. This feature of the capacity achieving scheme does not provide a clear intuition on the role of the RV U_{2pb}. For this reason we next present two examples of deterministic channels where the encoders can choose their respective codebooks in a way that allows binning of the interference without rate splitting. To make these examples more interesting, we choose them so that they do *not* fall into the category of the "very strong interference" regime of Theorem 5, which, in the deterministic case, reduces to

$$H(Y_1|X_2) \leq H(Y_2|X_2), \; H(Y_2) \leq H(Y_1) \tag{24}$$

for all p_{X_1,X_2}. Unfortunately, checking for the "very weak interference" condition of Theorem 4 is not possible as no cardinality bound on the alphabet of U is available.

A. Example I: The "Asymmetric Clipper" Channel

Consider the channel in Fig. 3. The input and output alphabets are $\mathcal{X}_1 = \mathcal{Y}_1 = [0 : 3]$ and $\mathcal{X}_2 = \mathcal{Y}_2 = [0 : 7]$ and the input/output relationships are

$$Y_1 = X_1 \oplus_4 X_2, \tag{25a}$$
$$Y_2 = 1_{\{2,3\}}(X_1) \oplus_8 X_2 \tag{25b}$$

where $1_{\{A\}}(x) = 1$ if $x \in A$ and zero otherwise and $\oplus_N$ denotes the addition modulo N. Also let $\mathcal{U}(\mathcal{S})$ be the uniform distribution over the set $\mathcal{S}$. First we show that the channel in (25) does not fall in the "very strong interference" class. For the input distribution:

$$X_1 \sim \mathcal{U}(\{1\}), \; X_2 \sim \mathcal{U}(\mathcal{X}_2)$$

we have

$$Y_1 \sim \mathcal{U}(\mathcal{Y}_1), \; Y_2 \sim \mathcal{U}(\mathcal{Y}_2)$$

so that

$$H(Y_2) = \log(|\mathcal{Y}_2|) = 3 > 2 = \log(|\mathcal{Y}_1|) = H(Y_1)$$

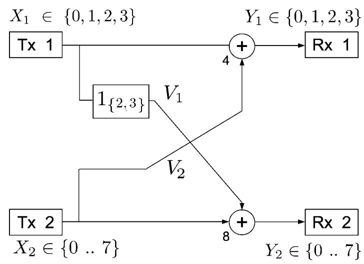

Fig. 3. The "asymmetric clipper" channel considered in Section IX-A.

which contradicts the "very strong interference" condition in (24). For this channel the outer bound in Theorem 12 is included in

$$R_1 \le H(Y_1|X_2) \le H(Y_1) \le \log(|\mathcal{Y}_1|) = 2, \tag{26a}$$
$$R_2 \le H(Y_2) \le \log(|\mathcal{Y}_2|) = 3, \tag{26b}$$
$$R_1 + R_2 \le H(Y_2) + H(Y_1|X_2, Y_2) \le 4 \tag{26c}$$

where $H(Y_1|X_2, Y_2) \le H(X_1|1_{\{2,3\}}(X_1)) \le 1$ follows from the multiplicity of the solutions of an addition in a Galois field. We now show that the region in (26) indeed corresponds to the capacity region in Theorem 12. Indeed, the corner point $(R_1, R_2) = (1, 3)$ in (26) is obtained in Theorem 12 with the input distribution:

$$X_1 \sim \mathcal{U}(\{0,1\}),\ X_2 \sim \mathcal{U}(\mathcal{X}_2)$$

while the corner point $(R_1, R_2) = (2, 2)$ in (26) is obtained in Theorem 12 with the input distribution:

$$X_1 \sim \mathcal{U}(\mathcal{X}_1),\ X_2 \sim \mathcal{U}(\mathcal{X}_2).$$

Time sharing between the two corner points shows that the region in (26) and the region in Theorem 12 coincide.

For the achievability of the corner point $(R_1, R_2) = (1, 3)$ consider the following strategy:

- transmitter 2 sends $x_2 = w_2$;
- transmitter 1 sends $x_1 = [w_1 + w_2]_2$;
- receiver 1 decodes $\hat{w}_1 = [y_1]_2$;
- receiver 2 decodes $\hat{w}_2 = y_2 = x_2$.

It can be verified by inspection of Table I, which shows for each possible message pair (w_1, w_2) the corresponding channel inputs (x_1, x_2), channel outputs (y_1, y_2) and decoded messages $(\hat{w}_1, \hat{w}_2)$, that the rate pair $(R_1, R_2) = (1, 3)$ is indeed achievable.

For the achievability of the corner point $(R_1, R_2) = (2, 2)$ consider the following strategy:

- transmitter 2 sends $x_2 = 2w_2$;
- transmitter 1 sends $x_1 = [w_1 + 2w_2]_4$;
- receiver 1 decodes $\hat{w}_1 = y_1$;
- receiver 2 decodes $\hat{w}_2 = \lfloor \frac{y_2}{2} \rfloor$.

It can be verified by the inspection of Table II, which uses the same convention as Table I, that the rate pair $(R_1, R_2) = (2, 2)$ is indeed achievable.

In this example,we see how the two senders jointly design their codebooks to achieve the outer bound and in particular how the cognitive transmitter adapts its strategy to the transmission of the primary transmitter so as to avoid interfering with it. Also notice that the capacity achieving strategy achieves zero-error in a single channel use, hence the capacity region coincides with the zero-error capacity.

TABLE I
ACHIEVABILITY OF THE RATE POINT $(R_1, R_2) = (1, 3)$ IN EXAMPLE I IN SECTION IX-A: FOR EACH POSSIBLE MESSAGE PAIR (w_1, w_2), WE INDICATE THE CORRESPONDING CHANNEL INPUTS (x_1, x_2), CHANNEL OUTPUTS (y_1, y_2) AND DECODED MESSAGES $(\hat{w}_1, \hat{w}_2)$

	w_1	w_2	x_1	x_2	y_1	y_2	$\widehat{w}_1$	$\widehat{w}_2$
0	0	0	0	0	0	0	0	0
1	0	1	1	1	2	1	0	1
2	0	2	0	2	2	2	0	2
3	0	3	1	3	0	3	0	3
4	0	4	0	4	0	4	0	4
5	0	5	1	5	2	5	0	5
6	0	6	0	6	2	6	0	6
7	0	7	1	7	0	7	0	7
8	1	0	1	0	1	0	1	0
9	1	1	0	1	1	1	1	1
10	1	2	1	2	3	2	1	2
11	1	3	0	3	3	3	1	3
12	1	4	1	4	1	4	1	4
13	1	5	0	5	1	5	1	5
14	1	6	1	6	3	6	1	6
15	1	7	0	7	3	7	1	7

TABLE II
ACHIEVABILITY OF THE RATE POINT $(R_1, R_2) = (2, 2)$ IN EXAMPLE I IN SECTION IX-A: FOR EACH POSSIBLE MESSAGE PAIR (w_1, w_2), WE INDICATE THE CORRESPONDING CHANNEL INPUTS (x_1, x_2), CHANNEL OUTPUTS (y_1, y_2) AND DECODED MESSAGES $(\hat{w}_1, \hat{w}_2)$

	w_1	w_2	x_1	x_2	y_1	y_2	$\widehat{w}_1$	$\widehat{w}_2$
0	0	0	0	0	0	0	0	0
1	0	1	2	2	0	3	0	1
2	0	2	0	4	0	4	0	2
3	0	3	2	6	0	7	0	3
4	1	0	1	0	1	0	1	0
5	1	1	3	2	1	3	1	1
6	1	2	1	4	1	4	1	2
7	1	3	3	6	1	7	1	3
8	2	0	2	0	2	1	2	0
9	2	1	0	2	2	2	2	1
10	2	2	2	4	2	5	2	2
11	2	3	0	6	2	6	2	3
12	3	0	3	0	3	1	3	0
13	3	1	1	2	3	2	3	1
14	3	2	3	4	3	5	3	2
15	3	3	1	6	3	6	3	3

In achieving the point $(R_1, R_2) = (1, 3)$, transmitter 2's strategy is that of a point-to-point channel. The cognitive transmitter chooses its codewords so as not to interfere with the primary transmission. Only two codewords do not interfere: it alternatively picks one of these two codewords to produce the desired channel output. For example, when the primary sends $w_2 = 0$ (line 0 and 8 in Table I) transmitter 1 can send either 1 or 0 without creating interference at receiver 2. On the other hand, these two values produce a different output at receiver 1, allowing the transmission at rate $R_1 = 1$ bit.

In achieving the point $(R_1, R_2) = (2, 2)$, the primary transmitter picks its codewords so as to tolerate one unit of interference. Transmitter 1 again chooses its codewords in order to create at most one unit of interference at the primary decoder. By adapting its transmission to the primary user, the cognitive transmitter is able to always find four such codewords. It is interesting to notice the tension at transmitter 1 between the interference it creates at the primary decoder and its own rate. There is an optimal trade-off between these two quantities that is achieved by

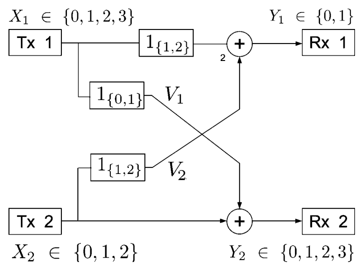

Fig. 4. "Symmetric Clipper" channel considered in Section IX-B.

TABLE III
INPUT DISTRIBUTION THAT ACHIEVES THE OUTER BOUND IN THEOREM 12 FOR THE CHANNEL IN EXAMPLE II IN SECTION IX-B

$X_2,\ X_1$	1	2	3	4	p_{X_1}
0	1/8	1/8	1/8	1/8	1/2
1	1/8	1/8	0	0	1/4
2	1/8	1/8	0	0	1/4
p_{X_2}	3/8	3/8	1/8	1/8	

carefully picking the codewords at the cognitive transmitter. For example, when the primary transmitter sends $w_2 = 0$ (lines 0, 4, 8, and 12 in Table II), transmitter 1 can send $x_1 \in [0:3]$ and create at most one unit of interference at receiver 2. Each of these four values produces a different output at receiver 1, thus allowing the transmission at rate $R_1 = 2$ bits.

B. Example II: The "Symmetric Clipper" Channel

Consider the now channel in Fig. 4 whose input and output alphabets are $\mathcal{X}_1 = [0:3] = \mathcal{Y}_2$, $\mathcal{X}_2 = [0:2]$ and $\mathcal{Y}_1 = [0:1]$. The input/output relationships are

$$Y_1 = 1_{\{1,2\}}(X_1) \oplus_2 1_{\{1,2\}}(X_2), \tag{27a}$$

$$Y_2 = 1_{\{0,1\}}(X_1) \oplus X_2. \tag{27b}$$

The channel in (27) does not fall in the "very strong interference" class since for the input distribution:

$$X_1 \sim \mathcal{U}(\{3\}),\ X_2 \sim \mathcal{U}(\{1,2\})$$

we have $H(Y_1) = 0$ and $H(Y_2) = 1$, which contradicts the "very strong interference" condition in (24).

The outer bound of Theorem 12 is achieved by the input distribution p_{X_1,X_2} in Table III. This distribution produces $H(Y_1) = 1 = \log_2(|\mathcal{Y}_1|)$ and $H(Y_2) = 2 = \log(|\mathcal{Y}_2|)$, which are the largest possible output entropies given the cardinality of the output alphabets. We therefore conclude that the region in Theorem 12 is equivalent to

$$R_1 \leq 1, \tag{28a}$$

$$R_2 \leq 2. \tag{28b}$$

The region in (28) is achieved by using the transmission scheme described in Table IV, which shows for each possible message pair (w_1, w_2), the corresponding channel inputs (x_1, x_2) and channel outputs $(y_1, y_2) = (\hat{w}_1, \hat{w}_2)$. This transmission scheme achieves the proposed outer bound, thus showing capacity. The transmission scheme can be described as follows:

- encoder 2 sends $x_2 = \lfloor \frac{w_2}{2} \rfloor$,
- encoder 1 sends the value x_1 that simultaneously makes $y_1 = w_1$ and $y_2 = w_2$,
- receivers 1 and 2 decode $\hat{w}_1 = y_1$ and $\hat{w}_2 = y_2$, respectively.

TABLE IV
ACHIEVABILITY OF THE RATE POINT $(R_1, R_2) = (1, 2)$ IN EXAMPLE II IN SECTION IX-B: FOR EACH POSSIBLE MESSAGE PAIR (w_1, w_2), WE INDICATE THE CORRESPONDING CHANNEL INPUTS (x_1, x_2) AND CHANNEL OUTPUTS $(y_1, y_2) = (\hat{w}_1, \hat{w}_2)$

	w_1	w_2	x_1	x_2	y_1	y_2
0	0	0	3	0	0	0
1	0	1	0	0	0	1
2	0	2	1	1	0	2
3	0	3	1	1	0	3
4	1	0	2	0	1	0
5	1	1	1	0	1	1
6	1	2	0	1	1	2
7	1	3	0	1	1	3

This example is particularly interesting since both decoders obtain their intended message without suffering any interference. Here cognition allows the simultaneous cancelation of the interference at both decoders. Encoder 2 has only three codewords and relies on transmitter 1 to achieve its full rate of $R_2 = 2$. In fact encoder 1 is able to design its codebook to transmit two codewords for its decoder and still contribute to the rate of primary user by making the codewords corresponding to $w_2 \in \{2, 3\}$ distinguishable at the cognitive decoder. This feature of the capacity achieving scheme is intriguing: the primary transmitter needs the support of the cognitive transmitter to achieve $R_2 = 2$ since its input alphabet has cardinality three. That is, the primary pair achieves a larger rate thanks to the cognitive pair than it would in its absence. This shows that cognition may benefit both source-destination pairs.

For example consider the transmission of $w_2 = 2$ or 3 (lines 2, 3, 6 and 7 in Table IV). In this case transmitter 1 sends $x_1 = 0$ or $x_1 = 1$ to simultaneously influence both channel outputs so that both decoders receive the desired symbols. This simultaneous cancellation is possible due to the channel's deterministic nature and the extra message knowledge at the cognitive transmitter.

X. CONCLUSION

In this paper,we focused on the general memoryless cognitive interference channel. We proposed new inner and outer bounds, and derived the capacity for certain classes of channels. Our new outer bound builds on the fact that the capacity of channels without receiver cooperation only depends on the channel conditional marginal distributions and results in a bound that does not involve auxiliary RVs, which is thus easily computable. Our new inner bound generalizes all other known achievable rate regions and is the largest rate region known to date. We determined the capacity for a class of channels in the "better cognitive decoding" regime, which includes the "very weak" and the "very strong" interference regimes for which capacity was known and is the largest region where capacity is known to date. We also determined the capacity for the semi-deterministic channel where the cognitive receiver's output is a deterministic function of the inputs. Furthermore, for channels where both outputs are deterministic functions of the inputs, we showed the achievability of our new outer bound. Extensions of the results presented here to Gaussian channels are presented in [29].

APPENDIX A
PROOF OF THEOREM 6

The R_1-bound in (8a) is as in (1a). The R_2-bound in (8b) and the sum-rate bound in (8c) are looser than (1a) and (1c), respectively, as pointed out in Remark 1. This proves that the region in (8) is an outer bound for the general CIFC. Nonetheless, we offer a novel proof for the sum-rate bound in (8c) that uses the fact that the capacity region only depends on the conditional marginal distributions because the receivers do not cooperate [19].

By Fano's inequality, $H(W_i|Y_i^N) \le N\epsilon_N$, for some ϵ_N such that $\epsilon_N \to 0$ as $N \to 0$ for $i \in [1:2]$. Let Y_2' be a RV such that $P_{Y_2'|X_1,X_2} = P_{Y_2|X_1,X_2}$ but with any joint distribution $P_{Y_1,Y_2'|X_1,X_2}$. The sum-rate can then be bounded as

$$
\begin{aligned}
&N(R_1+R_2-2N\epsilon_N) \le I(W_1;Y_1^N)+I(W_2;Y_2^N)\\
&\overset{(a)}{\le} I(W_1;Y_1^N|W_2)+I(W_2;Y_2^N)\\
&\overset{(b)}{\le} I(W_1;Y_1^N,Y_2'^N|W_2)+I(W_2;Y_2^N)\\
&\overset{(c)}{=} I(W_2;Y_2^N)+I(W_1;Y_2'^N|W_2)+I(W_1;Y_1^N|Y_2'^N,W_2)\\
&\overset{(d)}{\le} H(Y_2^N)+\Big(-H(Y_2^N|W_2)+H(Y_2'^N|W_2)\Big)\\
&\quad -H(Y_2'^N|W_1,W_2)+H(Y_1^N|Y_2'^N,W_2)\\
&\quad -H(Y_1^N|Y_2'^N,W_1,W_2)\\
&\overset{(e)}{=} H(Y_2^N)+H(Y_1^N|W_2,X_2^N,Y_2'^N)\\
&\quad -H(Y_2'^N|W_1,W_2,X_1^N,X_2^N)\\
&\quad -H(Y_1^N|Y_2'^N,W_1,W_2,X_1^N,X_2^N)\\
&\overset{(f)}{=} H(Y_2^N)+H(Y_1^N|W_2,X_2^N,Y_2'^N)\\
&\quad -H(Y_2^N|X_1^N,X_2^N)-H(Y_1^N|Y_2'^N,X_1^N,X_2^N)\\
&\overset{(g)}{\le} H(Y_2^N)+H(Y_1^N|X_2^N,Y_2'^N)\\
&\quad -H(Y_2^N|X_1^N,X_2^N)-H(Y_1^N|Y_2'^N,X_1^N,X_2^N)\\
&\overset{(h)}{=} I(Y_2^N;X_1^N,X_2^N)\\
&\quad +\sum_{i=1}^{N} H(Y_{1i}|X_2^N,Y_2'^N,Y_1^{i-1})\\
&\quad -\sum_{i=1}^{N} H(Y_{1i}|X_1^N,X_2^N,Y_2'^N,Y_1^{i-1})\\
&\overset{(i)}{\le} I(Y_2^N;X_1^N,X_2^N)\\
&\quad +\sum_{i=1}^{N} H(Y_{1i}|X_{2i},Y_{2i}')-H(Y_{1i}|X_{1i},X_{2i},Y_{2i}')\\
&\overset{(j)}{=} N\left(I(Y_{2T};X_{1T},X_{2T})+H(Y_{1T}|X_{2T},Y_{2T}',T)\right.\\
&\quad \left.-H(Y_{1T}|X_{1T},X_{2T},Y_{2T}')\right)\\
&\overset{(k)}{\le} N\left(I(Y_{2T};X_{1T},X_{2T})+I(Y_{1T};X_{1T}|X_{2T},Y_{2T}')\right).
\end{aligned}
$$

Here the (in)equalities follow from (a) non-negativity of mutual information and independence of W_1, W_2, (b) addition of side-information $Y_2'^N$, (d) definition, (e) as Y_2 and Y_2' have the same marginals and the channel model where X_1^N depends on W_1 and W_2, while X_2^N depends only on W_2, (f) as $(W_1,W_2) \to (X_1^N,X_2^N) \to (Y_1^N)$ forms a Markov chain, (g) conditioning reduces entropy, (h) chain rule, (i) conditioning reduces entropy and memorylessness, (j) and (k) memorylessness of the channel, definition of the time-sharing RV T uniformly distributed over the set $[1:N]$ and independent of everything else.

APPENDIX B
ERROR ANALYSIS FOR THEOREM 7

Without loss of generality assume that the message $(w_{1c}, w_{2c}, w_{2pa}, w_{1pb}, w_{2pb}) = (1,1,1,1,1)$ was sent and let $(\bar{b}_0, \bar{b}_1, \bar{b}_2)$ be the triplet (b_0, b_1, b_2) chosen at encoder 1. Let $(\hat{w}_{1c}, \hat{w}_{2c}, \hat{w}_{2pa}, \hat{w}_{2pb}, \hat{b}_0, \hat{b}_2)$ be the estimate at the decoder 2 and $(\hat{\hat{w}}_{1c}, \hat{\hat{w}}_{2c}, \hat{\hat{w}}_{1pb}, \hat{\hat{b}}_0, \hat{\hat{b}}_1)$ be the estimate at the decoder 1. The probability of error at decoder $u \in [1,2]$ is bounded by

$$
\begin{aligned}
\Pr[\text{error } u] \le\ & \Pr[\text{error } u|\text{encoding successful}]\\
&+\Pr[\text{encoding NOT successful}].
\end{aligned}
$$

An encoding error occurs if encoder 1 is not able to find a tuple $(\bar{b}_0, \bar{b}_1, \bar{b}_2)$ that guarantees typicality. A decoding error is committed at decoder 1 when $(\hat{\hat{w}}_{1c}, \hat{\hat{w}}_{1pb}) \ne (1,1)$. A decoding error is committed at decoder 2 when $(\hat{w}_{2c}, \hat{w}_{2pa}, \hat{w}_{2pb}) \ne (1,1,1)$.

A) Encoding Error: Since the codebooks are generated iid according to

$$
\begin{aligned}
p^{(\text{codebook})} &= p_{U_{2c}}\, p_{X_2|U_{2c}}\\
&\quad p_{U_{1c}|U_{2c}}\, p_{U_{2pb}|U_{2c},U_{1c},X_2}\, p_{U_{1pb}|U_{2c},U_{1c}}
\end{aligned} \tag{29}
$$

but the encoding forces the actual transmitted codewords to look as if they were generated iid according to

$$
\begin{aligned}
p^{(\text{encoding})} &= p_{U_{2c}}\, p_{X_2|U_{2c}}\\
&\quad p_{U_{1c}|U_{2c},X_2}\, p_{U_{2pb}|U_{2c},U_{1c},X_2} p_{U_{1pb}|U_{2c},U_{1c},X_2,U_{2pb}}
\end{aligned} \tag{30}
$$

we thus expect the probability of encoding error to depend on

$$
\begin{aligned}
\mathbb{E}\left[\log\frac{p^{(\text{encoding})}}{p^{(\text{codebook})}}\right] &= \mathbb{E}\left[\log\frac{p_{U_{1c}|U_{2c},X_2}\, p_{U_{1pb}|U_{2c},U_{1c},X_2,U_{2pb}}}{p_{U_{1c}|U_{2c}}\, p_{U_{1pb}|U_{2c},U_{1c}}}\right]\\
&= I(U_{1c};X_2|U_{2c})+I(U_{1pb};X_2,U_{2pb}|U_{2c},U_{1c}).
\end{aligned}
$$

The probability that the encoding fails can be bounded as

$$
\begin{aligned}
\Pr[\text{encoding NOT successful}] &= P\Bigg[\bigcap_{b_0=1}^{2^{NR'_{1c}}}\bigcap_{b_1=1}^{2^{NR'_{1pb}}}\bigcap_{b_2=1}^{2^{NR'_{2pb}}}\\
&\quad \big(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0), U_{1pb}^N(1,1,b_0,1,b_1),\\
&\quad U_{2pb}^N(1,1,b_0,1,b_2)\big) \notin T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}})\Bigg]\\
&= \Pr[K=0] \le \frac{\mathbb{V}\mathrm{ar}[K]}{\mathbb{E}^2[K]}
\end{aligned}
$$

where

$$
K \triangleq \sum_{b_0=1}^{2^{NR'_{1c}}}\sum_{b_1=1}^{2^{NR'_{1pb}}}\sum_{b_2=1}^{2^{NR'_{2pb}}} K_{b_0,b_1,b_2}
$$

and

$$K_{b_0,b_1,b_2} \triangleq 1_{\{\mathcal{E}\}},$$
$$\mathcal{E} = \{(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0), U_{1pb}^N(1,1,b_0,1,b_1), U_{2pb}^N(1,1,b_0,1,b_2)) \in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}})\}$$

where $1_{\{\mathcal{E}\}} = 1$ if the condition expressed by $\mathcal{E}$ is true and zero otherwise.

The mean value of K (neglecting all terms that depend on ϵ and that eventually go to zero as $N \to \infty$) is

$$\mathbb{E}[K] = \sum_{b_0=1}^{2^{NR'_{1c}}} \sum_{b_1=1}^{2^{NR'_{1pb}}} \sum_{b_2=1}^{2^{NR'_{2pb}}} \Pr[K_{b_0,b_1,b_2} = 1] = 2^{N(R'_{1c}+R'_{1pb}+R'_{2pb}-A)}$$

with

$$\begin{aligned}2^{-NA} &= \Pr[K_{b_0,b_1,b_2} = 1] = \mathbb{E}[K_{b_0,b_1,b_2}]\\ &= \Pr[\big(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0), U_{1pb}^N(1,1,b_0,1,b_1), U_{2pb}^N(1,1,b_0,1,b_2)\big) \in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}})]\\ &= \sum_{(u_{1c}^N,u_{1pb}^N,u_{2pb}^N)\in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}}|u_{2c}^N,x_2^N)} p^N_{U_{1c}|U_{2c}}\, p^N_{U_{2pb}|U_{2c},U_{1c},X_2}\, p^N_{U_{1pb}|U_{2c},U_{1c}}\\ &\geq 2^{-N[I(U_{1c};X_2|U_{2c})+I(U_{1pb};X_2,U_{2pb}|U_{1c},U_{2c})]}.\end{aligned}$$

Here p_X^N denotes the N-th memoryless extension of the density p_X for a RV X.

The variance of K (neglecting all terms that depend on ϵ and that eventually go to zero as $N \to \infty$) is

$$\begin{aligned}\mathbb{V}\mathrm{ar}[K] &= \sum_{b_0=1}^{2^{NR'_{1c}}} \sum_{b_1=1}^{2^{NR'_{1pb}}} \sum_{b_2=1}^{2^{NR'_{2pb}}} \sum_{b'_0=1}^{2^{NR'_{1c}}} \sum_{b'_1=1}^{2^{NR'_{1pb}}} \sum_{b'_2=1}^{2^{NR'_{2pb}}} \big(\Pr[K_{b_0,b_1,b_2} = 1, K_{b'_0,b'_1,b'_2} = 1]\\ &\qquad - \Pr[K_{b_0,b_1,b_2} = 1]\Pr[K_{b'_0,b'_1,b'_2} = 1]\big)\\ &= \sum_{b'_0=b_0,(b_1,b_2,b'_1,b'_2)} \big(\Pr[K_{b_0,b_1,b_2} = 1, K_{b_0,b'_1,b'_2} = 1]\\ &\qquad - \Pr[K_{b_0,b_1,b_2} = 1]\Pr[K_{b_0,b'_1,b'_2} = 1]\big)\\ &\leq \sum_{b_0,(b_1,b_2,b'_1,b'_2)} \Pr[K_{b_0,b_1,b_2} = 1, K_{b_0,b'_1,b'_2} = 1]\end{aligned}$$

because when $b_0 \neq b'_0$, that is, $U_{1c}^N(\ldots,b_0)$ and $U_{1c}^N(\ldots,b'_0)$ are independent (here the dots are in place of indices that are the same in both codewords), the RVs K_{b_0,b_1,b_2} and $K_{b'_0,b'_1,b'_2}$ are independent and they do not contribute to the summation. We thus can focus only on the case $b_0 = b'_0$.

We can write

$$\begin{aligned}\mathbb{V}\mathrm{ar}[K] \leq &\underbrace{\sum_{b_0,\, b_1=b'_1,\, b_2=b'_2} \Pr[K_{b_0,b_1,b_2} = 1]}_{=\mathbb{E}[K]}\\ &+ \underbrace{\sum_{b_0,\, b_1=b'_1,\, b_2\neq b'_2} \Pr[K_{b_0,b_1,b_2} = 1]\Pr[K_{b_0,b_1,b'_2} = 1|K_{b_0,b_1,b_2} = 1]}_{=\mathbb{E}[K]\,2^{N(R'_{2pb}-B)}}\\ &+ \underbrace{\sum_{b_0,\, b_1\neq b'_1,\, b_2=b'_2} \Pr[K_{b_0,b_1,b_2} = 1]\Pr[K_{b_0,b'_1,b_2} = 1|K_{b_0,b_1,b_2} = 1]}_{=\mathbb{E}[K]\,2^{N(R'_{1pb}-C)}}\\ &+ \underbrace{\sum_{b_0,\, b_1\neq b'_1,\, b_2\neq b'_2} \Pr[K_{b_0,b_1,b_2} = 1]\Pr[K_{b_0,b'_1,b'_2} = 1|K_{b_0,b_1,b_2} = 1]}_{=\mathbb{E}[K]\,2^{N(R'_{1pb}+NR'_{2pb}-D)}}\end{aligned}$$

and

$$\begin{aligned}2^{-NB} &= \Pr[K_{b_0,b_1,b'_2} = 1|K_{b_0,b_1,b_2} = 1]\\ &= \Pr[\big(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0), U_{1pb}^N(1,1,b_0,1,b_1), U_{2pb}^N(1,1,b_0,1,b'_2)\big) \in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}})|\\ &\quad \big(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0), U_{1pb}^N(1,1,b_0,1,b_1), U_{2pb}^N(1,1,b_0,1,b_2)\big) \in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}})]\\ &= \sum_{u_{2pb}^N\in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}}|u_{2c}^N,x_2^N,u_{1c}^N,u_{1pb}^N)} p^N_{U_{2pb}|U_{2c},U_{1c},X_2}\\ &= 2^{-NI(U_{2pb};U_{1pb}|U_{2c},U_{1c},X_2)}\end{aligned}$$

and

$$\begin{aligned}2^{-NC} &= \Pr[K_{b_0,b'_1,b_2} = 1|K_{b_0,b_1,b_2} = 1]\\ &= \Pr[\big(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0), U_{1pb}^N(1,1,b_0,1,b'_1), U_{2pb}^N(1,1,b_0,1,b_2)\big) \in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}})|\\ &\quad \big(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0), U_{1pb}^N(1,1,b_0,1,b_1), U_{2pb}^N(1,1,b_0,1,b_2)\big) \in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}})]\\ &= \sum_{u_{1pb}^N\in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}}|u_{2c}^N,x_2^N,u_{1c}^N,u_{2pb}^N)} p^N_{U_{1pb}|U_{2c},U_{1c}}\\ &= 2^{-NI(U_{1pb};X_2,U_{2pb}|U_{1c},U_{2c})}\end{aligned}$$

and

$$\begin{aligned}2^{-ND} &= \Pr[K_{b_0,b'_1,b'_2} = 1|K_{b_0,b_1,b_2} = 1]\\ &= \Pr[\big(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0), U_{1pb}^N(1,1,b_0,1,b'_1), U_{2pb}^N(1,1,b_0,1,b'_2)\big) \in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}})|\\ &\quad \big(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0), U_{1pb}^N(1,1,b_0,1,b_1), U_{2pb}^N(1,1,b_0,1,b_2)\big) \in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}})]\\ &= \sum_{(u_{1pb}^N,u_{2pb}^N)\in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}}|u_{2c}^N,x_2^N,u_{1c}^N)} p^N_{U_{2pb}|U_{2c},U_{1c},X_2}\, p^N_{U_{1pb}|U_{2c},U_{1c}}\\ &= 2^{-NI(U_{1pb};X_2,U_{2pb}|U_{1c},U_{2c})} = 2^{-NC}.\end{aligned}$$

Hence, we can bound $\Pr[K = 0]$ as

$$\Pr[K = 0] \leq \frac{1 + 2^{N(R'_{1pb}-C)} + 2^{N(R'_{2pb}-B)} + 2^{N(R'_{1pb}+R'_{2pb}-C)}}{2^{N(R'_{1c}+R'_{1pb}+R'_{2pb}-I(U_{1c};X_2|U_{2c})-C)}}$$

TABLE V
ERROR EVENTS AT DECODER 2

Event	w_{2c}	(w_{1c}, b_1)	w_{2pa}	w_{2pb}	$p_{Y_2\|\star}$
$E_{2,1}$	X	$\cdots$	$\cdots$	$\cdots$	p_{Y_2}
$E_{2,2a}$	1	X	X	$\cdots$	$p_{Y_2\|U_{2c}}$
$E_{2,2b}$	1	1	X	$\cdots$	$p_{Y_2\|U_{2c},U_{1c}}$
$E_{2,3a}$	1	X	1	$\cdots$	$p_{Y_2\|U_{2c},X_2}$
$E_{2,3b}$	1	1	1	X	$P_{Y_2\|U_{2c},U_{1c},X_2}$

and $\Pr[K=0] \to 0$ if

$$\begin{aligned}&R'_{1c}+R'_{1pb}+R'_{2pb}-I(U_{1c};X_2|U_{2c})-C>0\\&R'_{1c}+R'_{1pb}+R'_{2pb}-I(U_{1c};X_2|U_{2c})-C-(R'_{2pb}-B)>0\\&R'_{1c}+R'_{1pb}+R'_{2pb}-I(U_{1c};X_2|U_{2c})-C-(R'_{1pb}-C)>0\\&R'_{1c}+R'_{1pb}+R'_{2pb}-I(U_{1c};X_2|U_{2c})-C\\&\quad-(R'_{1pb}+R'_{2pb}-C)>0\end{aligned}$$

that is, if the inequality in (11a)–(11c) hold. Note that the second to last constraint in the above expression is redundant.

B) Decoding Errors at Decoder 2: If decoder 2 decodes a $(\hat{w}_{2c}, \hat{w}_{2pa}, \hat{w}_{2pb}) \neq (1,1,1)$, then an error is committed. The probability of error at decoder 2 is bounded as

$$\Pr[\text{error 2}|\text{encoding successful}] \le \sum_{i\in\{1,2a,2b,3a,3b\}} \Pr[E_{2,i}], \tag{31}$$

where $E_{2,i},\ i\in\{1,2a,2b,3a,3b\}$, are the error events defined in Table V. In Table V, an “X” means that the corresponding message is in error (when the header of the column contains two indices, an “X” indicates that at least one of the two indexes is wrong), a “1” means that the corresponding message is correct, while the dots “$\cdots$” indicate that “*it does not matter whether the corresponding message is correct or not, because of superposition coding; in this case the most restrictive case is when the message is actually in error.*” The last column of Table V specifies the $p_{Y_2|\star}$ to be used in (32) defined below.

Depending on which messages are wrongly decoded at decoder 2, the generated sequences and the received Y_2^n are generated iid according to

$$p_{2|\star} \triangleq p_{U_{2c}}\, p_{X_2|U_{2c}}\, p_{U_{1c}|U_{2c}}\, p_{U_{2pb}|U_{2c},U_{1c},X_2}\, p_{Y_2|\star} \tag{32}$$

where “$\star$” indicates the messages decoded correctly. However, the actual transmitted sequences and the received Y_2^n considered at decoder 2 look as if they were generated iid according to

$$\begin{aligned}p_2 \triangleq\ & p_{U_{2c}}\, p_{X_2|U_{2c}}\\ & p_{U_{1c}|U_{2c},X_2}\, p_{U_{2pb}|U_{2c},U_{1c},X_2}\, p_{Y_2|U_{2c},U_{1c},X_2,U_{2pb}}.\end{aligned} \tag{33}$$

Hence we expect the probability of error at decoder 2 to depend on terms of the type

$$\begin{aligned}I_{2|\star} &= \mathbb{E}\left[\log\frac{p_2}{p_{2|\star}}\right]\\ &= \mathbb{E}\left[\log\frac{p_{U_{1c}|U_{2c},X_2}\, p_{Y_2|U_{2c},U_{1c},X_2,U_{2pb}}}{p_{U_{1c}|U_{2c}}\, p_{Y_2|\star}}\right]\\ &= I(U_{1c};X_2|U_{2c}) + I(Y_2;U_{2c},U_{1c},X_2,U_{2pb}|\star).\end{aligned} \tag{34}$$

We now proceed to bound the probability of all the events in (31). We have that $\Pr[\text{error 2}|\text{encoding successful}] \to 0$ when $N\to\infty$ if

- When the event $E_{2,1}$ occurs we have $\hat{w}_{2c}\neq 1$. In this case the received Y_2^N is independent of the transmitted sequences. This follows from the fact that the codewords U_{2c}^N are generated in an iid fashion and all the other codewords are generated independently conditioned on U_{2c}^N. Hence, when decoder 2 finds a wrong U_{2c}^N, all the decoded codewords are independent of the transmitted ones. We can bound the error probability of $E_{2,1}$ as

$$\begin{aligned}\Pr[E_{2,1}] =,\ & P\Bigg[\bigcup_{\tilde{w}_{2c}\neq 1,\tilde{w}_{2pa},\tilde{w}_{1c},\tilde{w}_{2pb},b_0,b_2}\\ & (Y_2^N, U_{2c}^N(\tilde{w}_{2c}), U_{1c}^N(\tilde{w}_{1c},\tilde{w}_{2c},b_0), X_2^N(\tilde{w}_{2c},\tilde{w}_{2pa}),\\ & U_{2pb}^N(\tilde{w}_{2c},\tilde{w}_{2pa},\tilde{w}_{1c},b_0,\tilde{w}_{2pb},b_2)) \in T_\epsilon^N\big(p_{Y_2,U_{2c},U_{1c},X_2,U_{2pb}}\big)\Bigg]\\ &\le 2^{N(R_{2c}+R_{2pa}+R_{1c}+R'_{1c}+R_{2pb}+R'_{2pb})}\\ &\quad \sum_{(y_2^N,u_{2c}^N,u_{1c}^N,x_2^N,u_{2pb}^N)\in T_\epsilon^N\left(p_{Y_2,U_{2c},U_{1c},X_2,U_{2pb}}\right)} p_{2|\star}^N|_{\star=\emptyset}\\ &\le 2^{N(R_{2c}+R_{2pa}+R_{1c}+R'_{1c}+R_{2pb}+R'_{2pb}-I_{2|\star}|_{\star=\emptyset})}\end{aligned}$$

  for $p_{2|\star}$ given in (33) and $I_{2|\star}$ given in (34). Hence $\Pr[E_{2,1}] \to 0$ as $N\to\infty$ if (11d) is satisfied.
- When the event $E_{2,2}$ occurs, i.e., either $E_{2,2a}$ or $E_{2,2b}$, we have $\hat{w}_{2c}=1$ but $\hat{w}_{2pa}\neq 1$. Whether $\hat{w}_{1c}$ is correct or not, it does not matter since decoder 2 is not interested in $\hat{w}_{1c}$. However we need to consider whether the pair $(\hat{w}_{1c},\hat{b}_0)$ is equal to the transmitted one or not because this affects the way the joint probability among all involved RVs factorizes. We have
  - Case $E_{2,2a}$: either $\hat{w}_{1c}\neq 1$ or $\hat{b}_0 \neq \bar{b}_0$. In this case, conditioned on the (correct) decoded sequence U_{2c}^N, the output Y_2^N is independent of the (wrong) decoded sequences $(U_{1c}^N, X_2^N, U_{2pb}^N)$ (because U_{2pb}^N is superimposed to the wrong pair (U_{1c}^N, X_2^N)). It is easy to see that the most stringent error event is when both $\hat{w}_{1c}\neq 1$ and $\hat{b}_0\neq\bar{b}_0$. Thus we have

$$\begin{aligned}\Pr[E_{2,2a}] =\ & P\Bigg[\bigcup_{\tilde{w}_{2pa}\neq 1,\tilde{w}_{1c}\neq 1,b_0\neq\bar{b}_0,\tilde{w}_{2pb},b_2}\\ & (Y_2^N, U_{2c}^N(1), U_{1c}^N(1,\tilde{w}_{1c},b_0), X_2^N(1,\tilde{w}_{2pa}),\\ & \quad U_{2pb}^N(1,\tilde{w}_{2pa},\tilde{w}_{1c},b_0,\tilde{w}_{2pb},b_2))\\ & \quad \in T_\epsilon^N\left(p_{Y_2,U_{2c},U_{1c},X_2,U_{2pb}}\right)\Bigg]\\ &\le 2^{N(R_{2pa}+R_{1c}+R'_{1c}+R_{2pb}+R'_{2pb})}\\ &\quad\sum_{(y_2^N,u_{2c}^N,u_{1c}^N,x_2^N,u_{2pb}^N)\in T_\epsilon^N\left(p_{Y_2,U_{2c},U_{1c},X_2,U_{2pb}}\right)}\\ &\quad p_{2|\star}^N|_{\star=U_{2c}}\\ &\le 2^{N(R_{2pa}+R_{1c}+R'_{1c}+R_{2pb}+R'_{2pb}-I_{2|\star}|_{\star=U_{2c}})}\end{aligned}$$

  for $p_{2|\star}$ given in (33) and $I_{2|\star}$ given in (34). Hence $\Pr[E_{2,2a}]\to 0$ as $N\to\infty$ if (11g) is satisfied.

— Case $E_{2,2b}$: both $\hat{w}_{1c}=1$ and $\hat{b}_0=\bar{b}_0$. In this case, conditioned on the (correct) decoded (U_{2c}^N, U_{1c}^N), the output Y_2^N is independent of the (wrong) decoded sequences (X_2^N, U_{2pb}^N). Thus we have

$$\begin{aligned}\Pr[E_{2,2b}] &= P\Bigg[\bigcup_{\tilde{w}_{2pa}\neq 1, \tilde{w}_{2pb}, b_2} (Y_2^N, U_{2c}^N(1), U_{1c}^N(1,1,\bar{b}_0), X_2^N(1,\tilde{w}_{2pa}), \\ &\quad U_{2pb}^N(1,\tilde{w}_{2pa},1,\bar{b}_0,\tilde{w}_{2pb},b_2)) \in T_\epsilon^N\left(p_{Y_2,U_{2c},U_{1c},X_2,U_{2pb}}\right)\Bigg] \\ &\leq 2^{N(R_{2pa}+R_{2pb}+R'_{2pb})} \sum_{(y_2^N,u_{2c}^N,u_{1c}^N,x_2^N,u_{2pb}^N)\in T_\epsilon^N\left(p_{Y_2,U_{2c},U_{1c},X_2,U_{2pb}}\right)} p_{2|\star}^N|_{\star=(U_{2c},U_{1c})} \\ &\leq 2^{N(R_{2pa}+R_{2pb}+R'_{2pb}-I_{2|\star}|_{\star=(U_{2c},U_{1c})})}\end{aligned}$$

for $p_{2|\star}$ given in (33) and $I_{2|\star}$ given in (34). Hence $\Pr[E_{2,2b}]\to 0$ as $N\to\infty$ if (11f) is satisfied.

- When the event $E_{2,3}$ occurs, i.e., either $E_{2,3a}$ or $E_{2,3b}$, we have $\hat{w}_{2c}=1$, $\hat{w}_{2pa}=1$ but $\hat{w}_{2pb}\neq 1$. Again, whether $\hat{w}_{1c}$ is correct or not, it does not matter since decoder 2 is not interested in $\hat{w}_{1c}$. However we need to consider whether the pair $(\hat{w}_{1c},\hat{b}_0)$ is equal to the transmitted one or not because this affects the way the joint probability among all involved RVs factorizes. The analysis proceeds as for the event $E_{2,2}$.
We have

— Case $E_{2,3a}$: either $\hat{w}_{1c}\neq 1$ or $\hat{b}_0\neq\bar{b}_0$. In this case, conditioned on the (correct) decoded sequences (U_{2c}^N, X_2^N), the output Y_2^N is independent of the (wrong) decoded sequences (U_{1c}^N, U_{2pb}^N). It is easy to see that the most stringent error event is when both $\hat{w}_{1c}\neq 1$ and $\hat{b}_0\neq\bar{b}_0$. Thus we have

$$\begin{aligned}\Pr[E_{2,3a}] &= P\Bigg[\bigcup_{\tilde{w}_{1c}\neq 1, b_0\neq\bar{b}_0, \tilde{w}_{2pb}, b_2} (Y_2^N, U_{2c}^N(1), U_{1c}^N(1,\tilde{w}_{1c},b_0), X_2^N(1,1), \\ &\quad U_{2pb}^N(1,1,\tilde{w}_{1c},b_0,\tilde{w}_{2pb},b_2)) \in T_\epsilon^N\left(p_{Y_2,U_{2c},U_{1c},X_2,U_{2pb}}\right)\Bigg] \\ &\leq 2^{N(R_{1c}+R'_{1c}+R_{2pb}+R'_{2pb})} \sum_{(y_2^N,u_{2c}^N,u_{1c}^N,x_2^N,u_{2pb}^N)\in T_\epsilon^N\left(p_{Y_2,U_{2c},U_{1c},X_2,U_{2pb}}\right)} p_{2|\star}^N|_{\star=(U_{2c},X_2)} \\ &\leq 2^{N(R_{2pa}+R_{1c}+R'_{1c}+R_{2pb}+R'_{2pb}-I_{2|\star}|_{\star=(U_{2c},X_2)})}\end{aligned}$$

for $p_{2|\star}$ given in (33) and $I_{2|\star}$ given in (34). Hence $\Pr[E_{2,3a}]\to 0$ as $N\to\infty$ if (11g) is satisfied.

— Case $E_{2,3b}$: both $\hat{w}_{1c}=1$ and $\hat{b}_0=\bar{b}_0$. In this case, conditioned on the (correct) decoded sequences $(U_{2c}^N, X_2^N, U_{1c}^N)$, the output Y_2^N is independent of the (wrong) decoded sequence U_{2pb}^N. However, since $(U_{2c}^N, X_2^N, U_{1c}^N)$ is the triplet that passed the encoding binning step, they are jointly typical. Hence, in this case we cannot use the factorization in $p_{2|\star}$ given in (33), but we need to replace $p_{U_{1c}|U_{2c}}$ in (33) with $p_{U_{1c}|U_{2c},X_2}$. Thus we have

$$\begin{aligned}\Pr[E_{2,3b}] &= P\Bigg[\bigcup_{\tilde{w}_{2pb}, b_2} Y_2^N, U_{2c}^N(1), U_{1c}^N(1,1,\bar{b}_0), X_2^N(1,1), \\ &\quad U_{2pb}^N(1,1,1,\bar{b}_0,\tilde{w}_{2pb},b_2)) \in T_\epsilon^N\left(p_{Y_2,U_{2c},U_{1c},X_2,U_{2pb}}\right)\Bigg] \\ &\leq 2^{N(R_{2pb}+R'_{2pb})} \sum_{(y_2^N,u_{2c}^N,u_{1c}^N,x_2^N,u_{2pb}^N)\in T_\epsilon^N\left(p_{Y_2,U_{2c},U_{1c},X_2,U_{2pb}}\right)} p_{U_{2c}}^N\, p_{X_2|U_{2c}}^N\, p_{U_{1c}|U_{2c},X_2}^N\, p_{U_{2pb}|U_{2c},U_{1c},X_2}^N\, p_{Y_2|U_{1c},U_{2c},X_2}^N \\ &\leq 2^{N(R_{2pb}+R'_{2pb}-I(Y_2;U_{2pb}|U_{1c},U_{2c},X_2))}.\end{aligned}$$

Hence $\Pr[E_{2,3b}]\to 0$ as $N\to\infty$ if (11h) is satisfied.

C) Decoding Errors at Decoder 1: The probability of error at decoder 1 is bounded as

$$\Pr[\text{error } 1|\text{encoding successful}] \leq \sum_{i=1}^{3}\Pr[E_{1,i}] \tag{35}$$

where $\Pr[E_{1,i}]$, for $i\in[1:3]$, is the error event defined in Table VI. The meaning of the symbols in Table VI is as for Table V.

TABLE VI
ERROR EVENTS AT DECODER 1

Event	w_{2c}	(w_{1c},b_1)	w_{1pb}	$p_{Y_1\|\star}$
$E_{1,1}$	X	$\cdots$	$\cdots$	p_{Y_1}
$E_{1,2}$	1	X	$\cdots$	$p_{Y_1\|U_{2c}}$
$E_{1,3}$	1	1	X	$P_{Y_1\|U_{2c},U_{1c}}$

Depending on which messages are wrongly decoded at decoder 1, the generated sequences and the received Y_1^N are generated iid according to

$$p_{1|\star} \triangleq p_{U_{2c}}\, p_{U_{1c}|U_{2c}}\, p_{U_{1pb}|U_{2c},U_{1c}}\, p_{Y_1|\star} \tag{36}$$

where "$\star$" indicates the messages decoded correctly. However, the actual transmitted sequences and the received Y_1^N considered at decoder 1 look as if they were generated iid according to

$$p_1 \triangleq p_{U_{2c}}\, p_{U_{1c}|U_{2c}}\, p_{U_{1pb}|U_{2c},U_{1c}}\, p_{Y_1|U_{2c},U_{1c},U_{1pb}}. \tag{37}$$

Hence we expect the probability of error at decoder 1 to depend on terms of the type

$$\begin{aligned}I_{1|\star} &= \mathbb{E}\left[\log\frac{p_1}{p_{1|\star}}\right] \\ &= \mathbb{E}\left[\log\frac{p_{Y_1|U_{2c},U_{1c},U_{1pb}}}{p_{Y_1|\star}}\right] \\ &= I(Y_1;U_{2c},U_{1c},U_{1pb}|\star).\end{aligned} \tag{38}$$

We now proceed to bound the probability of all the events in (35). We have that $\Pr[\text{error 1}|\text{encoding successful}] \to 0$ when $N \to \infty$ if

- When the event $E_{1,1}$ occurs we have $\hat{w}_{2c} \neq 1$. In this case the received Y_1^N is independent of the transmitted sequences. We can bound the error probability of $E_{1,1}$ as

$$\Pr[E_{1,1}] = P\Big[\bigcup_{\tilde{w}_{2c}\neq 1, \tilde{w}_{1c}, \tilde{w}_{1pb}, b_0, b_1} (Y_1^N, U_{2c}^N(\tilde{w}_{2c}), U_{1c}^N(\tilde{w}_{1c}, \tilde{w}_{2c}, b_0), U_{1pb}^N(\tilde{w}_{2c}, \tilde{w}_{2pa}, \tilde{w}_{1c}, b_0, \tilde{w}_{2pb}, b_1)) \in T_\epsilon^N\left(p_{Y_1,U_{2c},U_{1c},U_{1pb}}\right)\Big]$$
$$\leq 2^{N(R_{2c}+R_{1c}+R'_{1c}+R_{1pb}+R'_{1pb})} \sum_{(y_1^N, u_{2c}^N, u_{1c}^N, u_{1pb}^N)\in T_\epsilon^N\left(p_{Y_1,U_{2c},U_{1c},U_{1pb}}\right)} p_{1|\star}^N|_{\star=\emptyset}$$
$$\leq 2^{N(R_{2c}+R_{2pa}+R_{1c}+R'_{1c}+R_{2pb}+R'_{2pb}-I_{1|\star}|_{\star=\emptyset})}$$

for $p_{1|\star}$ given in (33) and $I_{1|\star}$ given in (38). Hence $\Pr[E_{1,1}] \to 0$ as $N \to \infty$ if (11i) is satisfied.

- When the event $E_{1,2}$ occurs, either $\hat{w}_{1c} \neq 1$, $\hat{b}_0 \neq \bar{b}_0$ or both. In this case, conditioned on the (correct) decoded sequence U_{2c}^N, the output Y_1^N is independent of the (wrong) decoded sequences (U_{1c}^N, U_{1pb}^N) . It is easy to see that the most stringent error event is when both $\hat{w}_{1c} \neq 1$ and $\hat{b}_0 \neq \bar{b}_0$. Thus we have

$$\Pr[E_{1,2}] = P\Bigg[\bigcup_{\tilde{w}_{1c}\neq 1, b_0\neq\bar{b}_0, \tilde{w}_{1pb}, b_1} (Y_1^N, U_{2c}^N(1), U_{1c}^N(1, \tilde{w}_{1c}, b_0), U_{1pb}^N(1, \tilde{w}_{1c}, b_0, \tilde{w}_{1pb}, b_1)) \in T_\epsilon^N\left(p_{Y_1,U_{2c},U_{1c},U_{1pb}}\right)\Bigg]$$
$$\leq 2^{N(R_{1c}+R'_{1c}+R_{1pb}+R'_{1pb})} \sum_{(y_1^N, u_{2c}^N, u_{1c}^N, u_{1pb}^N)\in T_\epsilon^N\left(p_{Y_1,U_{2c},U_{1c},U_{1pb}}\right)} p_{1|\star}^N|_{\star=U_{2c}}$$
$$\leq 2^{N(R_{1c}+R'_{1c}+R_{1pb}+R'_{1pb}-I_{1|\star}|_{\star=U_{2c}})}$$

for $p_{1|\star}$ given in (37) and $I_{1|\star}$ given in (38). Hence $\Pr[E_{1,2}] \to 0$ as $N \to \infty$ if (11j) is satisfied.

- When the event $E_{1,3}$ occurs, either $\hat{w}_{1pb} \neq 1$ or $\hat{b}_1 \neq \bar{b}_1$ or both. In this case, conditioned on the (correct) decoded sequence (U_{2c}^N, U_{1c}^N), the output Y_1^N is independent of the (wrong) decoded sequences U_{1pb}^N. It is easy to see that the most stringent error event is when both $\hat{w}_{1pb} \neq 1$ and $\hat{b}_1 \neq \bar{b}_1$. Thus we have

$$\Pr[E_{1,3}] = P\Bigg[\bigcup_{\tilde{w}_{1pb}\neq 1, b_1\neq\bar{b}_1} (Y_1^N, U_{2c}^N(1), U_{1c}^N(1,1,\bar{b}_0), U_{1pb}^N(1,1,\bar{b}_0, \tilde{w}_{1pb}, b_1)) \in T_\epsilon^N\left(p_{Y_1,U_{2c},U_{1c},U_{1pb}}\right)\Bigg]$$
$$\leq 2^{N(R_{1pb}+R'_{1pb})} \sum_{(y_1^N, u_{2c}^N, u_{1c}^N, u_{1pb}^N)\in T_\epsilon^N\left(p_{Y_1,U_{2c},U_{1c},U_{1pb}}\right)} p_{1|\star}^N|_{\star=(U_{2c},U_{1c})}$$
$$\leq 2^{N(R_{1c}+R'_{1c}+R_{1pb}+R'_{1pb}-I_{1|\star}|_{\star=(U_{2c},U_{1c})})}$$

for $p_{1|\star}$ given in (37) and $I_{1|\star}$ given in (38). Hence $\Pr[E_{1,3}] \to 0$ as $N \to \infty$ if (11k) is satisfied.

Appendix C
Proof of Lemma 8

An encoding error is committed with sequential/two-step binning if we cannot find a b_0 in the first step or if, upon finding the correct b_0 in the first encoding step, we cannot find the correct (b_1, b_2) in the second step. Let $E_{e,0}$ the probability of the error at the first step and $E_{e,12}$ the probability of the error at the second step, the

$$\Pr[\text{encoding NOT successful}] \leq \Pr[E_{e,0}] + \Pr[E_{e,12}|E_{e,0}^c]$$

where

$$\Pr[E_{e,0}] = \Pr\left[\bigcap_{b_0=1}^{2^{NR'_{1c}}} \left(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0)\right) \notin T_\epsilon^N(p_{U_{2c},X_2,U_{1c}})\right]$$
$$= \left(1 - \Pr[\left(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0)\right) \notin T_\epsilon^N(p_{U_{2c},X_2,U_{1c}})]\right)^{2^{NR'_{1c}}}.$$

Using standard typicality arguments we have

$$\Pr[\left(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0)\right) \notin T_\epsilon^N(p_{U_{2c},X_2,U_{1c}})] \geq (1-\epsilon)2^{N(I(U_{1c};X_2|U_{2c})+\delta)}.$$

Now we can write

$$\Pr[E_{e,0}] \leq (1-(1-\epsilon)2^{N(I(U_{1c};X_2|U_{2c})+\delta)})^{2^{NR'_{1c}}}$$
$$\leq \exp\left(1-(1-\epsilon)2^{N(R'_{1c}-I(U_{1c};X_2|U_{2c})+\delta)}\right)$$

so that $\Pr[E_{e,0}] \to 0$ when $N \to 0$ if (12a) is satisfied.

The error event $E_{e,12}$ can be divided in three error events

- $E_{e,21\ a}$: it is not possible to find b_1 such that $(U_{2c}^N, X_2^N, U_{1c}^N, U_{1pb}^N) \in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb}})$
- $E_{e,21\ b}$: it is not possible to find b_2 such that $(U_{2c}^N, X_2^N, U_{1c}^N, U_{2pb}^N) \in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{2pb}})$.
- $E_{e,21\ c}$: given that we can find b_1 and b_2 satisfy the first two conditions, we cannot find a couple (b_1, b_2) such that $(U_{2c}^N, X_2^N, U_{1c}^N, U_{1pb}^N, U_{2pb}^N) \in T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}})$.

We now establish the conditions that guarantee that the probability of error of each of these events goes to zero as $N \to \infty$.

For $E_{e,21\ a}$ we have

$$P[E_{e,21\ a}] = (1 - \Pr[\left(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0), U_{1pb}^N(1,1,b_0,1,b_1)\right) \notin T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb}})])^{2^{NR'_{1pb}}}$$

where

$$\Pr[\left(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0), U_{1pb}^N(1,1,b_0,1,b_1)\right) \notin T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb}})] \geq (1-\epsilon)2^{-N(I(X_2;U_{1pb}|U_{2c},U_{1c})+\delta)}.$$

As for $E_{e,0}$, this implies that $\Pr[E_{e,21\ a}] \to 0$ when $N \to \infty$ if (12b) is satisfied.

For $E_{e,21\ b}$, we have that the probability of this event goes to zero for large N given that $(U_{2c}^N, X_2^N, U_{1c}^N)$ appear to be generated according to the distribution $p_{U_{2c},X_2,U_{1c}}$ and U_{2pb} is generated according to $p_{U_{2pb}|U_{2c},X_2,U_{1c}}$.

For $E_{e,21\ c}$ we have

$$P[E_{e,21\ c}] = (1 - \Pr[\big(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0), U_{1pb}^N(1,1,b_0,1,b_1), U_{1pb}^N(1,1,b_0,1,b_2)\big) \notin T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}})])^{2^{N(R'_{1pb}+R'_{2pb})}}$$

where

$$\Pr[\big(U_{2c}^N(1), X_2^N(1,1), U_{1c}^N(1,1,b_0), U_{1pb}^N(1,1,b_0,1,b_1), U_{2pb}^N(1,1,b_0,1,b_2)\big) \notin T_\epsilon^N(p_{U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}})] \geq (1-\epsilon)2^{-N(I(U_{1pb};X_2,U_{2pb}|U_{2c},U_{1c})+\delta)}.$$

This implies that $\Pr[E_{e,21\ c}] \to 0$ when $N \to \infty$ if (12c) is satisfied.

Appendix D
Containment of [16, Th. 1] in $\mathcal{R}_{\mathrm{RTD}}$

We refer to the region in [16, Th. 1] as $\mathcal{R}_{\mathrm{DMT}}$ for brevity. We show this inclusion of $\mathcal{R}_{\mathrm{DMT}}$ in $\mathcal{R}_{\mathrm{RTD}}$ with the following steps.

- We enlarge the region $\mathcal{R}_{\mathrm{DMT}}$ by removing some rate constraints.
- We further enlarge the region by enlarging the set of possible input distributions. This allows us to remove the V_{11} and Q from the inner bound. We refer to this region as $\mathcal{R}_{\mathrm{DMT}}^{\mathrm{out}}$ since is enlarges the original achievable rate region.
- We make a correspondence between the RVs and corresponding rates of $\mathcal{R}_{\mathrm{DMT}}^{\mathrm{out}}$ and $\mathcal{R}_{\mathrm{RTD}}$.
- We choose a particular subset of $\mathcal{R}_{\mathrm{RTD}}$, $\mathcal{R}_{\mathrm{RTD}}^{\mathrm{in}}$, for which we can more easily show $\mathcal{R}_{\mathrm{DMT}} \subseteq \mathcal{R}_{\mathrm{DMT}}^{\mathrm{out}} \subset \mathcal{R}_{\mathrm{RTD}}^{\mathrm{in}} \subseteq \mathcal{R}_{\mathrm{RTD}}$, since $\mathcal{R}_{\mathrm{DMT}}^{\mathrm{out}}$ and $\mathcal{R}_{\mathrm{RTD}}^{\mathrm{in}}$ have identical input distribution decompositions and similar rate bound equations.

Enlarge the region $\mathcal{R}_{\mathrm{DMT}}$

We first enlarge the rate region of [16, Th. 1] $\mathcal{R}_{\mathrm{DMT}}$ by removing a number of constraints (specifically, we remove (2.6, 2.8, 2.10, 2.13, 2.14, 2.16, 2.17) of [16, Th. 1]). Also, following the line of thoughts in [29, App. D], it is possible to show that without loss of generality we can set X_1 to be a deterministic function of V_{11} and V_{12}, allowing us insert X_1 next to V_{11}, V_{12}. With these consideration we can enlarge the original region and define $\mathcal{R}_{\mathrm{DMT}}^{\mathrm{out}}$ as in (39), taken over the union of distributions.

$$p_W p_{V_{11}} p_{V_{12}} p_{X_1|V_{11},V_{12}} p_{V_{21}|V_{11}V_{12}} p_{V_{22}|V_{11},V_{12}} p_{X_2|V_{11},V_{12},V_{21},V_{22}}. \tag{40}$$

For (39c) we have

$$\begin{aligned} R_{11} &\leq I(Y_1, V_{12}, V_{21}; V_{11}|W) \\ &= I(Y_1, V_{21}; V_{11}|V_{12}, W) + I(V_{12}; V_{11}|W) \\ &= I(Y_1, V_{21}; V_{11}|V_{12}, W) \\ &= I(Y_1, V_{21}; X_1, V_{11}|V_{12}, W) \\ &= I(Y_1; X_1, V_{11}|V_{12}, V_{21}, W) + I(V_{21}; X_1, V_{11}|V_{12}, W). \end{aligned}$$

For (39e) we have

$$\begin{aligned} R_{11} + R_{21} + R'_{21} &\leq I(Y_1, V_{12}; V_{11}, V_{21}|W) \\ &\quad + I(V_{11}; V_{21}|W) \\ &= I(Y_1; V_{11}, V_{21}|V_{12}, W) + I(V_{12}; V_{11}, V_{21}|W) \\ &\quad + I(V_{11}; V_{21}|W) \\ &= I(Y_1; V_{11}, V_{21}|V_{12}, W) + I(V_{12}; V_{21}|V_{11}, W) \\ &\quad + I(V_{11}; V_{21}|W) \\ &= I(Y_1; V_{11}, V_{21}|V_{12}, W) + I(V_{11}, V_{12}; V_{21}|W) \\ &= I(Y_1; X_1, V_{11}, V_{21}|V_{12}, W) + I(X_1, V_{11}, V_{12}; V_{21}|W). \end{aligned}$$

The original region is thus equivalent to the region in (41), taken over the union of all distributions that factor as in (40).

Enlarge the class of input distribution and eliminate V_{11} and W. Now increase the set of possible input distributions of equation (40) by letting V_{11} have any joint distribution with V_{12}.

$$R'_{21} = I(V_{21}; X_1, V_{11}, V_{12}|W) \tag{39a}$$
$$R'_{22} = I(V_{22}; X_1, V_{11}, V_{12}|W) \tag{39b}$$
$$R_{11} \leq I(Y_1, V_{12}, V_{21}; V_{11}|W) \tag{39c}$$
$$R_{21} + R'_{21} \leq I(Y_1, X_1, V_{11}, V_{12}; V_{21}|W) \tag{39d}$$
$$R_{11} + R_{21} + R'_{21} \leq I(Y_1, V_{12}; V_{11}, V_{21}|W) + I(V_{11}; V_{21}|W) \tag{39e}$$
$$R_{11} + R_{21} + R'_{21} + R_{12} \leq I(Y_1; X_1, V_{11}, V_{12}, V_{21}|W) + I(X_1, V_{11}, V_{12}; V_{21}|W) \tag{39f}$$
$$R_{22} + R'_{22} \leq I(Y_2, V_{12}, V_{21}; V_{22}|W) \tag{39g}$$
$$R_{22} + R'_{22} + R_{21} + R'_{21} \leq I(Y_2, V_{12}; V_{22}, V_{21}|W) + I(V_{22}; V_{21}|W) \tag{39h}$$
$$R_{22} + R'_{22} + R_{21} + R'_{21} + R_{12} \leq I(Y_2; V_{22}, V_{21}, V_{12}|W) + I(V_{22}, V_{21}; V_{12}|W) \tag{39i}$$

TABLE VII
ASSIGNMENT OF RVS OF APPENDIX C

RV, rate of Theorem 7	RV, rate of [16, Thm. 1]	Comments
U_{2c}, R_{2c}	V_{12}, R_{12}	TX 2 → RX 1, RX 2
U_{1c}, R_{1c}	V_{21}, R_{21}	TX 1 → RX 1, RX 2
U_{1pb}, R_{1pb}	V_{22}, R_{22}	TX 1 → RX 1
X_2, R_{2pa}	X_1', R_{11}	TX 2 → RX 2
$U_{2pb} = \emptyset, R'_{2pb} = 0$	–	TX 1 → RX 2
$R'_{1c} = I(U_{1c}; X_2 \vert U_{2c})$	$L_{21} - R_{21} = I(V_{21}; V_{11}, V_{12})$	Binning rate
$R'_{1pb} = I(U_{1pb}; X_2 \vert U_{1c}, U_{2c})$	$L_{22} - R_{22} = I(V_{22}; V_{11}, V_{12})$	Binning rate
X_1	X_2	

This is done by substituting $p_{V_{11}}$ with $p_{V_{11}|V_{12}}$ in the expression of the input distribution. With this substitution we have

$$\begin{aligned} & p_W p_{V_{11}|V_{12}} p_{V_{12}} p_{X_1|V_{11},V_{12}} p_{V_{21}|X_1,V_{11}V_{12}} p_{V_{22}|X_1,V_{11},V_{12}} \\ &\quad p_{X_2|X_1,V_{11},V_{12},V_{21},V_{22}} \\ &= p_W p_{V_{12}} p_{V_{11},X_1|V_{12}} p_{V_{21}|X_1,V_{11}V_{12}} p_{V_{22}|X_1,V_{11},V_{12}} \\ &\quad p_{X_2|X_1,V_{11},V_{12},V_{21},V_{22}} \\ &= p_W p_{V_{12}} p_{X_1'|V_{12}} p_{V_{21}|X_1',V_{12}} p_{V_{22}|X_1',V_{12}} \\ &\quad p_{X_2|X_1',V_{12},V_{21},V_{22}} \end{aligned}$$

with $X_1' = (X_1, V_{11})$. Since V_{12} is decoded at both decoders, the time sharing random W may be incorporated with V_{12} without loss of generality and thus can be dropped. The region described in (41) is convex and thus time sharing is not needed. With these simplifications, the region $\mathcal{R}_{\mathrm{DMT}}^{\mathrm{out}}$ is now defined as the region in (42), taken over the union of all distributions

$$p_{V_{12}} p_{X_1'|V_{12}} p_{V_{21}|X_1',V_{12}} p_{V_{22}|X_1',V_{12}} p_{X_2|X_1',V_{12},V_{21},V_{22}}.$$

Correspondence between the random variables and rates. When referring to [16] please note that the index of the primary and cognitive user are reversed with respect to our notation (i.e., $1 \to 2$ and vice-versa). Consider the correspondences between the variables of [16, Th. 1] and those of Theorem 7 in Table VII to obtain the region $\mathcal{R}_{\mathrm{DMT}}^{\mathrm{out}}$ defined as the set of rate pairs satisfying the inequalities in (43), taken over the union of all distributions

$$p_{U_{2c}} p_{X_2|U_{2c}} p_{U_{1c}|X_2} p_{U_{1pb}|X_2} p_{X_1|X_2,U_{1c},U_{1pb}}. \tag{44}$$

$$\begin{aligned} R'_{21} &= I(V_{21}; X_1, V_{11}, V_{12}|W) && \text{(41a)} \\ R'_{22} &= I(V_{22}; X_1, V_{11}, V_{12}|W) && \text{(41b)} \\ R_{11} &\le I(Y_1; X_1, V_{11}|V_{12}, V_{21}|W) + I(V_{21}; X_1|V_{12}, W) && \text{(41c)} \\ R_{21} + R'_{21} &\le I(Y_1, X_1, V_{11}, V_{12}; V_{21}|W) && \text{(41d)} \\ R_{11} + R_{21} + R'_{21} &\le I(Y_1; X_1, V_{11}, V_{21}|V_{12}, W) + I(X_1; V_{21}|W) && \text{(41e)} \\ R_{11} + R_{21} + R'_{21} + R_{12} &\le I(Y_1; X_1, V_{11}, V_{21}, V_{12}|W) + I(X_1, V_{11}, V_{12}; V_{21}|W) && \text{(41f)} \\ R_{22} + R'_{22} &\le I(Y_2, V_{12}, V_{21}; V_{22}|W) && \text{(41g)} \\ R_{22} + R'_{22} + R_{21} + R'_{21} &\le I(Y_2, V_{12}; V_{22}, V_{21}|W) + I(V_{22}; V_{21}|W) && \text{(41h)} \\ R_{22} + R'_{22} + R_{21} + R'_{21} + R_{12} &\le I(Y_2; V_{22}, V_{21}, V_{12}|W) + I(V_{22}, V_{21}; V_{12}|W) && \text{(41i)} \end{aligned}$$

$$\begin{aligned} R'_{21} &= I(V_{21}; X_1', V_{12}) && \text{(42a)} \\ R'_{22} &= I(V_{22}; X_1', V_{12}) && \text{(42b)} \\ R_{11} &\le I(Y_1; X_1'|V_{12}, V_{21}) + I(V_{21}; X_1|V_{12}) && \text{(42c)} \\ R_{21} + R'_{21} &\le I(Y_1, X_1', V_{12}; V_{21}) && \text{(42d)} \\ R_{11} + R_{21} + R'_{21} &\le I(Y_1; X_1', V_{21}|V_{12}) + I(X_1; V_{21}) && \text{(42e)} \\ R_{11} + R_{21} + R'_{21} + R_{12} &\le I(Y_1; X_1', V_{21}, V_{12}) + I(X_1', V_{12}; V_{21}) && \text{(42f)} \\ R_{22} + R'_{22} &\le I(Y_2, V_{12}, V_{21}; V_{22}) && \text{(42g)} \\ R_{22} + R'_{22} + R_{21} + R'_{21} &\le I(Y_2, V_{12}; V_{22}, V_{21}) + I(V_{22}; V_{21}) && \text{(42h)} \\ R_{22} + R'_{22} + R_{21} + R'_{21} + R_{12} &\le I(Y_2; V_{22}, V_{21}, V_{12}) + I(V_{22}, V_{21}; V_{12}) && \text{(42i)} \end{aligned}$$

Next, we using the correspondences of the table and restrict the fully general input distribution of Theorem 7 to match the more constrained factorization of (44), obtaining a region $\mathcal{R}_{\rm RTD}^{\rm in} \subseteq \mathcal{R}_{\rm RTD}$ defined as the set of rate tuples satisfying the inequalities in (45), taken over the union of all distributions that factor as

$$p_{U_{2c},X_2} p_{U_{1c}|X_2} p_{U_{1pb}|X_2} p_{X_1|X_2,U_{1c},U_{1pb}}.$$

Equation-by-equation comparison. We now show that $\mathcal{R}_{\rm DMT}^{\rm out} \subseteq \mathcal{R}_{\rm RTD}^{\rm in}$ by fixing an input distribution (which are the same for these two regions) and comparing the rate regions equation by equation. We refer to the equation numbers directly and look at the difference between the corresponding equations in the two new regions.

- Equations (45c)–(45a) versus (43c)–(43a): Noting the cancelation/interplay between the binning rates, we see that

$$((45\text{c}) - (45\text{a})) - ((43\text{d}) - (43\text{a})) = 0.$$

- Equations (45d)–(45a) versus (43d)–(43a):

$$\begin{aligned}
&((45\text{d}) - (45\text{a})) - ((43\text{d}) - (43\text{a}))\\
&\quad = -I(X_2; U_{1c}) + I(U_{1c}; X_2, U_{2c})\\
&\quad = I(U_{2c}; U_{1c}|X_2)\\
&\quad = 0
\end{aligned}$$

- Equations (45e)–(45a) versus (43e)–(43a): again noting the cancellations,

$$((45\text{e}) - (45\text{a})) - ((43\text{e}) - (43\text{a})) = 0$$

- Equations (45f) versus (43f):

$$(45\text{f}) - (43\text{f}) = 0$$

- Equations (45g)–(45b) versus (43g)–(43b)–(43a),

$$\begin{aligned}
&((45\text{g}) - (45\text{b})) - ((43\text{g}) - (43\text{b}) - (43\text{a}))\\
&\quad = -I(X_2; U_{1c}, U_{1pb}|U_{2c})\\
&\qquad - I(U_{1pb}, U_{1c}; U_{2c}) + I(U_{1c}; U_{2c}, X_2)\\
&\qquad + I(U_{1pb}; U_{2c}, X_2)\\
&\quad = -I(U_{1pb}, U_{1c}; X_2, U_{2c}) + I(U_{1c}; U_{2c}, X_2)\\
&\qquad + I(U_{1pb}; U_{2c}, X_2)\\
&\quad = -I(U_{1pb}; X_2, U_{2c}) - I(U_{1c}; X_2, U_{2c}|U_{1pb})\\
&\qquad + I(U_{1c}; U_{2c}, X_2) + I(U_{1pb}; U_{2c}, X_2)\\
&\quad = -I(U_{1c}; X_2, U_{2c}|U_{1pb}) + I(U_{1c}; U_{2c}, X_2)\\
&\quad = -H(U_{1c}|U_{1pb}) + H(U_{1c}|X_2, U_{2c}, U_{1pb})\\
&\qquad + H(U_{1c}) - H(U_{1c}|X_2, U_{2c})\\
&\quad = I(U_{1c}; U_{1pb}) > 0
\end{aligned}$$

$$\begin{aligned}
R'_{1c} &= I(U_{1c}; X_2, U_{2c}) && (43\text{a})\\
R'_{1pb} &= I(U_{1pb}; X_2, U_{2c}) && (43\text{b})\\
R_{2pa} + R_{1c} + R'_{1c} + R_{2c} &\le I(Y_2; U_{1c}, U_{2c}, X_2) + I(X_2, U_{2c}; U_{1c}) && (43\text{c})\\
R_{2pa} + R_{1c} + R'_{1c} &\le I(Y_2; X_2, U_{1c}|U_{2c}) + I(X_2; U_{1c}) && (43\text{d})\\
R_{1c} + R'_{1c} &\le I(Y_2, X_2, U_{2c}; U_{1c}) && (43\text{e})\\
R_{2pa} &\le I(Y_2; X_2|U_{2c}, U_{1c}) + I(U_{1c}; X_2|U_{2c}) && (43\text{f})\\
R_{1pb} + R'_{1pb} + R_{1c} + R'_{1c} + R_{2c} &\le I(Y_1; U_{1pb}, U_{1c}, U_{2c}) + I(U_{1pb}, U_{1c}; U_{2c}) && (43\text{g})\\
R_{1c} + R_{1pb} + R'_{1c} + R'_{1pb} &\le I(Y_1, U_{2c}; U_{1pb}, U_{1c}) + I(U_{1pb}; U_{1c}) && (43\text{h})\\
R_{1pb} + R'_{1pb} &\le I(Y_1, U_{2c}, U_{1c}; U_{1pb}) && (43\text{i})
\end{aligned}$$

$$\begin{aligned}
R'_{1c} &= I(U_{1c}; X_2|U_{2c}) && (45\text{a})\\
R'_{1c} + R'_{1pb} &= I(X_2; U_{1c}, U_{1pb}|U_{2c}) && (45\text{b})\\
R_{2c} + R_{1c} + R_{2pa} + R'_{1c} &\le I(Y_2; U_{2c}, U_{1c}, X_2) + I(U_{1c}; X_2|U_{2c}) && (45\text{c})\\
R_{2pa} + R_{1c} + R'_{1c} &\le I(Y_2; U_{1c}, X_2|U_{2c}) + I(U_{1c}; X_2|U_{2c}) && (45\text{d})\\
R_{1c} + R'_{1c} &\le I(Y_2; U_{1c}|U_{2c}, X_2) + I(U_{1c}; X_2|U_{2c}) && (45\text{e})\\
R_{2pa} &\le I(Y_2; X_2|U_{2c}, U_{1c}) + I(U_{1c}; X_2|U_{2c}) && (45\text{f})\\
R_{1pb} + R'_{1pb} + R_{1c} + R'_{1c} + R_{2c} &\le I(Y_1; U_{2c}, U_{1c}, U_{1pb}) && (45\text{g})\\
R_{1c} + R_{1pb} + R'_{1c} + R'_{1pb} &\le I(Y_1; U_{1c}, U_{1pb}|U_{2c}) && (45\text{h})\\
R_{1pb} + R'_{1pb} &\le I(Y_1; U_{1pb}|U_{2c}, U_{1c}) && (45\text{i})
\end{aligned}$$

where we have used the fact that U_{1c} and U_{1pb} are conditionally independent given (U_{2c}, X_2).

- Equations (45i)–(45b)–(45a) versus (43i)–(43b):

$$
\begin{aligned}
&((45h) - (45b)) - ((45h) - (45b) - (45a)) \\
&= -I(X_2; U_{1c}, U_{1pb}|U_{2c}) - I(U_{2c}; U_{1c}, U_{1pb}) \\
&\quad + I(U_{1pb}; U_{2c}, X_2) - I(U_{1pb}; U_{1c}) \\
&\quad + I(U_{1c}; X_2, U_{2c}) \\
&= -I(X_2, U_{2c}; U_{1c}, U_{1pb}) + I(U_{1pb}; U_{2c}, X_2) \\
&\quad - I(U_{1pb}; U_{1c}) + I(U_{1c}; X_2, U_{2c}) \\
&= -I(X_2, U_{2c}; U_{1pb}) - I(U_{1c}; X_2, U_{2c}|U_{1pb}) \\
&\quad + I(U_{1pb}; U_{2c}, X_2) - I(U_{1pb}; U_{1c}) \\
&\quad + I(U_{1c}; X_2, U_{2c}) \\
&= -I(U_{1c}; X_2, U_{2c}, U_{1pb}) + I(U_{1c}; X_2, U_{2c}) \\
&= -I(U_{1c}; X_2, U_{2c}) - I(U_{1c}; U_{1pb}|X_2, U_{2c}) \\
&\quad + I(U_{1c}; X_2, U_{2c}) \\
&= 0
\end{aligned}
$$

where we have used the fact that U_{1c} and U_{1pb} are conditionally independent given (U_{2c}, X_2).

- Equations (45i)–(45b) versus (43i)–(43b)—(45a)(45a):

$$
\begin{aligned}
&((45i) - (45b) + (45a)) - ((43i) - (43b)) \\
&= -I(U_{1pb}; X_2|U_{2c}, U_{1c}) - I(U_{1pb}; U_{2c}, U_{1c}) \\
&\quad + I(U_{1pb}; X_2, U_{2c}) \\
&= -I(U_{1pb}; X_2, U_{2c}, U_{1c}) + I(U_{1pb}; U_{2c}, X_2) \\
&= -I(U_{1pb}; U_{1c}|U_{2c}, X_2) \\
&= 0
\end{aligned}
$$

APPENDIX E
CONTAINMENT OF [12, TH. 2] IN $\mathcal{R}_{\mathrm{RTD}}$

The independently derived region in [12, Th. 2] uses a similar encoding structure as that of $\mathcal{R}_{\mathrm{RTD}}$ with two exceptions: a) the binning is done sequentially rather than jointly as in $\mathcal{R}_{\mathrm{RTD}}$ leading to binning constraints (43)–(45) in [12, Th. 2] as opposed to (11a)–(11c) in Theorem 7 . Notable is that both schemes have adopted a Marton-like binning scheme at the cognitive transmitter, as first introduced in the context of the CIFC in [12]. b) While the cognitive messages are rate-split in identical fashions, the primary message is split into two parts in [12, Th. 2] ($R_1 = R_{11} + R_{10}$, note the reversal of indices) while we explicitly split the primary message into three parts $R_2 = R_{2c} + R_{2pa} + R_{2pb}$. We show that the region of [12, Th. 2], denoted as $\mathcal{R}_{\mathrm{CC}} \subseteq \mathcal{R}_{\mathrm{RTD}}$ in two steps.

- We first show that we may without loss of generality set $U_{11} = \emptyset$ in [12, Th. 2], creating a new region $\mathcal{R}'_{\mathrm{CC}}$.
- We next make a correspondence between our RVs and those of [12, Th. 2] and obtain identical regions.

We note that the primary and cognitive indices are permuted in [12].

We first show that U_{11} in [12, Th. 2] may be dropped without loss of generality. Consider the region $\mathcal{R}_{\mathrm{CC}}$ of [12, Th. 2], defined as the union over all distributions $p_{U_{10},U_{11},V_{11},V_{20},V_{22},X_1,X_2} p_{Y_1,Y_2|X_1,X_2}$ of all rate tuples satisfying

$$
\begin{aligned}
R_1 &\le I(Y_1; V_{11}, U_{11}, V_{20}, U_{10}) && (46a)\\
R_2 &\le I(Y_2; V_{20}, V_{22}|U_{10}) \\
&\quad - I(V_{22}, V_{20}; U_{11}|U_{10}) && (46b)\\
R_1 + R_2 &\le I(Y_1; V_{11}, U_{11}|V_{20}, U_{10}) \\
&\quad + I(Y_2; V_{22}, V_{20}, U_{10}) \\
&\quad - I(V_{22}; U_{11}, V_{11}|V_{20}, U_{10}) && (46c)\\
R_1 + R_2 &\le I(Y_1; V_{11}, U_{11}, V_{20}, U_{10}) \\
&\quad + I(Y_2; V_{22}|V_{20}, U_{10}) \\
&\quad - I(V_{22}; U_{11}, V_{11}|V_{20}, U_{10}) && (46d)\\
2R_2 + R_1 &\le I(Y_1; V_{11}, U_{11}, V_{20}|U_{10}) \\
&\quad + I(Y_2; V_{22}|V_{20}, U_{10}) \\
&\quad + I(Y_2; V_{20}, V_{22}, U_{10}) \\
&\quad - I(V_{22}, V_{20}; U_{11}|U_{10}) \\
&\quad - I(V_{22}; U_{11}, V_{11}|V_{20}, U_{10}). && (46e)
\end{aligned}
$$

Now let $\mathcal{R}'_{\mathrm{CC}}$ be the region obtained by setting $U'_{11} = \emptyset$ and $V'_{11} = (V_{11}, U_{11})$ while keeping all remaining RVs identical. Then $\mathcal{R}'_{\mathrm{CC}}$ is the union over all distributions $p_{U_{10},V'_{11},V_{20},V_{22},X_1,X_2} p_{Y_1,Y_2|X_1,X_2}$, with $V'_{11} = (V_{11}, U_{11})$ in $\mathcal{R}_{\mathrm{CC}}$, of all rate tuples satisfying

$$
\begin{aligned}
R_1 &\le I(Y_1; V_{11}, U_{11}, V_{20}, U_{10}) && (47a)\\
R_2 &\le I(Y_2; V_{20}, V_{22}|U_{10}) && (47b)\\
R_1 + R_2 &\le I(Y_1; V_{11}, U_{11}|V_{20}, U_{10}) + I(Y_2; V_{22}, V_{20}, U_{10}) \\
&\quad - I(V_{22}; U_{11}, V_{11}|V_{20}, U_{10}) && (47c)\\
R_1 + R_2 &\le I(Y_1; V_{11}, U_{11}, V_{20}, U_{10}) + I(Y_2; V_{22}|V_{20}, U_{10}) \\
&\quad - I(V_{22}; U_{11}, V_{11}|V_{20}, U_{10}) && (47d)\\
2R_2 + R_1 &\le I(Y_1; V_{11}, U_{11}, V_{20}|U_{10}) + I(Y_2; V_{22}|V_{20}, U_{10}) \\
&\quad + I(Y_2; V_{20}, V_{22}, U_{10}) - I(V_{22}; U_{11}, V_{11}|V_{20}, U_{10}). && (47e)
\end{aligned}
$$

Comparing the two regions equation by equation, we see that

- Equation (46a) = (47a).
- Equation (46b) < (47b) as this choice of RVs sets the generally positive mutual information to 0.
- Equation (46c) = (47c).
- Equation (46d) = (47d).
- Equation (46e) < (47e) as this choice of RVs sets the generally positive mutual information to 0.

From the previous, we may set $U_{11} = \emptyset$ in the region $\mathcal{R}_{\mathrm{CC}}$ of [12, Th. 1] without loss of generality, obtaining the region $\mathcal{R}'_{\mathrm{CC}}$ defined in (47a) –(47e). We show that $\mathcal{R}'_{\mathrm{CC}}$ may be obtained from the region $\mathcal{R}_{\mathrm{RTD}}$ with the assigment of RVs, rates and binning rates in Table VIII.

Evaluating $\mathcal{R}'_{\mathrm{CC}}$ defined by (47a) –(47e) with the above assignment, translating all RVs into the notation used here, we obtain the region

$$
\begin{aligned}
R'_{1c} &\ge 0 && (48a)\\
R'_{1pb} + R'_{2pb} &\ge I(U_{1pb}; U_{2pb}|U_{2c}, U_{1c}) && (48b)
\end{aligned}
$$

TABLE VIII
ASSIGNMENT OF RVs OF SECTION V-B

RV, rate of Theorem 7	RV, rate of [12, Thm. 1]	Comments
U_{2c}, R_{2c}	U_{10}, R_{10}	TX 2 $\rightarrow$ RX 1, RX 2
$X_2 = U_{2c}, R_{2pa} = 0$	$U_{11} = \emptyset$	TX 2 $\rightarrow$ RX 2
U_{1c}, R_{1c}	V_{20}, R_{20}	TX 1 $\rightarrow$ RX 1, RX 2
U_{1pb}, R_{1pb}	V_{22}, R_{22}	TX 1 $\rightarrow$ RX 1
U_{2pb}, R_{2pb}	V_{11}, R_{11}	TX 1 $\rightarrow$ RX 2
R'_{1c}	$L_{20} - R_{20}$	
R'_{1pb}	$L_{22} - R_{22}$	
R'_{2pb}	$L_{11} - R_{11}$	
X_1	X_2	
X_2	X_1	

$$R_{2pb} + R'_{2pb} \leq I(Y_2; U_{2pb}|U_{2c}, U_{1c}) \tag{48c}$$

$$R_{2pb} + R'_{2pb} + R_{1c} + R'_{1c} \leq I(Y_2; U_{1c}, U_{2pb}|U_{2c}) \tag{48d}$$

$$R_{2pb} + R'_{2pb} + R_{1c} + R'_{1c} + R_{2c} \leq I(Y_2; U_{1c}, U_{2c}, U_{2pb}) \tag{48e}$$

$$R_{1pb} + R'_{1pb} \leq I(Y_1; U_{1pb}|U_{2c}, U_{1c}) \tag{48f}$$

$$R_{1pb} + R'_{1pb} + R_{1c} + R'_{1c} \leq I(Y_1; U_{1pb}, U_{1c}|U_{2c}) \tag{48g}$$

$$R_{1pb} + R'_{1pb} + R_{1c} + R'_{1c} + R_{2c} \leq I(Y_1; U_{1pb}, U_{1c}, U_{2c}). \tag{48h}$$

Note that we may take binning rate equations $R'_{1c} \geq 0$ and $R'_{1pb} + R'_{2pb} \geq I(U_{1pb}; U_{2pb}|U_{2c}, U_{1c})$ to be equality without loss of generality—the largest region will take $R'_{1c}, R'_{1pb}, R'_{2pb}$ as small as possible. The region $\mathcal{R}_{\text{RTD}}$ with $R_{2pa} = 0$

$$R'_{1c} \geq 0 \tag{49a}$$

$$R'_{1c} + R'_{1pb} \geq 0 \tag{49b}$$

$$R'_{1c} + R'_{1pb} + R'_{2pb} \geq I(U_{1pb}; U_{2pb}|U_{2c}, U_{1c}) \tag{49c}$$

$$R_{2pb} + R'_{2pb} \leq I(Y_2; U_{2pb}|U_{2c}, U_{1c}) \tag{49d}$$

$$R_{2pb} + R'_{2pb} + R_{1c} + R'_{1c} \leq I(Y_2; U_{1c}, U_{2pb}|U_{2c}) \tag{49e}$$

$$R_{2pb} + R'_{2pb} + R_{1c} + R'_{1c} + R_{2c} \leq I(Y_2; U_{1c}, U_{2c}, U_{2pb}) \tag{49f}$$

$$R_{1pb} + R'_{1pb} \leq I(Y_1; U_{1pb}|U_{2c}, U_{1c}) \tag{49g}$$

$$R_{1pb} + R'_{1pb} + R_{1c} + R'_{1c} \leq I(Y_1; U_{1pb}, U_{1c}|U_{2c}) \tag{49h}$$

$$R_{1pb} + R'_{1pb} + R_{1c} + R'_{1c} + R_{2c} \leq I(Y_1; U_{1pb}, U_{1c}, U_{2c}). \tag{49i}$$

For $R'_{1c} = 0$ these two regions are identical, showing that $\mathcal{R}_{\text{RTD}}$ is surely no smaller than $\mathcal{R}_{\text{CC}}$. For $R'_{1c} > 0, \mathcal{R}_{\text{RTD}}$, the binning rates of the region $\mathcal{R}_{\text{RTD}}$ are looser than the ones in $\mathcal{R}_{\text{CC}}$. This is probably due to the fact that the first one uses joint binning and latter one sequential binning. Therefore $\mathcal{R}_{\text{RTD}}$ may produce rates larger than $\mathcal{R}_{\text{CC}}$. However, in general, no strict inclusion of $\mathcal{R}_{\text{CC}}$ in $\mathcal{R}_{\text{RTD}}$ has been shown.

APPENDIX F
CONTAINMENT OF [13, TH. 4.1] IN $\mathcal{R}_{\text{RTD}}$

In this scheme, the common messages are created independently instead of having the common message from transmitter 1 being superposed to the common message from transmitter 2. The former choice introduces more rate constraints than the latter and allows us to show inclusion in $\mathcal{R}_{\text{RTD}}$.

Again, following the argument of [29, App. D], we can show that without loss of generality we can take X_1 and X_2 to be deterministic functions. With this consideration we can express the region of [13, Th. 4.1] as

$$R'_{22} \geq I(W_2; V_1, X_1|U_1, U_2) \tag{50a}$$

$$R'_{11} + R'_{22} \geq I(W_2; W_1, V_1, X_1|U_1, U_2) \tag{50b}$$

$$R_{11} + R'_{11} \leq I(V_1, X_1, W_1; Y_1|U_1, U_2) \tag{50c}$$

$$R_{12} + R_{11} + R'_{11} \leq I(U_1, V_1, X_1, W_1; Y_1|U_2) \tag{50d}$$

$$R_{21} + R_{11} + R'_{11} \leq I(U_2, V_1, X_1, W_1; Y_1|U_1) \tag{50e}$$

$$R_{12} + R_{21} + R_{11} + R'_{11} \leq I(U_1, V_1, X_1, W_1, U_2; Y_1) \tag{50f}$$

$$R_{22} + R'_{22} \leq I(W_2; Y_2|U_1, U_2) \tag{50g}$$

$$R_{21} + R_{22} + R'_{22} \leq I(U_2, W_2; Y_2|U_1) \tag{50h}$$

$$R_{12} + R_{22} + R'_{22} \leq I(U_1, W_2; Y_2|U_2) \tag{50i}$$

$$R_{12} + R_{21} + R_{22} + R'_{22} \leq I(U_1, U_2, W_2; Y_2) \tag{50j}$$

taken over the union of all distributions

$$p_{U_1} p_{V_1|U_1} p_{X_1|V_1,U_1} p_{U_2} p_{W_1,W_2|V_1,U_1,U_2} p_{X_0|W_1,W_2,V_1,U_1,U_2} p_{Y_1,Y_2|X_1,X_0}$$

for $(R'_{11}, R'_{22}, R_{11}, R_{12}, R_{21}, R_{22}) \in \mathbb{R}^6_+$.

We can now eliminate one RV by noticing that

$$\begin{aligned} &p_{U_1} p_{V_1|U_1} p_{X_1|V_1,U_1} p_{U_2} p_{W_1,W_2|V_1,U_1,U_2} p_{X_0|W_1,W_2,V_1,U_1,U_2} p_{Y_1,Y_2|X_1,X_0} \\ &= p_{U_1} p_{V_1,X_1|U_1} p_{U_2} p_{W_1,W_2|V_1,U_1,X_1,U_2} p_{X_0|W_1,W_2,V_1,U_1,X_1,U_2} p_{Y_1,Y_2|X_1,X_0} \end{aligned}$$

and setting $V'_1 = [V_1, X_1]$, to obtain the region

$$R'_{22} \geq I(W_2; V'_1|U_1, U_2) \tag{51a}$$

$$R'_{11} + R'_{22} \geq I(W_2; W_1, V'_1|U_1, U_2) \tag{51b}$$

$$R_{11} + R'_{11} \leq I(V'_1, W_1; Y_1|U_1, U_2) \tag{51c}$$

$$R_{12} + R_{11} + R'_{11} \leq I(U_1, V'_1, W_1; Y_1|U_2) \tag{51d}$$

$$R_{21} + R_{11} + R'_{11} \leq I(U_2, V'_1, W_1; Y_1|U_1) \tag{51e}$$

$$R_{12} + R_{21} + R_{11} + R'_{11} \leq I(U_1, V'_1, W_1, U_2; Y_1) \tag{51f}$$

$$R_{22} + R'_{22} \leq I(W_2; Y_2|U_1, U_2) \tag{51g}$$

$$R_{21} + R_{22} + R'_{22} \leq I(U_2, W_2; Y_2|U_1) \tag{51h}$$

TABLE IX
ASSIGNMENT OF RVS OF APPENDIX F

RV, rate of Theorem 7	RV, rate of [13, Th.4.1] in (51)	Comments
U_{2c}, R_{2c}	U_1, R_{12}	TX 2 → RX 1, RX 2
X_2, R_{2pa}	V_1', R_{11}	TX 2 → RX 2
U_{1c}, R_{1c}	U_2, R_{21}	TX 1 → RX 1, RX 2
U_{1pb}, R_{1pb}	W_2, R_{22}	TX 1 → RX 1
$U_{2pb}, R_{2pb}=0$	W_1	TX 1 → RX 2
R_{1c}'	$L_{20}-R_{20}$	
R_{1pb}'	$R_{22}'=L_{22}-R_{22}$	
R_{2pb}'	$R_{11}'=L_{11}-R_{11}$	
X_1	X_0	
X_2	X_1	

$$R_{12}+R_{22}+R_{22}' \le I(U_1,W_2;Y_2|U_2) \quad (51i)$$
$$R_{12}+R_{21}+R_{22}+R_{22}' \le I(U_1,U_2,W_2;Y_2) \quad (51j)$$

taken over the union of all distributions of the form

$$p_{U_1}p_{V_1'|U_1}p_{U_2}p_{W_1,W_2|V_1',U_1,U_2}p_{X_0|W_1,W_2,V_1',U_1,U_2}p_{Y_1,Y_2|V_1',X_0}.$$

We equate the RVs in the region of [13] with the RVs in Theorem 7 as in Table IX.

With the substitutions of Table IX in the achievable rate region of (51), we obtain the region

$$R_{1pb}' \ge I(U_{1pb};X_2|U_{2c},U_{1c}) \quad (52a)$$
$$R_{1pb}'+R_{2pb}' \ge I(U_{1pb};U_{2pb},X_2|U_{2c},U_{1c}) \quad (52b)$$
$$R_{2pa}+R_{2pb}' \le I(X_2,U_{2pb};Y_2|U_{2c},U_{1c}) \quad (52c)$$
$$R_{2c}+R_{2pa}+R_{2pb}' \le I(U_{2c},X_2,U_{2pb};Y_2|U_{1c}) \quad (52d)$$
$$R_{1c}+R_{2pa}+R_{2pb}' \le I(U_{1c},X_2,U_{2pb};Y_2|U_{2c}) \quad (52e)$$
$$R_{2c}+R_{1c}+R_{2pa}+R_{2pb}' \le I(U_{2c},X_2,U_{1c},U_{2pb};Y_2) \quad (52f)$$
$$R_{1pb}+R_{1pb}' \le I(U_{1pb};Y_1|U_{2c},U_{1c}) \quad (52g)$$
$$R_{1c}+R_{1pb}+R_{1pb}' \le I(U_{1c},U_{1pb};Y_1|U_{2c}) \quad (52h)$$
$$R_{2c}+R_{1pb}+R_{1pb}' \le I(U_{2c},U_{1pb};Y_1|U_{1c}) \quad (52i)$$
$$R_{2c}+R_{1c}+R_{1pb}+R_{1pb}' \le I(U_{2c},U_{1c},U_{1pb};Y_1). \quad (52j)$$

taken over the union of all distributions of the form

$$p_{U_{1c}}p_{U_{2c}}p_{X_2|U_{2c}}p_{U_{1pb},U_{2pb}|U_{1c},U_{2c},X_2}p_{X_1|U_{2c},U_{1c},U_{1pb},U_{2pb}}.$$

Set $R_{2pb}=0$ and $R_{1c}'=I(U_{1c};X_2|U_{2c})$ in the achievable scheme of Theorem 7 and consider the factorization of the remaining RVs as in the scheme of (52), that is, according to

$$p_{U_{1c}}p_{U_{2c}}p_{X_2|U_{2c}}p_{U_{1pb},U_{2pb}|U_{1c},U_{2c},X_2}p_{X_1|U_{2c},X_2,U_{1c},U_{1pb},U_{2pb}}.$$

With this factorization of the distributions, we obtain the achievable rate region

$$R_{1c}'=I(U_{1c};X_2|U_{2c}) \quad (53a)$$
$$R_{1pb}' \ge I(U_{1pb};X_2|U_{2c},U_{1c}) \quad (53b)$$
$$R_{1pb}'+R_{2pb}' \ge I(U_{1pb};X_2,U_{2pb}|U_{2c},U_{1c}) \quad (53c)$$
$$R_{2pa}+R_{2pb}' \le I(Y_2;X_2,U_{2pb}|U_{2c},U_{1c}) + I(U_{1c};X_2|U_{2c}) \quad (53d)$$
$$R_{1c}+R_{2pa}+R_{2pb}' \le I(Y_2;U_{1c},X_2,U_{2pb}|U_{2c}) \quad (53e)$$
$$R_{2c}+R_{1c}+R_{2pa}+R_{2pb}' \le I(Y_2;U_{2pb},U_{1c},U_{2c},X_2) \quad (53f)$$
$$R_{1pb}+R_{1pb}' \le I(Y_1;U_{1pb}|U_{2c},U_{1c}) \quad (53g)$$
$$R_{1c}+R_{1pb}+R_{1pb}' \le I(Y_1;U_{1c},U_{1pb}|U_{2c}) \quad (53h)$$
$$R_{2c}+R_{1c}+R_{1pb}+R_{1pb}' \le I(Y_1;U_{2c},U_{1c},U_{1pb}). \quad (53i)$$

Note that with this particular factorization we have that $I(U_{1c};X_2|U_{2c})=0$, since X_2 is conditionally independent of U_{1c} given U_{2c}.

We now compare the region of (52) and (53) for a fixed input distribution, equation by equation:

$$(53b)=(52a)$$
$$(53c)=(52b)$$
$$(53d)=(52c)$$
$$(53e)=(52e)$$
$$(53f)=(52f)$$
$$(53g)=(52g)$$
$$(53h)=(52h)$$
$$(53i)=(52j).$$

We see that (52d) and (52i) are extra bounds that further restrict the region in [13] to be contained in the region of Theorem 7.

REFERENCES

[1] S. Rini, D. Tuninetti, and N. Devroye, "The capacity region of the Gaussian cognitive radio channels at high SNR," in *Proc. IEEE Information Theory Workshop (ITW)*, Taormina, Italy, 2009, pp. 505–509.

[2] S. Rini, D. Tuninetti, and N. Devroye, "State of the cognitive interference channel: A new unified inner bound," in *Int. Zurich Seminar on Communications (IZS)*, Mar. 2010, pp. 57–62.

[3] S. Rini, D. Tuninetti, and N. Devroye, "The capacity of the semi-deterministic cognitive interference channel and its application to constant gap results for the Gaussian channel," in *IEEE Int. Conf. Communications, (ICC)*, Kyoto, Japan, 2011.

[4] A. Goldsmith, S. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, 2009.

[5] I. Maric, R. Yates, and G. Kramer, "The strong interference channel with unidirectional cooperation," in *The Information Theory and Applications (ITA) Inaugural Workshop*, La Jolla, CA, Feb. 2006.

[6] W. Wu, S. Vishwanath, and A. Arapostathis, "Capacity of a class of cognitive radio channels: Interference channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4391–4399, Nov. 2007.

[7] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1813–1827, May 2006.

[8] A. Jovicic and P. Viswanath, "Cognitive radio: An information-theoretic perspective," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Washington, DC, Jul. 2006, pp. 2413–2417.

[9] Y. Liang, A. Somekh-Baruch, V. Poor, S. Shamai, and S. Verdú, "Cognitive interference channels with confidential messages," in *Proc. 45th Annual Allerton Conf.*, 2007.

[10] I. Maric, A. Goldsmith, G. Kramer, and S. Shamai, "On the capacity of interference channels with one cooperating transmitter," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 405–420, 2008.

[11] C. Nair and A. El Gamal, "An outer bound to the capacity region of the broadcast channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 350–355, Jan. 2007.

[12] Y. Cao and B. Chen, "Interference channel with one cognitive transmitter," in *Signals, Systems and Computers, 42nd Asilomar Conf.*, 2009, pp. 1593–1597.
[13] J. Jiang, I. Maric, A. Goldsmith, and S. Cui, "Achievable rate regions for broadcast channels with cognitive relays," in *Proc. IEEE Information Theory Workshop (ITW)*, Taormina, Italy, 2009, pp. 500–504.
[14] I. Maric, R. Yates, and G. Kramer, "Capacity of interference channels with partial transmitter cooperation," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3536–3548, Oct. 2007.
[15] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.
[16] N. Devroye, "Information Theoretic Limits of Cognition and Cooperation in Wireless Networks," Ph. D. dissertation, Harvard University, , 2007.
[17] J. Jiang and Y. Xin, "On the achievable rate regions for interference channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4707–4712, 2008.
[18] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 2, pp. 166–169, Mar. 1979.
[19] H. Sato, "An outer bound to the capacity region of broadcast channels (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 374–377, May 1978.
[20] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.
[21] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Probl. Contr. Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.
[22] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 3, pp. 306–311, May 1979.
[23] I. Maric, A. Goldsmith, G. Kramer, and S. Shamai, "On the capacity of interference channels with a cognitive transmitter," in *Information Theory and Applications (ITA) Workshop, 2007*, La Jolla, CA, Feb. 2, 2007, pp. 268–273.
[24] S. Lall, "Advanced topics in computation for control," in *Lecture Notes for Engr210b*. Stanford, CA: Stanford Univ., 2005.
[25] N. Liu, I. Marie, A. Goldsmith, and S. Shamai, "Bounds and capacity results for the cognitive Z-interference channel," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Seoul, Korea, 2009, pp. 2422–2426.
[26] S. Rini, D. Tuninetti, and N. Devroye, "Outer bounds for the interference channel with a cognitive relay," in *Proc. IEEE Information Theory Workshop (ITW)*, Dublin, Ireland, 2010.
[27] J. Jiang, Y. Xin, and H. Garg, "The capacity region of a class of deterministic interference channels with common information," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2007, vol. 3, pp. III-681–III-684.
[28] A. El Gamal and M. Costa, "The capacity region of a class of deterministic interference channels," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 343–346, Mar. 1982.
[29] S. Rini, D. Tuninetti, and N. Devroye, "Inner and outer bounds for the Gaussian cognitive interference channel and new capacity results," *IEEE Trans. Inf. Theory*, submitted for publication.
[30] F. Willems and E. Van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 313–327, 1985.

Stefano Rini received the B.A. degree in computer science from the Politecnico di Milano, Italy, in 2005, the M.A. degree in statistics from the University of Illinois at Chicago (UIC) in 2008 and the Ph.D. degree in electrical and computer engineering from UIC in 2010.

He is currently a postdoctoral fellow with the Department of Electrical Engineering at Stanford University, Stanford, CA. His research interests are in multi-user information theory with special emphasis on cognitive wireless networks.

Daniela Tuninetti received the M.S. degree in telecommunication engineering from the Politecnico di Torino, Turin, Italy, in 1998 and the Ph.D. in electrical engineering from ENST/Telecom ParisTech, Paris, France (work done at the Eurecom Institute in Sophia Antipolis, France) in 2002.

From 2002 to 2004, she was a postdoctoral research associate at the School of Communication and Computer Science, EPFL/Swiss Federal Institute of Technology, Lausanne. Since January 2005, she is an Assistant Professor with the Department of Electrical and Computer Engineering, University of Illinois at Chicago. Her research interests are in the ultimate performance limits of wireless interference networks, with special emphasis on cognition and user cooperation.

Dr. Tuninetti was the editor-in-chief of the IEEE Information Theory Society Newsletter from 2006 to 2008 and an associate editor for the IEEE COMMUNICATION LETTERS from 2006 to 2009. She currently is an editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. She regularly serves on the Technical Program Committee of IEEE workshops and conferences and she was the Communication Theory symposium co-chair of the 2010 IEEE International Conference on Communications (ICC 2010). She received the best student paper award at the European Wireless Conference in 2002 and was the recipient of an NSF CAREER award in 2007.

Natasha Devroye received the a Honors B.Eng. degree in electrical engineering from McGill University, Montreal, QC, Canada, in 2001, the M.Sc. degree from Harvard University, Cambridge, MA, in 2003, and the Ph.D. degree in engineering sciences from the School of Engineering and Applied Sciences at Harvard in 2007.

She has been an Assistant Professor in the Department of Electrical and Computer Engineering, University of Illinois at Chicago, since January 2009. From July 2007 until July 2008, she was a Lecturer at Harvard. She has worked as a research intern at Intel Corporation, Santa Clara, CA, as well as with Mitsubishi Electric Research Labs, in Cambridge. Her research focuses on multi-user information theory and applications to cognitive and software-defined radio, radar, two-way and wireless communications in general.

Dr. Devroye was a recipient of an NSF CAREER award in 2011.