

New inner and outer bounds for the discrete memoryless cognitive interference channel and some capacity results

Stefano Rini, Daniela Tuninetti, and Natasha Devroye

Department of Electrical and Computer Engineering
University of Illinois at Chicago
Email: {srini2, danielat, devroye}@uic.edu

Abstract—The cognitive interference channel is an 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 for certain classes of channels have been proved. In this paper we present 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 is known. Consequently it is not possible to evaluate such outer bounds explicitly for a given channel model. The outer bound we derive is based on an idea originally devised by Sato for the broadcast channel and does not contain auxiliary random variables, allowing it to be more easily evaluated. The inner bound we derive is the largest known to date and is explicitly shown to include all previously proposed achievable rate regions. This comparison highlights which features of the transmission scheme—which includes rate-splitting, superposition coding, a broadcast channel-like binning scheme, and Gel’fand Pinsker coding—are most effective in approaching capacity. We next present new capacity result for a class of discrete memoryless channels that we term the “better cognitive decoding regime” which includes all previously regimes in which capacity was known. Finally, we determine the capacity region of the semi-deterministic cognitive interference channel, in which the signal at the cognitive receiver is a deterministic function of the channel inputs.

I. INTRODUCTION

The rapid advancement of wireless technology in the past years has started what some commentators call the “wireless revolution” [3]. This revolution envisions a world where one can access telecommunication services on a global scale without the deployment of local infrastructure. By increasing the adaptability, communication and cooperation capabilities of wireless devices, it may be possible to realize this revolution. 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 and well interconnected 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 ultimate performance limits of a collaborating network. Given the complexity of this task in its fullest generality, researchers have focussed on simpler models with idealized assumptions.

One of the most well studied and simplest collaborative models is the genie aided cognitive interference channel. This channel is similar to the classical 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 secondary (or cognitive) pairs/users. Over the channel each transmitted message interferes with the other, creating undesired interference at the intended receiver. 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 (or non-causal) knowledge of the primary message. This assumption is referred to as genie aided cognition¹. The model was firstly posed from an information theoretic perspective in [5], 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

D. Tuninetti is with the Electrical and Computer Engineering Department of the University of Illinois at Chicago, Chicago, IL 60607, USA. Parts of this work were presented at [1], [2]. This work was partially funded by NSF under award number 0643954. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the NSF.

¹This has also been termed “unidirectional cooperation” or transmission with a “degraded message sets”.

than the classical interference channel. [5] also presents an outer bound for the Gaussian channel based on the broadcast channel; another outer bound was derived in [6], together with the first capacity result for a class of channels termed “very weak interference” in which (in Gaussian noise) treating interference at the primary receiver as noise is optimal. The same achievable rate region was simultaneously derived in [7], where the authors further characterized the maximum rate achievable by the cognitive user without degrading the rate achievable by the primary user. Another capacity result was proved in [8] for the so-called “very strong interference case”, where, without loss of optimality, both receivers can decode both messages. The capacity is also known for the case where the cognitive user decodes both messages [9] with and without confidentiality constraints.

However, the capacity region of the cognitive interference channel, both for discrete memoryless as well as Gaussian noise channels, remains unknown in general. Tools such as rate-splitting, binning, cooperation and superposition coding have been used to derive different achievable rate regions. The authors of [10] proposed an achievable 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 in [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 more general broadcast channel with cognitive relays, which contains the cognitive interference channel as special case.

A natural extension of the cognitive interference channel model is the so called “broadcast channel with cognitive relays” or “interference” channel with one cognitive relay”. In this channel model, a cognitive relay is inserted in a classical interference channel. The cognitive relay has knowledge of the two messages and thus cooperates with the two encoders in the transmission of these two messages. The model contains both the interference channel and the cognitive interference channel when removing one of the transmitters and message knowledge (for the interference channel) and thus can reveal the optimal cooperation trade off between entities in a larger network. This model was first introduced in [14], where an achievable rate region was derived. In [15] the authors introduced a larger achievable rate region and derived an outer bound for the sum capacity. In [13] a yet larger inner bound is derived by having the cognitive transmitter send a private message to both receivers as in a broadcast channel.

A. Main contributions

In this paper we establish a series of new results for the discrete memoryless cognitive interference channel. Section II introduces the basic definitions and notation. Section III summarizes the known results including general inner bounds, outer bounds and capacity in the “very weak interference” [6], [16] and “very strong interference” [17] regimes. Our contributions start in Section IV and may be summarized as follows:

- A new outer bound for the capacity region is presented in Section IV: this outer bound is looser than previously derived outer bounds but it does not include auxiliary random variables and thus it can be more easily evaluated.
- In Section V we present a new inner bound.
- We show that the newly derived region encompasses all previously presented regions in Section VI.
- We derive the capacity region of the cognitive interference channel in the “better cognitive decoding” regime in Section VII: this regime includes the “very weak interference” and the “very strong interference” regimes and is thus the largest set of channels for which capacity is known.
- Section VIII focuses on the semi-deterministic cognitive interference channel in which the output at the cognitive receiver is a deterministic function of the channel inputs. We determine capacity for this channel model by showing the achievability of the outer bound first derived in [6].
- In Section IX we consider the deterministic cognitive interference channel: in this case both channel outputs are deterministic functions of the inputs. This channel is a subcase of the semi-deterministic case for which capacity is known. For this channel model we show the achievability of the outer bound proposed in section IV, thus showing that this outer bound is tight for this class of channels.
- The paper concludes with some examples in Section X which provide insight on the role of cognition. We consider two deterministic cognitive interference channel and show the achievability of the outer bound of Section IV with transmission strategies over one channel use. The achievable scheme we propose provides interesting insights on the capacity achieving scheme in this channel model - the extra non-causal message knowledge at one of the transmitters allows a partial joint design of the codebooks and transmission strategies - and is easily appreciated in these simple deterministic models.

II. CHANNEL MODEL, NOTATION AND DEFINITIONS

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’ messages. 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 probabilities $p_{Y_1, Y_2 | X_1, X_2}(x_1, x_2)$.

Transmitter $i \in [1, 2]$ wishes to communicate a message W_i , uniformly distributed on $[1 : 2^{NR_i}]$, to receiver i in N

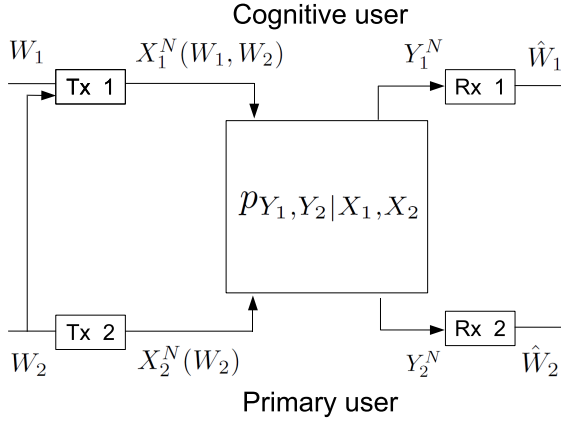


Fig. 1. The CIFIC model.

channel uses at rate R_i . 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

$$\begin{aligned} X_1^N &= X_1^N(W_1, W_2), \\ X_2^N &= X_2^N(W_2), \end{aligned}$$

and a sequence of decoding functions

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

such that

$$\lim_{N \rightarrow \infty} \max_{i \in [1, 2]} \mathbb{P} [\hat{W}_i \neq W_i] \rightarrow 0.$$

The capacity region is defined as the closure of the region of all achievable (R_1, R_2) pairs [18].

III. EXISTING RESULTS FOR THE DM-CIFIC

We now present the existing outer bounds and the capacity results available for the DM-CIFIC. The first outer bound for the DM-CIFIC was obtained in [6, Thm 3.2] by the introduction of an auxiliary Random Variable (RV).

Theorem III.1. [6, Thm 3.2]: *If (R_1, R_2) lies in the capacity region of the DM-CIFIC then*

$$R_1 \leq I(X_1; Y_1 | X_2), \quad (1a)$$

$$R_2 \leq I(X_2, U; Y_2), \quad (1b)$$

$$R_1 + R_2 \leq I(X_2, U; Y_2) + I(X_1; Y_1 | X_2, U), \quad (1c)$$

taken over the union of distributions that factor as

$$p_{U, X_1, X_2} p_{Y_1, Y_2 | X_1, X_2}.$$

The expression of the outer bound of Theorem III.1 can be simplified in two instances called “weak” and “strong interference”.

Corollary III.2. Weak interference outer bound of [6, Thm 3.4]:

When the condition

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

is satisfied, the outer bound of Theorem III.1 can be equivalently expressed as

$$R_1 \leq I(Y_1; X_1 | U, X_2), \quad (3a)$$

$$R_2 \leq I(U, X_2; Y_2), \quad (3b)$$

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

We refer to the condition in (2) as the “weak interference condition”.

Corollary III.3. Strong interference outer bound of [17, Thm 5]:

When the condition

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

is satisfied, the outer bound of Theorem III.1 can be equivalently expressed as

$$R_1 \leq I(Y_1; X_1 | X_2), \quad (5a)$$

$$R_1 + R_2 \leq I(Y_2; X_1, X_2), \quad (5b)$$

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

We refer to the condition in (4) as the “strong interference condition”.

The outer bound of Theorem III.1 may be shown to be achievable in a subset of the “weak interference” (2) and of the “strong interference” (4) conditions. We refer to these subsets as the “very strong interference” and “very weak interference” regimes.

Theorem III.4. Very weak interference capacity of [6, Thm. 3.4] and [7, Thm. 4.1]. *The outer bound of Theorem III.1 is the capacity region if*

$$\begin{aligned} I(U; Y_2 | X_2) &\leq I(U; Y_1 | X_2), \\ I(X_2; Y_2) &\leq I(X_2; Y_1), \\ &\forall p_{X_1, X_2}. \end{aligned} \quad (6)$$

We refer to the condition in (6) as “very weak interference”. In this regime capacity is achieved by having encoder 2 transmit as in a point-to-point channel and encoder 1 perform Gelfand-Pinsker binning against the interference created by transmitter 2.

In a similar spirit, capacity may be obtained in “very strong interference”.

Theorem III.5. Very strong interference capacity of [17, Thm. 5]. *The outer bound of Theorem III.1 is the capacity region if*

$$\begin{aligned} I(X_1; Y_1 | X_2) &\leq I(X_1; Y_2 | X_2), \\ I(Y_2; X_1, X_2) &\leq I(Y_1; X_1, X_2), \\ &\forall p_{X_1, X_2}. \end{aligned} \quad (7)$$

We refer to the condition in (7) as “very strong interference”. In this regime, capacity is achieved by having both receivers decode both messages.

The outer bound presented in Theorem III.1 cannot be evaluated in general since it includes an auxiliary RV whose cardinality has not yet been bounded. In the following we propose a new outer bound, looser in general than the outer bound of Theorem III.1 but without auxiliary RVs.

IV. A NEW OUTER BOUND

Theorem IV.1. *If (R_1, R_2) lies in the capacity region of the DM-CIFC then*

$$R_1 \leq I(Y_1; X_1|X_2), \quad (8a)$$

$$R_2 \leq I(X_1, X_2; Y_2), \quad (8b)$$

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

taken over the union of all distributions p_{X_1, X_2} and $p_{Y_1, Y_2'|X_1, X_2}$, where Y_2' has the same marginal distribution as Y_2 , i.e., $p_{Y_2'|X_1, X_2} = p_{Y_2|X_1, X_2}$.

Proof: The proof of this theorem 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 marginal distributions $P_{Y_1|X_1, X_2}$ and $P_{Y_2|X_1, X_2}$ because the receivers do not cooperate. ■

Remark IV.2. The outer bound of Theorem IV.1 contains the outer bound of Theorem III.1. Indeed, for a fixed distribution p_{X_1, X_2} , the bounds on R_1 are the same ((1a) = (8a)) and the bound on R_2 , $I(X_2, U; Y_2) \leq I(X_1, X_2; Y_2)$ (or (1b) \leq (8b)) since

$$\begin{aligned} I(Y_2; X_2, U) &\stackrel{(a)}{\leq} I(Y_2; X_2, U) + I(Y_2; X_1|U, X_2) \\ &= I(Y_2; X_1, X_2, U) \\ &= I(Y_2; X_1, X_2), \end{aligned}$$

where the last equality follows from the Markov chain $U - X_1, X_2 - Y_1, Y_2$.

Consider Y_2' such that $p_{Y_2'|U, X_1, X_2} = p_{Y_2|U, X_1, X_2}$, which also implies $p_{Y_2'|U, X_2} = p_{Y_2|U, X_2}$. then:

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

where (a) follows from the non-negativity of mutual information, (b) follows from the Markov chain $U - X_1, X_2 - Y_1, Y_2$ and (c) from the fact that conditioning reduces entropy. Now the RV U does not appear in the outer bound expression (8c) and thus we can consider simply the RVs with $p_{\tilde{Y}_2|X_1, X_2} = p_{Y_2|X_1, X_2}$ which corresponds to the definition of Y_2' in Theorem IV.1. Equality of the outer bounds in Theorems III.1 and

IV.1 is verified when conditions (a) and (c) hold with equality, that is when

$$\begin{aligned} I(Y_2; X_1|U, X_2) &= 0, \\ I(Y_1; X_1|\tilde{Y}_2, U, X_2) &= I(Y_1; X_1|\tilde{Y}_2, X_2), \quad \forall p_U, \end{aligned}$$

for a given \tilde{Y}_2 . The first condition implies the Markov Chain (MC)

$$Y_2 - U, X_2 - X_1,$$

and the second condition the MC

$$Y_1, X_1 - \tilde{Y}_2 X_2 - U.$$

We currently cannot relate these conditions to any specific class of DM-CIFC.

Remark IV.3. The outer bound of Theorem IV.1 reduces to the “strong interference” outer bound in (5), in fact

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

implies

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

Now let $Y_2' = Y_2$ to obtain that $I(Y_1; X_1|Y_2, X_2) = 0$ yielding (8c) = (8b) so that the two outer bounds coincide.

V. A NEW INNER BOUND

As the DM-CIFC encompasses classical interference, multiple-access 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.** We employ a rate-splitting technique similar to that in Han and Kobayashi’s scheme of [19] for the interference-channel, also employed in the DM-CIFC regions of [10], [5], [20]. While rate-splitting may be useful in general, is not necessary in the “very weak” [6] and “very strong interference” [8] regimes of (6) and (7).

- **Superposition-coding.** Useful in multiple-access and broadcast channels [18], in the DM-CIFC the superposition of private messages on top of common ones, as in [10], [20], is known to be capacity achieving in “very strong interference” [8].

- **Binning.** Gel’fand-Pinsker coding [21], often simply referred to as binning, allows a transmitter to “cancel” (portions of) the interference known to be experienced at a receiver. Binning is also used by Marton in deriving the largest known achievable rate region [22] for the discrete memoryless broadcast channel.

We now present a new achievable rate region for the DM-CIFC which generalizes all the known achievable rate regions presented in [10], [6], [20], [23], [12] and [13]. In the next Section, VI 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 V.1. The \mathcal{R}_{RTD} region. A rate pair (R_1, R_2) such that

$$\begin{aligned} R_1 &= R_{1c} + R_{1pb}, \\ R_2 &= R_{2c} + R_{2pa} + R_{2pb}. \end{aligned} \quad (9)$$

is achievable for the DM-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 (10) for some input distribution

$$\begin{aligned} p_{Y_1, Y_2, X_1, X_2, U_{1c}, U_{2c}, U_{2pa}, U_{1pb}, U_{2pb}} \\ = p_{U_{1c}, U_{2c}, U_{2pa}, U_{1pb}, U_{2pb}, X_1, X_2} p_{Y_1, Y_2 | X_1, X_2}. \end{aligned}$$

Remark V.2. Moreover:

- (23c) can be dropped when $R_{2c} = R_{2pa} = R_{2pb} = R'_{2pb} = 0$;
- (10e) can be dropped when $R_{2pa} = R_{2pb} = R'_{2pb} = 0$;
- (10g) can be dropped when $R_{2pb} = R'_{2pb} = 0$;
- (10i) can be dropped when $R_{1c} = R'_{1c} = R_{1pb} = R'_{1pb} = 0$,

since they correspond to the event that a common message from the non-intended user is incorrectly decoded. This event is not an error event if no other intended message is incorrectly decoded.

Proof: The meaning of the RVs in Theorem V.1 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 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 V.1 can be found in Fig. 2. The formal description of the proposed encoding scheme is as follows:

A. 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:

$$\begin{aligned} W_1 &= (W_{1c}, W_{1pb}), \\ W_2 &= (W_{2c}, W_{2pb}, W_{2pa}), \end{aligned}$$

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

B. Codebook generation

Consider a distribution $p_{U_{1c}, U_{2c}, X_2, U_{1pb}, U_{1pb}, X_1, X_2}$. 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 $U_{2pb}^N(w_{2c}, w_{2pa}, w_{1c}, b_0, w_{2pb}, b_2)$, $w_{2pb} \in [1 : 2^{NR_{2pb}}]$ 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 $w_{2c} \in [1 : 2^{NR_{2c}}]$, $w_{2pa} \in [1 : 2^{NR_{2pa}}]$, $w_{1c} \in [1 : 2^{NR_{1c}}]$, $b_0 \in [1 : 2^{NR'_{1c}}]$, $w_{1pb} \in [1 : 2^{NR_{1pb}}]$, $b_1 \in [1 : 2^{NR'_{1pb}}]$, $w_{2pb} \in [1 : 2^{NR_{2pb}}]$, $b_2 \in [1 : 2^{NR'_{2pb}}]$, let the channel input $X_1^N(w_{2pa}, w_{2c}, w_{1c}, b_0, w_{1pb}, b_1, w_{2pb}, b_2)$ be any length- N sequence from the typical set $T_\epsilon^N(p_{X_1, U_{2c}, U_{1c}, X_2, U_{2pb}, U_{1pb}} | U_{2c}^N(w_{2c}), X_2^N(w_{2c}, w_{2pa}), U_{1c}^N(w_{2c}, w_{1c}, b_0), U_{2pb}^N(w_{2c}, w_{2pa}, w_{1c}, b_0, w_{2pb}, b_2), U_{1pb}^N(w_{2c}, w_{1c}, b_0, w_{1pb}, b_1))$.

C. Encoding

Given the message $w_2 = (w_{2c}, w_{2pb}, w_{2pa})$, encoder 2 sends the codeword $X_2^N(w_{2c}, w_{2pa})$.

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 triplet (b_0, b_1, b_2) such that:

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

$$\begin{aligned}
R'_{1c} &= I(U_{1c}; X_2|U_{2c}) & (10a) \\
R'_{1c} + R'_{1pb} &\geq I(U_{1pb}; X_2|U_{1c}, U_{2c}) + I(U_{1c}; X_2|U_{2c}) & (10b) \\
R'_{1c} + R'_{1pb} + R'_{2pb} &\geq I(U_{1pb}; X_2, U_{2pb}|U_{1c}, U_{2c}) + I(U_{1c}; X_2|U_{2c}) & (10c) \\
R_{2c} + R_{2pa} + (R_{1c} + R'_{1c}) + (R_{2pb} + R'_{2pb}) &\leq I(Y_2; U_{2pb}, U_{1c}, X_2, U_{2c}) + I(U_{1c}; X_2|U_{2c}) & (10d) \\
R_{2pa} + (R_{1c} + R'_{1c}) + (R_{2pb} + R'_{2pb}) &\leq I(Y_2; U_{2pb}, U_{1c}, X_2|U_{2c}) + I(U_{1c}; X_2|U_{2c}) & (10e) \\
R_{2pa} + (R_{2pb} + R'_{2pb}) &\leq I(Y_2; U_{2pb}, X_2|U_{1c}, U_{2c}) + I(U_{1c}; X_2|U_{2c}) & (10f) \\
(R_{1c} + R'_{1c}) + (R_{2pb} + R'_{2pb}) &\leq I(Y_2; U_{2pb}, U_{1c}|X_2, U_{2c}) + I(U_{1c}; X_2|U_{2c}) & (10g) \\
(R_{2pb} + R'_{2pb}) &\leq I(Y_2; U_{2pb}|U_{1c}, X_2, U_{2c}) & (10h) \\
R_{2c} + (R_{1c} + R'_{1c}) + (R_{1pb} + R'_{1pb}) &\leq I(Y_1; U_{1pb}, U_{1c}, U_{2c}), & (10i) \\
(R_{1c} + R'_{1c}) + (R_{1pb} + R'_{1pb}) &\leq I(Y_1; U_{1pb}, U_{1c}|U_{2c}), & (10j) \\
(R_{1pb} + R'_{1pb}) &\leq I(Y_1; U_{1pb}|U_{1c}, U_{2c}), & (10k)
\end{aligned}$$

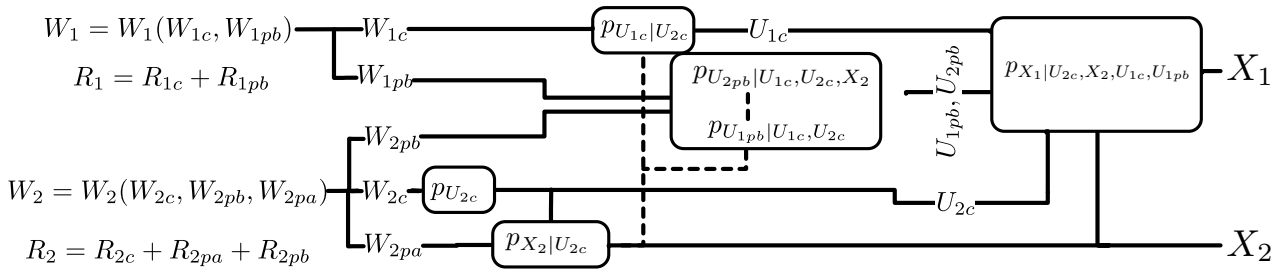


Fig. 2. The achievability encoding scheme of Thm. V.1. The ordering from left to right and the distributions demonstrate the codebook generation process. The dotted lines indicate binning. We see rate splits are used at both users, private messages W_{1pb} , W_{2pa} , W_{2pb} are superimposed on common messages W_{1c} , W_{2c} and U_{1c} , is binned against X_2 conditioned on U_{2c} , while U_{1pb} and U_{2pb} are binned against each and X_2 other in a Marton-like fashion (conditioned on other subsets of RVs).

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

D. Decoding

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

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

The probability of error at decoder 2 goes to zero if conditions (10e)-(10h) hold.

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

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

The probability of error at decoder 1 goes to zero if conditions (10i)-(10k) hold.

The detailed error analysis is found in Appendix B. ■

E. Two step binning

It is also possible to perform binning in a sequential manner. First, U_{1c} is binned against X_2 , and then U_{1pb} and

U_{2pb} are binned against each other conditioned on (U_{2c}, U_{1c}) and (U_{2c}, X_2, U_{1c}) respectively. With respect to the encoding operation of the previous section, this affects Section V-C 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 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 no such b_0 exists, it sets $b_0 = 1$. If more than one such b_0 exists, it picks one uniformly at random. For the selected b_0 , encoder 1 looks for (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 no such (b_1, b_2) exists, it sets $(b_1, b_2) = (1, 1)$. If more than one such (b_1, b_2) exists, it picks one uniformly at random from the found ones.

For the selected (b_0, b_1, b_2) , encoder 1 sends $X_1^N(w_{2pa}, w_{2c}, w_{1c}, b_0, w_{1pb}, b_1, w_{2pb}, b_2)$.

The next lemma states the condition under which this two step encoding procedure is successful with high probability.

Lemma V.3. *The two-step binning encoding procedure of Section V-E is successful if*

$$R'_{1c} \geq I(U_{1c}; X_2 | U_{2c}), \quad (11a)$$

$$R'_{1pb} \geq I(U_{1pb}; X_2 | U_{2c}, U_{1c}), \quad (11b)$$

$$R'_{1pb} + R'_{2pb} \geq I(U_{1pb}; X_2, U_{2pb} | U_{2c}, U_{1c}). \quad (11c)$$

The proof of the lemma is found in Appendix (F).

Remark V.4. Since the binning rate (10a) of Theorem V.1 can be taken with equality without loss of generality, the two step binning has the same performance as joint binning. In fact, by setting (11a) to hold with equality, we obtain the equality between the binning rate expression of the joint binning and the two step binning.

VI. COMPARISON WITH EXISTING ACHIEVABLE RATE REGIONS

We now show that the region of Theorem V.1 contains all other known achievable rate regions for the DM-CIFC. Showing inclusion of the rate regions [24, Thm.2], [23, Thm. 1] and [13, Thm. 4.1] is sufficient to demonstrate the largest known DM-CIFC region, since the region of [24, Thm.2] (first presented in [12]) is shown (in [24]) to contain those of [10, Thm. 1] and [20].

A. Devroye et al.'s region [23, Thm. 1]

In Appendix G we show that the region of [23, Thm. 1] \mathcal{R}_{DMT} , is contained in our new region \mathcal{R}_{RTD} along the lines:

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

B. Cao and Chen's region [24, Thm. 2]

The region in [24, Thm. 2] uses a similar encoding structure as that of \mathcal{R}_{RTD} with two exceptions:

1) The binning is done sequentially rather than jointly as in \mathcal{R}_{RTD} leading to binning constraints (43)–(45) in [24, Thm. 2] as opposed to (10a)–(10c) in Thm.V.1. 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) While the cognitive messages are rate-split in identical fashions, the primary message is split into 2 parts in [24, Thm. 2] ($R_1 = R_{11} + R_{10}$, note the reversal of indices) while we explicitly split the primary message into three parts $R_2 = R_{2c} + R_{2pa} + R_{2pb}$. In Appendix H we show that the region of [24, Thm.2], denoted as $\mathcal{R}_{CC} \subseteq \mathcal{R}_{RTD}$ in two steps:

- We first show that we may WLOG set $U_{11} = \emptyset$ in [24, Thm.2], creating a new region R'_{CC} .
- We next make a correspondence between our RVs and those of [24, Thm.2] and obtain identical regions.

C. Jiang et al.'s region [13, Thm. 4.1]

The scheme originally designed for the more general broadcast channel with cognitive relays (or interference-channel with a cognitive relay) may be tailored/reduced to derive a region for the cognitive interference channel. This scheme also incorporates a broadcasting strategy. However, the common messages are created independently instead of having the common message from transmitter 1 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}_{RTD} after equating random variables. The proof of the containment of the achievable region of [13, Thm. 4.1] in \mathcal{R}_{RTD} is found in Appendix I. We note that the region of [25], used to prove capacity in the cognitive Z-IFC when the interference-free component is noiseless, is a special case of the region in [24] and is thus also contained in this region. Also the outer bound of Liu et al. uses a technique tailored to the special Z channel and is not clear how it can be compared to the existing outer bounds.

VII. NEW CAPACITY RESULTS FOR THE DM-CIFC

We now look at the expression of the outer bound [6, Thm. 3.1] to gain insight into potentially capacity achieving achievable schemes. In particular we look at the expression of the corner points of the outer bound region for a fixed p_{U, X_1, X_2} and try to interpret the RVs as private and common messages to be decoded at the transmitter side. We then consider an achievable scheme inspired by these observations and show that schemes achieve capacity for a particular class of channels. This class of channels contains the “very strong” and the “very weak interference” regimes and thus corresponds to the largest class of channels for which capacity is currently known.

The outer bound region of [6, Thm. 3.1] has at most two corner points where both R_1 and R_2 are non zero:

$$\begin{aligned} (R_1^{out(a)}, R_2^{out(a)}) \\ = (I(Y_1; X_1 | U, X_2), I(Y_2; U, X_2)) \end{aligned} \quad (12)$$

and

$$\begin{aligned} (R_1^{out(b)}, R_2^{out(b)}) \\ = (I(Y_1; X_1 | U, X_2) + I(Y_2; U, X_2) - \Delta, \Delta) \end{aligned} \quad (13)$$

for

$$\Delta = [I(Y_2; U, X_2) - I(Y_1; U | X_2)]^+,$$

since

$$\begin{aligned} R_2^{out(a)} &= \min\{I(Y_2; U, X_2), I(Y_2; U, X_2) + I(Y_1; X_1 | U, X_2)\} \\ &= I(Y_2; U, X_2), \\ R_1^{out(a)} &= \min\{I(Y_1; X_1 | U, X_2), I(Y_1; X_1 | X_2)\} \\ &= I(Y_1; X_1 | U, X_2), \end{aligned}$$

and

$$\begin{aligned} R_2^{out(b)} &= \min\{I(Y_2; U, X_2), \\ &I(Y_2; U, X_2) + I(Y_1; X_1 | U, X_2) - I(Y_1; X_1 | X_2)\} \\ &= [I(Y_2; U, X_2) + \\ &\quad \min\{0, I(Y_1; X_1 | U, X_2) - I(Y_1; X_1 | X_2)\}]^+ \\ &= [I(Y_2; U, X_2) - I(Y_1; U | X_2)]^+ \triangleq \Delta, \end{aligned}$$

with

$$\begin{aligned} R_1^{out (b)} &\leq \min\{I(Y_1; X_1|X_2), \\ &\quad I(Y_2; U, X_2) + I(Y_1; X_1|U, X_2)\} \\ &= I(Y_1; X_1|U, X_2) + I(Y_2; U, X_2) \\ &\quad - \max\{I(Y_2; U, X_2) - I(Y_1; U|X_2), 0\} \\ &= I(Y_1; X_1|U, X_2) + I(Y_2; U, X_2) - \Delta. \end{aligned}$$

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 $(R_1^{(a)}, R_2^{(a)})$ we can interpret (U, X_2) as a common message from transmitter 2 to receiver 2 that is also decoded at receiver 1. X_1 is superposed to (U, X_2) since the decoding of X_1 follows the one of (U, X_2) at decoder 2.

The corner point $(R_1^{out (b)}, R_2^{out (b)})$ has two possible expressions:

1) If $I(Y_1; U|X_2) \leq I(Y_2; U, X_2)$ we have that

$$(R_1^{out (b)'}, R_2^{out (b)'}) = \quad (14)$$

$$(I(Y_1; X_1, U|X_2), I(Y_2; U, X_2) - I(Y_1; U|X_2)), \quad (15)$$

which suggests that X_2 is again the common primary message and the cognitive message is divided into a public and private part, U and X_1 respectively.

2) If $I(Y_1; U|X_2) > I(Y_2; U, X_2)$ we have that

$$(R_1^{out (b)'}, R_2^{out (b)'}) = (I(Y_2; U, X_2) + I(Y_1; X_1, U|X_2), 0). \quad (16)$$

In this case the outer bound has only one corner point where both rates are non zero. Note that we can always achieve the point

$$(R_1^{in (b)'}, R_2^{in (b)'}) = (I(Y_1; X_1, U|X_2), 0)$$

by having transmitter 2 send a known signal. In this case we have $R_2^{out (b)'}$ = $R_2^{in (b)'}$ and $R_1^{out (b)'}$ \leq $R_1^{in (b)'}$ since

$$\begin{aligned} I(Y_1; X_1, U|X_2) &\geq I(Y_2; U, X_2) + I(Y_1; X_1, U|X_2) \\ I(Y_1; U|X_2) &\geq I(Y_2; U, X_2). \end{aligned}$$

So in this case showing the achievability of the point in equation (13) is sufficient to show capacity.

Guided by these observations, we consider a scheme that has only the components U_{2c}, U_{1c} and U_{1pb} . That is, the primary message w_2 is common and the cognitive message w_1 is split into a private and a public message. With this scheme we are able to extend the capacity results in the ‘‘very weak interference’’ of Theorem III.4 and the ‘‘very strong interference’’ of Theorem III.5. This scheme coincides with the scheme of [26] which achieves capacity if the cognitive receiver is required to decode both messages (with and without the secrecy constraint).

Theorem VII.1. Capacity in the ‘‘better cognitive decoding’’ regime.

When the following condition holds

$$I(Y_1; X_2, U) \geq I(Y_2; X_2, U) \quad \forall p_{X_1, X_2, U}, \quad (17)$$

the capacity region of the DM-CIFC is given by Theorem III.1.

Proof: Consider the achievable rate region of Theorem V.1 when setting

$$\begin{aligned} X_1 &= U_{1pb} \\ X_2 &= U_{2c} = U_{2pb} \end{aligned}$$

so that

$$\begin{aligned} R_2 &= R_{2c} \\ R_{2pa} &= R_{2pb} = 0 \\ R'_{1c} &= R'_{1pb} = R'_{2pb} = 0. \end{aligned}$$

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 region of this sub-scheme is:

$$R_2 + R_{1c} \leq I(Y_2; U_{1c}, X_2), \quad (18a)$$

$$R_2 + R_{1c} + R_{1pb} \leq I(Y_1; U_{1c}, X_2), \quad (18b)$$

$$R_{1c} + R_{1pb} \leq I(Y_1; U_{1c}, X_1|X_2), \quad (18c)$$

$$R_{1pb} \leq I(Y_1; X_1|X_2, U_{1c}). \quad (18d)$$

By applying Fourier-Motzkin elimination [27] we obtain the achievable rate region

$$R_1 \leq I(Y_1; U_{1c}, X_1|X_2), \quad (19a)$$

$$R_2 \leq I(Y_2; U_{1c}, X_2), \quad (19b)$$

$$R_1 + R_2 \leq I(Y_2; U_{1c}, X_2) + I(Y_1; X_1|X_2, U_{1c}), \quad (19c)$$

$$R_1 + R_2 \leq I(Y_1; X_2, U_{1c}, X_1). \quad (19d)$$

By letting $U_{1c} = U$ we see that (1a) matches (19a), (1b) matches (19b), (1c) matches (19c), and (19d) is redundant when

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

or equivalently when

$$I(Y_1; U, X_2) \geq I(Y_2; U, X_2). \quad (20)$$

We term the condition in equation (20) ‘‘better cognitive decoding’’ since decoder 1 has a higher mutual information between its received channel output and the RVs U and X_2 than the primary receiver. ■

Remark VII.2. The ‘‘better cognitive decoding’’ in (20) is looser than both the ‘‘very weak interference’’ condition of (6) and the ‘‘very strong interference’’ condition of (7). In fact summing the two equations of condition (6) we have

$$\begin{aligned} I(U; Y_1|X_2) + I(X_2; Y_1) &\geq I(U; Y_2|X_2) + I(X_2; Y_2) \\ &\quad \Downarrow \\ I(Y_1; U, X_2) &\geq I(Y_2; U, X_2), \end{aligned}$$

which corresponds to condition (20). Similarly by summing the two equation of condition (7) we obtain

$$\begin{aligned} I(Y_1; X_1, X_2) + I(X_1; Y_2|X_2) &\geq I(Y_2; X_1, X_2) \\ + I(X_1; Y_1|X_2) &\iff \\ I(Y_1; X_1, X_2) - I(X_1; Y_1|X_2) &\geq I(Y_2; X_1, X_2) \\ - I(X_1; Y_2|X_2) &\iff \\ I(Y_1; X_1, X_2, U) - I(X_1; Y_1|X_2) &\geq I(Y_2; X_1, X_2, U) \\ - I(X_1; Y_2|X_2) &\iff \\ I(Y_1; X_2, U) &\geq I(Y_2; X_2, U) \end{aligned}$$

which again corresponds to condition (20).

Since both (6) and (7) imply the (20), we conclude that (20) is more general than the previous two.

The scheme that achieves capacity in “very weak interference” is obtained by setting $U_{1c} = X_2$ so that all the cognitive message is private and the primary message is common. The scheme that achieves capacity in “very strong interference” is obtained by setting $U_{1c} = X_1$ so that both transmitters send only public messages. The scheme that we use to show the achievability in the “strong cognitive decoding” regime mixes these two schemes by splitting the cognitive message into public and private messages. This relaxes the “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” achievability condition since it allows the cognitive encoder to decode part of the cognitive message and remove its unwanted effects. For this reason, the resulting achievability conditions are looser than both cases.

VIII. CAPACITY FOR THE SEMI-DETERMINISTIC CIFC

Consider the specific class of DM-CIFC for which the signal received at the cognitive receiver is a deterministic function of the channel inputs, that is

$$Y_1 = f_1(X_1, X_2), \quad (21)$$

for some function f_1 . This class of channels is termed semi-deterministic CIFC and it was first introduced in [24]. In [24] the capacity region is derived for the case $I(Y_1; X_2) \geq I(Y_2; X_2)$; we extend this result by determining the capacity region in general (no extra conditions). Note that the authors of [24] consider the case where f_1 is invertible; we do not require this condition.

Theorem VIII.1. *The capacity region of the semi-deterministic cognitive interference channel in (21) consists of all $(R_1, R_2) \in \mathbb{R}_+^2$ such that*

$$R_1 \leq H(Y_1|X_2), \quad (22a)$$

$$R_2 \leq I(Y_2; U, X_2), \quad (22b)$$

$$R_1 + R_2 \leq I(Y_2; U, X_2) + H(Y_1|U, X_2), \quad (22c)$$

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

Proof:

Converse: We consider the outer bound of Theorem III.1 with the additional deterministic assumption in (21) i.e. $H(Y_1|X_1, X_2) = 0$.

Achievability: The region of Theorem V.1 computed for $U_{2c} = U_{1c} = \emptyset$, $U_1 = U_{1pb}$ and $U_2 = U_{2pb}$, and $R_{2c} = R_{1c} = R_{1pb} = 0$ becomes:

$$R'_1 \geq I(U_1; X_2), \quad (23a)$$

$$R'_1 + R'_2 \geq I(U_1; U_2, X_2), \quad (23b)$$

$$R_2 + R'_2 \leq I(Y_2; U_2, X_2), \quad (23c)$$

$$R_1 + R'_1 \leq I(Y_1; U_1), \quad (23d)$$

for any $P_{U_1 U_2 X_1 X_2}$. After Fourier-Motzkin elimination, the region in (23) may be rewritten as

$$R_1 \leq I(Y_1; U_1) - I(U_1; X_2), \quad (24a)$$

$$\begin{aligned} R_2 &\leq I(Y_2; U_2, X_2) \\ &= I(Y_2; X_1, X_2) - I(Y_2; X_1|U_2, X_2), \end{aligned} \quad (24b)$$

$$\begin{aligned} R_1 + R_2 &\leq I(Y_2; U_2, X_2) + I(Y_1; U_1) - I(U_1; U_2, X_2) \\ &= (24a) + (24b) - I(U_1; U_2|X_2). \end{aligned} \quad (24c)$$

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 region in (24) reduces to the outer bound. ■

Remark VIII.2. The achievable scheme in (23) cannot be obtained as a special case of any previously known achievable scheme but [13]. The RV U_{2pb} , which broadcasts the private primary message from transmitter 1, appears in [24] as well. In this case it is possible to reobtain the scheme in (23) with a specific choice of the RVs. Here the same message is embedded in U_{2pb} and the private primary message, this performs strictly worse than using only U_{2pb} .

Remark VIII.3. The achievability scheme for the semi-deterministic C-IFC has been used (see [2], [28]) to provide a unified scheme which achieves a constant additive gap for the Gaussian C-IFC in the whole parameter space. This supports the notion that results for (semi)-deterministic channel models may carry over to noisy networks.

IX. CAPACITY FOR THE DETERMINISTIC CIFC

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

$$\begin{aligned} Y_1 &= f_1(X_1, X_2), \\ Y_2 &= f_2(X_1, X_2), \end{aligned} \quad (25a)$$

for some function f_1 and f_2 . This class of channels is a subclass of the semi-deterministic CIFC of Section VIII, and we already have obtained the capacity region for this case. However, we re-derive the capacity region in a new fashion for this channel, which illustrates two new ideas:

We show the achievability of the outer bound of Theorem IV.1 when letting $Y'_2 = Y_2$, instead of the outer bound of Theorem III.1.

Theorem IX.1. *The capacity region of the deterministic cognitive interference channel consists of all $(R_1, R_2) \in \mathbb{R}_+^2$ such that*

$$R_1 \leq H(Y_1|X_2), \quad (26a)$$

$$R_2 \leq H(Y_2), \quad (26b)$$

$$R_1 + R_2 \leq H(Y_2) + H(Y_1|Y_2, X_2), \quad (26c)$$

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

Proof: Outer bound: The outer bound is obtained from Theorem IV.1 using the deterministic conditions in (25a).

Achievability: Consider the scheme in (24) and let $U_{1pb} = Y_1$, $U_{2pb} = Y_2$ to achieve the region in (26). ■

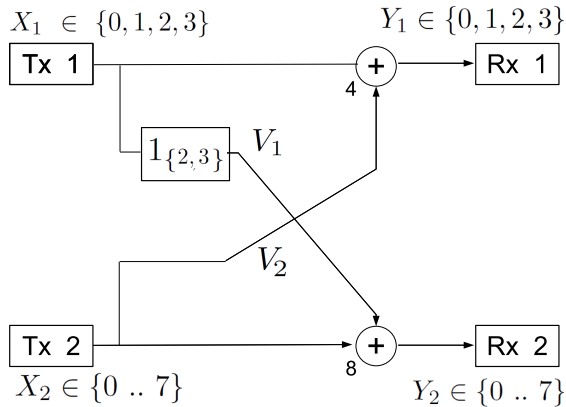


Fig. 3. The “asymmetric clipper” of Section X-A.

X. EXAMPLES

The scheme that achieves capacity in 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}) , but interestingly, with R_{2pb} equal to zero. 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 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 III.5 which, in the deterministic case, reduces to

$$\begin{aligned} H(Y_1|X_2) &\leq H(Y_2|X_2) \\ H(Y_2) &\leq H(Y_1) \quad \forall p_{X_1, X_2}. \end{aligned} \quad (27)$$

Unfortunately, checking for the “very weak interference condition” of Theorem III.4 is not possible as no cardinality bounds on U are available.

A. Example I: the “Asymmetric Clipper”

Consider the channel in Fig. 3. The input and output alphabets are $\mathcal{X}_1 = \mathcal{Y}_1 = [0, 1, 2, 3]$ and $\mathcal{X}_2 = \mathcal{Y}_2 = [0, 1, 2, 3, 4, 5, 6, 7]$ and the input/output relationships are

$$Y_1 = X_1 \oplus_4 X_2, \quad (28)$$

$$Y_2 = 1_{\{2,3\}}(X_1) \oplus_8 X_2, \quad (29)$$

where $1_{\{A\}}(x) = 1$ if $x \in A$ and zero otherwise, and \oplus_N denotes the addition operation over the ring $[1 : N]$. Also let $\mathcal{U}(\mathcal{S})$ be the uniform distribution over the set \mathcal{S} . First we show that the channel in (29) does not fall in the “very strong interference” class. Consider the input distribution:

$$\begin{aligned} X_2 &\sim \mathcal{U}(1) \implies P[X_1 = 0] = 1, \\ X_2 &\sim \mathcal{U}(\mathcal{X}_2). \end{aligned}$$

For this input distribution, we have $Y_1 \sim \mathcal{U}(\mathcal{Y}_1)$ and $Y_2 \sim \mathcal{U}(\mathcal{Y}_2)$, so that

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

which does not satisfy the “very strong interference” condition of (27).

For this channel we have:

$$H(Y_1|X_2) \leq H(Y_1) \leq \log(|\mathcal{Y}_1|) = 2,$$

$$H(Y_2) \leq \log(|\mathcal{Y}_2|) = 3,$$

$$H(Y_1|X_2, Y_2) \leq H(X_1|1_{\{2,3\}}(X_1)) \leq 1,$$

where the last bound follows from the multiplicity of the solutions of an addition in a Galois field. This shows that the outer bound in Theorem IX.1 is included in

$$R_1 \leq 2, \quad (30a)$$

$$R_2 \leq 3, \quad (30b)$$

$$R_1 + R_2 \leq 4. \quad (30c)$$

We now show that the region in (30) indeed corresponds to Theorem IX.1 when considering the union over all input distributions. The corner point $(R_1, R_2) = (1, 3)$ in (30) is obtained in Theorem IX.1 with the input distribution:

$$\begin{aligned} X_1 &\sim \mathcal{U}([0, 1]), \\ X_2 &\sim \mathcal{U}(\mathcal{X}_2). \end{aligned}$$

The corner point $(R_1, R_2) = (2, 2)$ in (30) is obtained in Theorem IX.1 by considering the input distribution:

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

Time sharing shows that the region of (30) and the region of Theorem IX.1 indeed coincide.

We next show the achievability of the corner point $(R_1, R_2) = (1, 3)$. Consider the following strategy:

- transmitter 2 sends symbols from $\mathcal{X}_2 = [0 : 7]$ with uniform probability,
- transmitter 1 transmits $[x_1 - x_2]_2$ (where the inverse of the difference operation is taken over the ring \mathcal{G}_2);
- receiver 1 decodes $\hat{w}_1 = \lfloor \frac{y_2}{2} \rfloor$;
- receiver 2 decodes $\hat{w}_2 = y_2$.

It can be verified by inspection of Table I that the rate pair $(R_1, R_2) = (1, 3)$ is indeed achievable.

TABLE I
ACHIEVABILITY TABLE FOR THE RATE POINT $(R_1, R_2) = (1, 3)$ IN EXAMPLE I IN SECTION X-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 DECODING MESSAGE (\hat{w}_1, \hat{w}_2) .

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

Now we show the achievability of the corner point $(R_1, R_2) = (2, 2)$. Consider the following strategy:

- transmitter 2 sends symbols from $x_2 \in [0, 2, 4, 6]$ with uniform probability;
- transmitter 1 transmits $[x_1 - x_2]_4$ (where the inverse of the difference operation is taken over the ring \mathcal{G}_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 that the rate pair $(R_1, R_2) = (2, 2)$ is indeed achievable.

TABLE II

ACHIEVABILITY TABLE FOR THE RATE POINT $(R_1, R_2) = (2, 2)$ IN EXAMPLE I IN SECTION X-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 DECODING MESSAGE (\hat{w}_1, \hat{w}_2) .

	w_1	w_2	x_1	x_2	y_1	y_2	\hat{w}_1	\hat{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	0	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 this example we see how the two senders jointly design the codebook to achieve the outer bound and in particular how the cognitive transmitter 1 adapts its strategy to the transmissions from the primary pair so as to avoid interfering with it.

In achieving the point $(R_1, R_2) = (1, 3)$, transmitter 2's strategy is that of a point to point channel. Transmitter 1 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 message is sending $w_2 = 0$ (line 0 and 8 in Table I) transmitter 1 can send either 1 or 2 without creating interference at receiver 2. On the other hand, these two values produce a different output at receiver 1, allowing the transmission of 1 bit.

In achieving the point $(R_1, R_2) = (2, 2)$, the primary receiver picks its codewords so as to tolerate 1 unit of interference. Transmitter 1 again chooses its input codewords in order to create at most 1 unit of interference at the primary decoder. By adapting its transmission to the primary symbol, 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 carefully picking the codewords at the primary transmitter. For example, when the primary receiver is sending $w_2 = 0$ (lines 0, 4, 8 and 12), transmitter 1 can send $x_1 \in [0, 1, 2, 3]$ and create at most 1 bit of interference at receiver 2. Each of these four values

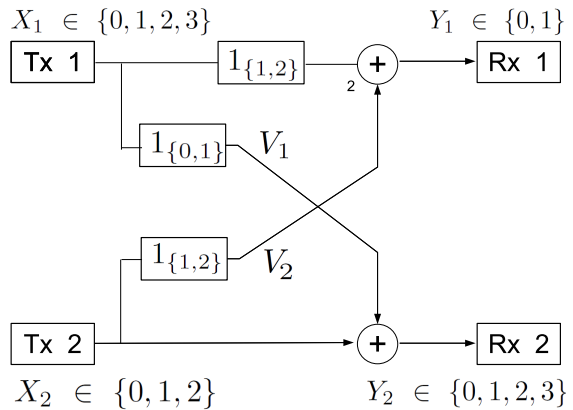


Fig. 4. “Symmetric Clipper” of Section X-B

TABLE III
THE INPUT DISTRIBUTION THAT ACHIEVES THE OUTER BOUND OF THEOREM IX.1 FOR THE CHANNEL IN EXAMPLE II

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	

produces a different output at receiver 1], thus allowing the transmission of 2 bits.

B. Example II: the “Symmetric Clipper”

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

$$\begin{aligned} Y_1 &= 1_{\{1,2\}}(X_1) \oplus 1_{\{1,2\}}(X_2), \\ Y_2 &= 1_{\{0,1\}}(X_1) \oplus X_2. \end{aligned}$$

Consider the input distribution: Consider the input distribution:

$$\begin{aligned} P[X_1 = 3] &= 1, \\ X_2 &\sim \mathcal{U}([1, 2]), \end{aligned}$$

in this case $H(Y_1) = 0$ and $H(Y_2) = 1$. This shows that there exists at least one input distribution for which $H(Y_2) > H(Y_1)$ and thus this channel is not in the “very strong interference” regime. The outer bound of Theorem IX.1 is achieved here by a single input distribution p_{X_1, X_2} : consider the distribution in Table III. This distribution produces $H(Y_1) = 1 = \log_2(|\mathcal{Y}_1|)$ and $H(Y_2) = 2 = \log_2(|\mathcal{Y}_2|)$ and clearly no larger outer bound can exist given the output cardinality. We therefore conclude that the region of Theorem IX.1 can be rewritten as:

$$\begin{aligned} R_1 &\leq 1, \\ R_2 &\leq 2. \end{aligned} \quad (31)$$

This region can be shown to be achievable using the transmission scheme described in Table IV. The decoding is simply $\hat{w}_i = Y_i$, $i \in [1, 2]$. This transmission scheme

TABLE IV
ACHIEVABILITY TABLE FOR THE RATE POINT $(R_1, R_2) = (1, 2)$ IN
EXAMPLE II: 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) .

	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

achieves the proposed outer bound, thus showing capacity. The transmission scheme can be described as follows:

- encoder 2 transmits $[x_2 - 1]^+$;
- encoder 1 transmits the value X_1 that simultaneously makes $Y_1 = w_1$ and $Y_2 = w_2$. For each w_1 and w_2 such a value always exists because X_2 takes on only three possible values;
- receivers 1 and 2 decode $\hat{w}_1 = Y_1$ and $\hat{w}_2 = Y_2$.

This example is particularly interesting since both decoders obtain the transmitted symbol without suffering any interference from the other user. 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. The transmitters optimally design their codebooks so to make the effect X_1 on both outputs the desired one.

For example consider the transmission of $w_2 = 2$ or 3 (lines 2, 3, 6 and 7). 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 cancelation is possible due to the channel's deterministic nature and the extra message knowledge at the cognitive transmitter.

XI. CONCLUSION

In this paper we focused on the discrete memoryless cognitive interference channel and derived new inner and outer bounds, derived the capacity region for a class of "better cognitive decoding" channels, and obtained the capacity region for the semi-deterministic cognitive interference channel. We proposed a new outer bound using an idea originally devised for the broadcast channel in [29]. This outer bound does not involve auxiliary RVs and is thus more easily computable. We also proposed a new inner bound that generalizes all other known achievable rate regions. We determined capacity for a class of channels that we term the "better cognitive decoding" regime. The conditions defining this regime are looser than the

"very weak interference condition" of [6] and the "very strong interference condition" of [17] and is the largest region where capacity is known. We also determined the capacity region for the class of semi-deterministic cognitive interference channels. Furthermore, for channels where both outputs are deterministic functions of the inputs, we showed the achievability of our new outer bound. The scheme that achieves capacity in the deterministic cognitive interference channel uses Gelf'and-Pinsker binning against the interference created at the primary receiver. This binning is performed by the cognitive encoder for the primary decoder. This feature of the transmission scheme was never known before to be capacity achieving. Extensions of the results presented here to Gaussian channels will be presented in [28].

REFERENCES

- [1] S. Rini, D. Tuninetti, and N. Devroye, "State of the cognitive interference channel: a new unified inner bound," in *International Zurich Seminar on Communications*, March 2010, p. 57.
- [2] —, "the capacity of the semi-deterministic cognitive interference channel and its application to constant gap results for the gaussian channel," *ICC Kyoto, Japan*, 2011, submitted.
- [3] M. Best, "The wireless revolution and universal access," *Trends in Telecommunications Reform*, pp. 1–24, 2003.
- [4] A. Goldsmith, S. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. of the IEEE*, 2009.
- [5] 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.
- [6] W. Wu, S. Vishwanath, and A. Arapostathis, "Capacity of a class of cognitive radio channels: Interference channels with degraded message sets," *Information Theory, IEEE Transactions on*, vol. 53, no. 11, pp. 4391–4399, Nov. 2007.
- [7] A. Jovicic and P. Viswanath, "Cognitive radio: An information-theoretic perspective," *Proc. IEEE ISIT Washington, U.S.A.*, pp. 2413–2417, July 2006.
- [8] R. D. Y. I. Maric and G. Kramer, "The strong interference channel with unidirectional cooperation," *The Information Theory and Applications (ITA) Inaugural Workshop*, Feb 2006, UCSD La Jolla, CA.
- [9] Y. Liang, A. Somekh-Baruch, H. V. Poor, S. Shamai, and S. Verdú, "Cognitive interference channels with confidential messages," *Proc. of the 45th Annual Allerton Conference*, vol. abs/0710.2018, 2007.
- [10] I. Maric, A. Goldsmith, G. Kramer, and S. Shamai, "On the capacity of interference channels with one cooperating transmitter," *Information Theory, IEEE Transactions on*, vol. 19, no. 4, pp. 405–420, 2008.
- [11] C. Nair and A. E. Gamal, "An outer bound to the capacity region of the broadcast channel," *Information Theory, IEEE Transactions on*, vol. 53, no. 1, pp. 350–355, Jan. 2007.
- [12] Y. Cao and B. Chen, "Interference channel with one cognitive transmitter," in *Asilomar Conference on Signals, Systems, and Computers*, Nov. 2008.
- [13] J. Jiang, I. Maric, A. Goldsmith, and S. Cui, "Achievable Rate Regions for Broadcast Channels With Cognitive Relays," *Proc. IEEE ITW, Taormina, Italy*, 2009.
- [14] O. Sahin and E. Erkip, "On achievable rates for interference relay channel with interference cancelation," pp. 805–809, 2008.
- [15] S. Sridharan, S. Vishwanath, S. Jafar, and S. Shamai, "On the capacity of cognitive relay assisted Gaussian interference channel," in *Proc. IEEE ISIT, Toronto, Canada*, 2008, pp. 549–553.
- [16] A. Jovicic and P. Viswanath, "Cognitive radio: An information-theoretic perspective," vol. 55, no. 9, pp. 3945–3958, Sep. 2009.
- [17] I. Maric, R. Yates, and G. Kramer, "Capacity of interference channels with partial transmitter cooperation," *Information Theory, IEEE Transactions on*, vol. 53, no. 10, pp. 3536–3548, Oct. 2007.
- [18] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley-Interscience, 1991.
- [19] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *Information Theory, IEEE Transactions on*, vol. 27, no. 1, pp. 49–60, Jan 1981.

- [20] J. Jiang and Y. Xin, "On the achievable rate regions for interference channels with degraded message sets," *Information Theory, IEEE Transactions on*, vol. 54, no. 10, pp. 4707–4712, Oct. 2008.
- [21] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Problems of control and information theory*, 1980.
- [22] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *Information Theory, IEEE Transactions on*, vol. 25, no. 3, pp. 306–311, May 1979.
- [23] N. Devroye, "Information theoretic limits of cognition and cooperation in wireless networks," Ph.D. dissertation, Harvard University, 2007.
- [24] Y. Cao and B. Chen, "Interference channels with one cognitive transmitter," in *Proc. Asilomar Conference on Signal, Systems and Computers*, 2009.
- [25] N. Liu, I. Marie, A. Goldsmith, and S. Shamai, "Bounds and capacity results for the cognitive Z-interference channel," in *ISIT 2009, Seoul, Korea*. IEEE, 2009, pp. 2422–2426.
- [26] J. Jiang, Y. Xin, and H. Garg, "The capacity region of a class of deterministic interference channels with common information," *IEEE ICASSP 2007*, vol. 3, pp. III–681–III–684, April 2007.
- [27] S. Lall, "Advanced topics in computation for control," *Lecture Notes for Engr210b at Stanford University, Stanford, CA*.
- [28] S. Rini, D. Tuninetti, and N. Devroye, "Inner and outer bounds for the gaussian cognitive interference channel and new capacity results," *Information Theory, IEEE Transactions on*, 2010, to be submitted.
- [29] H. Sato, "An outer bound to the capacity region of broadcast channels (Corresp.)," *Information Theory, IEEE Transactions on*, vol. 24, no. 3, pp. 374–377, 1978.
- [30] F. Willems and E. Van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *Information Theory, IEEE Transactions on*, vol. 31, no. 3, pp. 313–327, 1985.

APPENDIX

A. Proof of outer bound of Theorem IV.1

By Fano's inequality we have that $H(W_i|Y_i^N) \leq N\epsilon_N$, for some ϵ_N such that $\epsilon_N \rightarrow 0$ as $N \rightarrow 0$ for $i \in [1, 2]$. The rate of user 1 can be bounded as

$$\begin{aligned}
N(R_1 - \epsilon_N) &\leq I(W_1; Y_1^N) \\
&\stackrel{(a)}{\leq} I(W_1; Y_1^N | W_2) \\
&\stackrel{(b)}{=} I(W_1, X_1^N(W_1, W_2); Y_1^N | W_2, X_2^N(W_2)) \\
&\stackrel{(c)}{=} H(Y_1^N | W_2, X_2^N) - H(Y_1^N | W_2, W_1, X_1^N, X_2^N) \\
&\stackrel{(d)}{\leq} H(Y_1^N | X_2^N) - H(Y_1^N | W_2, W_1, X_1^N, X_2^N) \\
&\stackrel{(e)}{=} H(Y_1^N | X_2^N) - H(Y_1^N | X_1^N, X_2^N) \\
&\stackrel{(f)}{=} \sum_{i=1}^N H(Y_{1i} | X_2^N, (Y_1)_{1}^{i-1}) - H(Y_{1i} | X_2^N, X_2^N, (Y_1)_{1}^{i-1}) \\
&\stackrel{(g)}{\leq} \sum_{i=1}^N H(Y_{1i} | X_{2i}) - H(Y_{1i} | X_{1i}, X_{2i}) \\
&\stackrel{(h)}{=} NI(Y_{1T}; X_{1T} | X_{2T}, T) \\
&\stackrel{(i)}{=} N(H(Y_{1T} | X_{2T}, T) - H(Y_{1T} | X_{1T}, X_{2T}, T)) \\
&\stackrel{(j)}{=} N(H(Y_{1T} | X_{2T}, T) - H(Y_{1T} | X_{1T}, X_{2T})) \\
&\stackrel{(k)}{\leq} N(H(Y_{1T} | X_{2T}) - H(Y_{1T} | X_{1T}, X_{2T})) \\
&\stackrel{(l)}{=} NI(Y_{1T}; X_{1T} | X_{2T}), \tag{32a}
\end{aligned}$$

where T is the time sharing RV, informally distributed over the set $[1 : N]$. The (in)equalities follow from: (a) independence of messages W_1, W_2 , (b) the channel model where X_1^N depends on W_1 and W_2 , while X_2^N depends only on W_2 ,

(c) by definition, (d) as conditioning reduces entropy, (e) as $(W_1, W_2) \rightarrow (X_1^N, X_2^N) \rightarrow (Y_1^N)$ forms a Markov chain, (f) the chain rule, (g) memorylessness and conditioning reduces entropy, (h) and (i) by definition and introduction of the time-sharing RV T , (j) as the channel output depends only on channel inputs, (k) conditioning reduces entropy, (l) definition.

The rate of user 2 can be bounded as

$$\begin{aligned}
N(R_2 - \epsilon_N) &\leq I(Y_2^N; W_2) \\
&\stackrel{(a)}{\leq} I(Y_2^N; W_2, W_1) \\
&\stackrel{(b)}{=} H(Y_2^N) - H(Y_2^N | W_1, W_2, X_2^N(W_2), X_1^N(W_1, W_2)) \\
&\stackrel{(c)}{=} H(Y_2^N) - H(Y_2^N | X_2^N, X_1^N) \\
&\stackrel{(d)}{=} \sum_{i=1}^N H(Y_{2i} | (Y_2)_{1}^{i-1}) - H(Y_{2i} | X_1^N, X_2^N, (Y_2)_{1}^{i-1}) \\
&\stackrel{(e)}{\leq} \sum_{i=1}^N H(Y_{2i}) - H(Y_{2i} | X_{1i}, X_{2i}) \\
&\stackrel{(f)}{\leq} NI(Y_{2T}; X_{1T}, X_{2T} | T) \\
&\stackrel{(g)}{=} N(H(Y_{2T} | T) - H(Y_{2T} | X_{1T}, X_{2T}, T)) \\
&\stackrel{(h)}{\leq} N(H(Y_{2T}) - H(Y_{2T} | X_{1T}, X_{2T})) \\
&\stackrel{(i)}{=} NI(Y_{2T}; X_{1T}, X_{2T}). \tag{32b}
\end{aligned}$$

Here the (in)equalities follow from (a) non-negativity of mutual information, (b) definition and channel model, (c) as $(W_1, W_2) \rightarrow (X_1^N, X_2^N) \rightarrow (Y_1^N)$ forms a Markov chain, (d) the chain rule, (e) conditioning reduces entropy, (f) introduction of the time-sharing RV T uniformly distributed over the set $[1 : N]$, (g) definition, (h) conditioning reduces entropy and channel outputs depend only on channel inputs, (i) definition.

Next let Y_2' be any 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

$$\begin{aligned}
N(R_1 + R_2 - 2N\epsilon_N) &\leq I(W_1; Y_1) + I(W_2; Y_2) \\
&\stackrel{(a)}{\leq} I(W_1; Y_1^N | W_2) + I(W_2; Y_2^N) \\
&\stackrel{(b)}{\leq} I(W_1; Y_1^N, Y_2'^N | W_2) + I(W_2; Y_2^N) \\
&\stackrel{(c)}{=} I(W_2; Y_2^N) + I(W_1; Y_2'^N | W_2) + I(W_1; Y_1^N | Y_2'^N, W_2) \\
&\stackrel{(d)}{=} H(Y_2^N) + \left(-H(Y_2^N | W_2) + H(Y_2'^N | W_2) \right) \\
&\quad - H(Y_2'^N | W_1, W_2) + H(Y_1^N | Y_2'^N, W_2) \\
&\quad - H(Y_1^N | Y_2'^N, W_1, W_2) \\
&\stackrel{(e)}{=} H(Y_2^N) + H(Y_1^N | W_2, X_2^N, Y_2'^N) \\
&\quad - H(Y_2'^N | W_1, W_2, X_1^N, X_2^N) \\
&\quad - H(Y_1^N | Y_2'^N, W_1, W_2, X_1^N, X_2^N) \\
&\stackrel{(f)}{=} H(Y_2^N) + H(Y_1^N | W_2, X_2^N, Y_2'^N) \\
&\quad - H(Y_2^N | X_1^N, X_2^N) - H(Y_1^N | Y_2'^N, X_1^N, X_2^N) \\
&\stackrel{(g)}{\leq} H(Y_2^N) + H(Y_1^N | X_2^N, Y_2'^N) \\
&\quad - H(Y_2^N | X_1^N, X_2^N) - H(Y_1^N | Y_2'^N, X_1^N, X_2^N) \\
&\stackrel{(h)}{=} I(Y_2^N; X_1^N, X_2^N) \\
&\quad + \sum_{i=1}^N H(Y_{1i} | X_2^N, Y_2'^N, (Y_1)_1^{i-1}) \\
&\quad - \sum_{i=1}^N H(Y_{1i} | X_1^N, X_2^N, Y_2'^N, (Y_1)_1^{i-1}) \\
&\stackrel{(i)}{\leq} I(Y_2^N; X_1^N, X_2^N) \\
&\quad + \sum_{i=1}^N H(Y_{1i} | X_{2i}, Y_{2i}') - H(Y_{1i} | X_{1i}, X_{2i}, Y_{2i}') \\
&\stackrel{(j)}{=} N(I(Y_{2T}; X_{1T}, X_{2T}) + H(Y_{1T} | X_{2T}, Y_{2T}', T) \\
&\quad - H(Y_{1T} | X_{1T}, X_{2T}, Y_{2T}')) \\
&\stackrel{(k)}{\leq} N(I(Y_{2T}; X_{1T}, X_{2T}) + I(Y_{1T}; X_{1T} | X_{2T}, Y_{2T}')). \quad (32c)
\end{aligned}$$

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 $(W_1, W_2) \rightarrow (X_1^N, X_2^N) \rightarrow (Y_1^N)$ forms a Markov chain, (g) conditioning reduces entropy, (h) chain rule, (i) conditioning reduces entropy and memorylessness, (j) and (k) by definition and definition of the time-sharing RV T uniformly distributed over the set $[1 : N]$.

B. Error analysis of the achievable region \mathcal{R}_{RTD} of Theorem V.1

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 tuple (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 $(\hat{w}_{1c}, \hat{w}_{2c}, \hat{w}_{1pb}, \hat{b}_0, \hat{b}_1)$ be the estimate at the decoder 1.

The probability of error at decoder u , $u \in [1, 2]$, is bounded by

$$P[\text{error } u] \leq P[\text{error } u | \text{encoding successful}] + P[\text{encoding NOT successful}].$$

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 $(\hat{w}_{1c}, \hat{w}_{1pb}) \neq (1, 1)$. A decoding error is committed at decoder 2 when $(\hat{w}_{2c}, \hat{w}_{2pa}, \hat{w}_{2pb}) \neq (1, 1, 1)$.

C. Encoding Error

Since the codebooks are generated iid according to

$$p^{(\text{codebook})} = p_{U_{2c}} p_{X_2 | U_{2c}} p_{U_{1c} | U_{2c}} p_{U_{2pb} | U_{2c}, U_{1c}, X_2} p_{U_{1pb} | U_{2c}, U_{1c}} \quad (33)$$

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

$$p^{(\text{encoding})} = p_{U_{2c}} p_{X_2 | U_{2c}} p_{U_{1c} | U_{2c}, X_2} p_{U_{2pb} | U_{2c}, U_{1c}, X_2}, p_{U_{1pb} | U_{2c}, U_{1c}, X_2, U_{2pb}}, \quad (34)$$

We expect the probability of encoding error to depend on

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

The probability that the encoding fails can be bounded as:

$$\begin{aligned}
P[\text{encoding NOT successful}] &= P \left[\bigcap_{b_0=1}^{2^{NR'_{1c}}} \bigcap_{b_1=1}^{2^{NR'_{1pb}}} \bigcap_{b_2=1}^{2^{NR'_{2pb}}} \right. \\
&\quad \left. \left(U_{2c}^N(1), X_2^N(1, 1), U_{1c}^N(1, 1, b_0), U_{1pb}^N(1, 1, b_0, 1, b_1), \right. \right. \\
&\quad \left. \left. U_{2pb}^N(1, 1, b_0, 1, b_2) \right) \notin T_\epsilon^N(p_{U_{2c}, X_2, U_{1c}, U_{1pb}, U_{2pb}}) \right] \\
&= P[K = 0] \leq \frac{\text{Var}[K]}{E^2[K]}
\end{aligned}$$

where

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

and

$$K_{b_0, b_1, b_2} = 1 \left\{ \left(U_{2c}^N(1), X_2^N(1, 1), U_{1c}^N(1, 1, b_0), U_{1pb}^N(1, 1, b_0, 1, b_1), \right. \right. \\ \left. \left. U_{2pb}^N(1, 1, b_0, 1, b_2) \right) \in T_\epsilon^N(p_{U_{2c}, X_2, U_{1c}, U_{1pb}, U_{2pb}}) \right\}$$

where $1_{\{x \in A\}} = 1$ if $x \in A$ and zero otherwise.

The mean value of K (neglecting all terms that depend on ϵ and that eventually go to zero) is:

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

with

$$\begin{aligned}
2^{-NA} &= P[K_{b_0, b_1, b_2} = 1] = E[K_{b_0, b_1, b_2}] \\
&= P\left[\left(U_{2c}^N(1), X_2^N(1, 1), U_{1c}^N(1, 1, b_0), U_{1pb}^N(1, 1, b_0, 1, b_1), \right. \right. \\
&\quad \left. \left. U_{2pb}^N(1, 1, b_0, 1, b_2)\right) \in T_\epsilon^N(p_{U_{2c}, X_2, U_{1c}, U_{1pb}, U_{2pb}})\right] \\
&= \sum_{(u_{1c}^N, u_{1pb}^N, u_{2pb}^N) \in T_\epsilon^N(p_{U_{2c}, X_2, U_{1c}, U_{1pb}, U_{2pb}} | u_{2c}^N, x_2^N)} \\
&\quad \mathcal{P}_{U_{1c}|U_{2c}} \mathcal{P}_{U_{2pb}|U_{2c}, U_{1c}, X_2} \mathcal{P}_{U_{1pb}|U_{2c}, U_{1c}} \\
&\geq 2^{-N[I(U_{1c}; X_2|U_{2c}) + I(U_{1pb}; X_2, U_{2pb}|U_{1c}, U_{2c})]}.
\end{aligned}$$

The variance of K (neglecting all terms that depend on ϵ and that eventually go to zero) is:

$$\begin{aligned}
\text{Var}[K] &= \sum_{b_0=1}^{2^{NR'_{1c}}} \sum_{b_1=1}^{2^{NR'_{1pb}}} \sum_{b_2=1}^{2^{NR'_{2pb}}} \sum_{b'_0=1}^{2^{NR'_{1c}}} \\
&\quad \sum_{b'_1=1}^{2^{NR'_{1pb}}} \sum_{b'_2=1}^{2^{NR'_{2pb}}} (P[K_{b_0, b_1, b_2} = 1, K_{b'_0, b'_1, b'_2} = 1] \\
&\quad - P[K_{b_0, b_1, b_2} = 1]P[K_{b'_0, b'_1, b'_2} = 1]) \\
&= \sum_{b'_0=b_0, (b_1, b_2, b'_1, b'_2)} (P[K_{b_0, b_1, b_2} = 1, K_{b_0, b'_1, b'_2} = 1] \\
&\quad - P[K_{b_0, b_1, b_2} = 1]P[K_{b_0, b'_1, b'_2} = 1]) \\
&\leq \sum_{b_0, (b_1, b_2, b'_1, b'_2)} P[K_{b_0, b_1, b_2} = 1, K_{b_0, b'_1, b'_2} = 1]
\end{aligned}$$

because when $b_0 \neq b'_0$, that is, $U_{1c}^N(\cdot, b_0)$ and $U_{1c}^N(\cdot, b'_0)$ are independent, 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:

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

and

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

and

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

and

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

Hence, we can bound $P[K = 0]$ as:

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

and $P[K = 0] \rightarrow 0$ if

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

that is, if

$$\begin{aligned}
R'_{1c} + R'_{1pb} + R'_{2pb} &> I(U_{1c}; X_2|U_{2c}) + I(U_{1pb}; X_2, U_{2pb}|U_{1c}, U_{2c}) \\
&= I(U_{1c}, U_{1pb}; X_2|U_{2c}) + I(U_{1pb}; U_{2pb}|U_{1c}, U_{2c}, X_2) \\
R'_{1c} + R'_{1pb} &> I(U_{1c}; X_2|U_{2c}) + I(U_{1pb}; X_2|U_{1c}, U_{2c}) \\
&= I(U_{1c}, U_{1pb}; X_2|U_{2c}) \\
R'_{1c} + R'_{2pb} &> I(U_{1c}; X_2|U_{2c}), \\
R'_{1c} &> I(U_{1c}; X_2|U_{2c})
\end{aligned}$$

as in (10a)-(10c), because the second to last equation is redundant.

D. Decoding Errors at decoder 2

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

$$p_{2|\star} \triangleq \mathcal{P}_{U_{2c}} \mathcal{P}_{X_2|U_{2c}} \mathcal{P}_{U_{1c}|U_{2c}} \mathcal{P}_{U_{2pb}|U_{2c}, U_{1c}, X_2} \mathcal{P}_{Y_2|\star}, \quad (35)$$

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	p_{Y_2}
$E_{2,2a}$	1	X	X	...	$p_{Y_2 U_{2c}}$
$E_{2,2b}$	1	1	X	...	$p_{Y_2 U_{2c}, U_{1c}}$
$E_{2,3a}$	1	X	1	X	$p_{Y_2 U_{2c}, X_2}$
$E_{2,3b}$	1	1	1	X	$p_{Y_2 U_{2c}, U_{1c}, X_2}$

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

$$p_2 \triangleq p_{U_{2c}} p_{X_2|U_{2c}} p_{U_{1c}|U_{2c}, X_2} p_{U_{2pb}|U_{2c}, U_{1c}, X_2} p_{Y_2|U_{2c}, U_{1c}, X_2, U_{2pb}} \quad (36)$$

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

$$\begin{aligned} I_{2|\star} &= E \left[\log \frac{p_2}{p_{2|\star}} \right] \\ &= E \left[\log \frac{p_{U_{1c}|U_{2c}, X_2} p_{Y_2|U_{2c}, U_{1c}, X_2, U_{2pb}}}{p_{U_{1c}|U_{2c}} p_{Y_2|\star}} \right] \\ &= I(U_{1c}; X_2|U_{2c}) + I(Y_2; U_{2c}, U_{1c}, X_2, U_{2pb}|\star). \quad (37) \end{aligned}$$

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

$$P[\text{error 2}|\text{encoding successful}] \leq \sum_{i \in \{1, 2a, 2b, 3a, 3b\}} P[E_{2,i}],$$

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 “...” indicates that “it does not matter whether the corresponding message is correct or not; in this case the most restrictive case is when the message is actually wrong.” The last column of Table V specifies the $p_{Y_2|\star}$ to be used in (35).

We have that $P[\text{error 2}|\text{encoding successful}] \rightarrow 0$ when $N \rightarrow \infty$ 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:

$$\begin{aligned} P[E_{2,1}] &= P \left[\bigcup_{\substack{\tilde{w}_{2c} \neq 1, \tilde{w}_{2pa}, \tilde{w}_{1c}, \tilde{w}_{2pb}, b_0, b_2 \\ (Y_2^N, U_{2c}^N(\tilde{w}_{2c}), U_{1c}^N(\tilde{w}_{1c}, \tilde{w}_{2c}, b_0), X_2^N(\tilde{w}_{2c}, \tilde{w}_{2pa}), \\ U_{2pb}^N(\tilde{w}_{2c}, \tilde{w}_{2pa}, \tilde{w}_{1c}, b_0, \tilde{w}_{2pb}, b_2)) \in T_\epsilon^N(p_{Y_2, U_{2c}, U_{1c}, X_2, U_{2pb}})}} \right] \\ &\leq 2^{N(R_{2c} + R_{2pa} + R_{1c} + R'_{1c} + R_{2pb} + R'_{2pb})} \\ &\quad \sum_{(y_2^N, u_{2c}^N, u_{1c}^N, x_2^N, u_{2pb}^N) \in T_\epsilon^N(p_{Y_2, U_{2c}, U_{1c}, X_2, U_{2pb}})} p_{2|\star} \Big|_{\star=0} \\ &\leq 2^{N(R_{2c} + R_{2pa} + R_{1c} + R'_{1c} + R_{2pb} + R'_{2pb} - I_{2|\star} \Big|_{\star=0})} \end{aligned}$$

for $p_{2|\star}$ given in (36) and $I_{2|\star}$ given in (37). Hence $P[E_{2,1}] \rightarrow 0$ as $N \rightarrow \infty$ if (23c) 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 and also of 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

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

for $p_{2|\star}$ given in (36) and $I_{2|\star}$ given in (37). Hence $P[E_{2,2a}] \rightarrow 0$ as $N \rightarrow \infty$ if (10e) 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

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

for $p_{2|\star}$ given in (36) and $I_{2|\star}$ given in (37). Hence $P[E_{2,2b}] \rightarrow 0$ as $N \rightarrow \infty$ if (10f) 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 \widehat{w}_{1c} . However we need to consider whether the pair $(\widehat{w}_{1c}, \widehat{b}_0)$ 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 $\widehat{w}_{1c} \neq 1$ or $\widehat{b}_0 \neq \bar{b}_0$. 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 $\widehat{w}_{1c} \neq 1$ and $\widehat{b}_0 \neq \bar{b}_0$. Thus we have

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

for $p_{2|\star}$ given in (36) and $I_{2|\star}$ given in (37). Hence $P[E_{2,3a}] \rightarrow 0$ as $N \rightarrow \infty$ if (10g) is satisfied.

- Case $E_{2,3b}$: both $\widehat{w}_{1c} = 1$ and $\widehat{b}_0 = \bar{b}_0$. 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 (36), but we need to replace $p_{U_{1c}|U_{2c}}$ in (36) with $p_{U_{1c}|U_{2c}, X_2}$. Thus we have

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

Hence $P[E_{2,3b}] \rightarrow 0$ as $N \rightarrow \infty$ if (10h) is satisfied.

E. Decoding Errors at Decoder 1

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

$$p_{1|\star} \triangleq p_{U_{2c}} p_{U_{1c}|U_{2c}} p_{U_{1pb}|U_{2c}, U_{1c}} p_{Y_1|\star}, \quad (38)$$

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

$$p_1 \triangleq p_{U_{2c}} p_{U_{1c}|U_{2c}} p_{U_{1pb}|U_{2c}, U_{1c}} p_{Y_1|U_{2c}, U_{1c}, U_{1pb}}. \quad (39)$$

TABLE VI
ERROR EVENTS AT DECODER 1.

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

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

$$\begin{aligned} I_{1|\star} &= E \left[\log \frac{p_1}{p_{1|\star}} \right] \\ &= E \left[\log \frac{p_{Y_1|U_{2c}, U_{1c}, U_{1pb}}}{p_{Y_1|\star}} \right] \\ &= I(Y_1; U_{2c}, U_{1c}, U_{1pb}|\star). \end{aligned} \quad (40)$$

The probability of error at decoder 1 is bounded as:

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

where $P[E_{1,i}]$ is the error event defined in Table VI. The meaning of the symbols in Table VI is as for Table V. We have that $P[\text{error} | \text{encoding successful}] \rightarrow 0$ when $N \rightarrow \infty$ if:

- When the event $E_{1,1}$ occurs we have $\widehat{w}_{2c} \neq 1$. 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:

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

for $p_{1|\star}$ given in (36) and $I_{1|\star}$ given in (40). Hence $P[E_{1,1}] \rightarrow 0$ as $N \rightarrow \infty$ if (10i) is satisfied.

- When the event $E_{1,2}$ occurs, either $\widehat{w}_{1c} \neq 1$, $\widehat{b}_0 \neq \bar{b}_0$ 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 and U_{1pb}^N . It is easy to see that the most stringent error event is when both $\widehat{w}_{1c} \neq 1$ and $\widehat{b}_0 \neq \bar{b}_0$. Thus we have

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

for $p_{1|\star}$ given in (39) and $I_{1|\star}$ given in (40). Hence $P[E_{1,2}] \rightarrow 0$ as $N \rightarrow \infty$ if (10j) is satisfied.

- When the event $E_{1,3}$ occurs, either $\widehat{w}_{1pb} \neq 1$, $\widehat{b}_1 \neq \bar{b}_1$ or both. In this case, conditioned on the (correct) decoded sequence U_{2c}^N and 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

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

for $p_{1|\star}$ given in (39) and $I_{1|\star}$ given in (40). Hence $P[E_{1,3}] \rightarrow 0$ as $N \rightarrow \infty$ if (10k) is satisfied.

F. Proof of Lemma V.3

An encoding error is committed 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 first event and $E_{e,12}$ of the latter, than:

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

where

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

Using standard typicality arguments we have

$$\begin{aligned} P[(U_{2c}^N(1), X_2^N(1, 1), U_{1c}^N(1, 1, b_0)) \notin T_\epsilon^N(p_{U_{2c}, X_2, U_{1c}})] \\ = \sum_{u_{1c} \in T_\epsilon^N(p_{U_{2c}, X_2, U_{1c}}|U_{2c}, X_2)} \\ \geq (1 - \epsilon) 2^{N(I(U_{1c}; X_2|U_{2c}) + \delta)}. \end{aligned}$$

Now we can write

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

so that $P[E_{e,0}] \rightarrow 1$ when $N \rightarrow 0$ if (11a) is satisfied.

Now the error event $E_{e,12}$ can be divided in three distinct 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 equations, 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 rate bounds that guarantee that the probability of error of each of these events goes to zero.

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

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

where

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

As for $E_{e,0}$, this implies that $P[E_{e,21 a}] \rightarrow 1$ when $N \rightarrow 0$ if (11b) is satisfied.

For $E_{e,21 b}$, we have that the probability of this event goes to one for large N given that (U_{2c}, X_2, U_{1c}) 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:

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

where

$$\begin{aligned} P[(U_{2c}^N(1), X_2^N(1, 1), U_{1c}^N(1, 1, b_0), U_{1pb}^N(1, 1, b_0, 1, b_1), \\ U_{2pb}^N(1, 1, b_0, 1, b_2)) \notin T_\epsilon^N(p_{U_{2c}, X_2, U_{1c}, U_{1pb}, U_{2pb}})] \leq 2^{I(\cdot) + \delta}. \end{aligned}$$

This implies that $P[E_{e,21 c}] \rightarrow 1$ when $N \rightarrow 0$ if (11c) is satisfied.

G. Containment of [23, Thm. 1] in \mathcal{R}_{RTD} of Section VI-A

We refer to the region in [23, Thm. 1] as \mathcal{R}_{DMT} for brevity. We show this inclusion of \mathcal{R}_{DMT} in \mathcal{R}_{RTD} with the following steps:

- We enlarge the region \mathcal{R}_{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}_{DMT}^{out} since it enlarges the original achievable region.
- We make a correspondence between the RVs and corresponding rates of \mathcal{R}_{DMT}^{out} and \mathcal{R}_{RTD} .
- We choose a particular subset of \mathcal{R}_{RTD} , \mathcal{R}_{RTD}^{in} , for which we can more easily show $\mathcal{R}_{DMT} \subseteq \mathcal{R}_{DMT}^{out} \subseteq \mathcal{R}_{RTD}^{in} \subseteq \mathcal{R}_{RTD}$, since \mathcal{R}_{DMT}^{out} and \mathcal{R}_{RTD}^{in} have identical input distribution decompositions and similar rate bound equations.

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

We first enlarge the rate region of [23, Thm. 1], \mathcal{R}_{DMT} by removing a number of constraints (specifically, we remove equations (2.6, 2.8, 2.10, 2.13, 2.14, 2.16 2.17) of [23, Thm. 1]). Also, following the line of thoughts in [30, Appendix 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}_{DMT}^{out} as in (41). taken over the union of distributions

$$p_W p_{V_{11}} p_{V_{12}} p_{X_1|V_{11}, V_{12}} p_{V_{21}|V_{11}, V_{12}} p_{V_{22}|V_{11}, V_{12}} p_{X_2|V_{11}, V_{12}, V_{21}, V_{22}}. \quad (42)$$

Using the factorization of the auxiliary RVs in [23, Thm. 1], we may insert X_1 next to V_{11} in equation (41f).

$$R'_{21} = I(V_{21}; X_1, V_{11}, V_{12}|W) \quad (41a)$$

$$R'_{22} = I(V_{22}; X_1, V_{11}, V_{12}|W) \quad (41b)$$

$$R_{11} \leq I(Y_1, V_{12}, V_{21}; V_{11}|W) \quad (41c)$$

$$R_{21} + R'_{21} \leq I(Y_1, X_1, V_{11}, V_{12}; V_{21}|W) \quad (41d)$$

$$R_{11} + R_{21} + R'_{21} \leq I(Y_1, V_{12}; V_{11}, V_{21}|W) + I(V_{11}; V_{21}|W) \quad (41e)$$

$$R_{11} + R_{21} + R'_{21} + R_{12} \leq I(Y_1; X_1, V_{11}, V_{12}, V_{21}|W) + I(X_1, V_{11}, V_{12}; V_{21}|W) \quad (41f)$$

$$R_{22} + R'_{22} \leq I(Y_2, V_{12}, V_{21}; V_{22}|W) \quad (41g)$$

$$R_{22} + R'_{22} + R_{21} + R'_{21} \leq I(Y_2, V_{12}; V_{22}, V_{21}|W) + I(V_{22}; V_{21}|W) \quad (41h)$$

$$R_{22} + R'_{22} + R_{21} + R'_{21} + R_{12} \leq I(Y_2; V_{22}, V_{21}, V_{12}|W) + I(V_{22}, V_{21}; V_{12}|W) \quad (41i)$$

For equation (41c):

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

For equation (41e) 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 (43) union over all distributions that factor as in (42).

Enlarge the class of input distribution and eliminate V_{11} and W

Now increase the set of possible input distributions of equation (42) by letting V_{11} have any joint distribution with V_{12} . 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 (43) is convex and thus time sharing is not needed. With these simplifications, the region \mathcal{R}_{DMT}^{out} is now defined as the region in (44), 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 [23] please note that the index of the primary and cognitive user are reversed with respect

to our notation (i.e $1 \rightarrow 2$ and vice-versa). Consider the correspondences between the variables of [23, Thm. 1] and those of Theorem V.1 in Table VII to obtain the region \mathcal{R}_{DMT}^{out} defined as the set of rate pairs satisfying the inequalities in (45),

taken over the union of all distributions

$$p_{U_{2c}} p_{X_2|U_{2c}} p_{U_{1c}|X_2} p_{U_{1pb}|X_2} p_{X_1|X_2, U_{1c}, U_{1pb}}. \quad (46)$$

Next, we using the correspondences of the table and restrict the fully general input distribution of Theorem V.1 to match the more constrained factorization of (46), obtaining a region $\mathcal{R}_{RTD}^{in} \subseteq \mathcal{R}_{RTD}$ defined as the set of rate tuples satisfying the inequalities in (47) 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 $\mathcal{R}_{DMT}^{out} \subseteq \mathcal{R}_{RTD}^{in}$ 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.

- (47c)-(47a) vs (45c)-(45a): Noting the cancelation / interplay between the binning rates, we see that

$$((47c) - (47a)) - ((45d) - (45a)) = 0.$$

- (47d)-(47a) vs. (45d)-(45a):

$$\begin{aligned} &((47d) - (47a)) - ((45d) - (45a)) \\ &= -I(X_2; U_{1c}) + I(U_{1c}; X_2, U_{2c}) \\ &= I(U_{2c}; U_{1c}|X_2) \\ &= 0 \end{aligned}$$

- (47e)-(47a) vs. (45e)-(45a): again noting the cancelations,

$$((47e) - (47a)) - ((45e) - (45a)) = 0$$

- (47f) vs. (45f):

$$(47f) - (45f) = 0$$

$$R'_{21} = I(V_{21}; X_1, V_{11}, V_{12}|W) \quad (43a)$$

$$R'_{22} = I(V_{22}; X_1, V_{11}, V_{12}|W) \quad (43b)$$

$$R_{11} \leq I(Y_1; X_1, V_{11}|V_{12}, V_{21}|W) + I(V_{21}; X_1|V_{12}, W) \quad (43c)$$

$$R_{21} + R'_{21} \leq I(Y_1, X_1, V_{11}, V_{12}; V_{21}|W) \quad (43d)$$

$$R_{11} + R_{21} + R'_{21} \leq I(Y_1; X_1, V_{11}, V_{21}|V_{12}, W) + I(X_1; V_{21}|W) \quad (43e)$$

$$R_{11} + R_{21} + R'_{21} + R_{12} \leq I(Y_1; X_1, V_{11}, V_{21}, V_{12}|W) + I(X_1, V_{11}, V_{12}; V_{21}|W) \quad (43f)$$

$$R_{22} + R'_{22} \leq I(Y_2, V_{12}, V_{21}; V_{22}|W) \quad (43g)$$

$$R_{22} + R'_{22} + R_{21} + R'_{21} \leq I(Y_2, V_{12}; V_{22}, V_{21}|W) + I(V_{22}; V_{21}|W) \quad (43h)$$

$$R_{22} + R'_{22} + R_{21} + R'_{21} + R_{12} \leq I(Y_2; V_{22}, V_{21}, V_{12}|W) + I(V_{22}, V_{21}; V_{12}|W) \quad (43i)$$

$$R'_{21} = I(V_{21}; X'_1, V_{12}) \quad (44a)$$

$$R'_{22} = I(V_{22}; X'_1, V_{12}) \quad (44b)$$

$$R_{11} \leq I(Y_1; X'_1|V_{12}, V_{21}) + I(V_{21}; X_1|V_{12}) \quad (44c)$$

$$R_{21} + R'_{21} \leq I(Y_1, X'_1, V_{12}; V_{21}) \quad (44d)$$

$$R_{11} + R_{21} + R'_{21} \leq I(Y_1; X'_1, V_{21}|V_{12}) + I(X_1; V_{21}) \quad (44e)$$

$$R_{11} + R_{21} + R'_{21} + R_{12} \leq I(Y_1; X'_1, V_{21}, V_{12}) + I(X'_1, V_{12}; V_{21}) \quad (44f)$$

$$R_{22} + R'_{22} \leq I(Y_2, V_{12}, V_{21}; V_{22}) \quad (44g)$$

$$R_{22} + R'_{22} + R_{21} + R'_{21} \leq I(Y_2, V_{12}; V_{22}, V_{21}) + I(V_{22}; V_{21}) \quad (44h)$$

$$R_{22} + R'_{22} + R_{21} + R'_{21} + R_{12} \leq I(Y_2; V_{22}, V_{21}, V_{12}) + I(V_{22}, V_{21}; V_{12}) \quad (44i)$$

TABLE VII
ASSIGNMENT OF RVs OF APPENDIX G

RV, rate of Theorem V.1	RV, rate of [23, Thm. 1]	Comments
U_{2c}, R_{2c}	V_{12}, R_{12}	TX 2 \rightarrow RX 1, RX 2
U_{1c}, R_{1c}	V_{21}, R_{21}	TX 1 \rightarrow RX 1, RX 2
U_{1pb}, R_{1pb}	V_{22}, R_{22}	TX 1 \rightarrow RX 1
X_2, R_{2pa}	X'_1, R_{11}	TX 2 \rightarrow RX 2
$U_{2pb} = \emptyset, R'_{2pb} = 0$	-	TX 1 \rightarrow RX 2
$R'_{1c} = I(U_{1c}; X_2 U_{2c})$	$L_{21} - R_{21} = I(V_{21}; V_{11}, V_{12})$	Binning rate
$R'_{1pb} = I(U_{1pb}; X_2 U_{1c}, U_{2c})$	$L_{22} - R_{22} = I(V_{22}; V_{11}, V_{12})$	Binning rate
X_1	X_2	

$$R'_{1c} = I(U_{1c}; X_2, U_{2c}) \quad (45a)$$

$$R'_{1pb} = I(U_{1pb}; X_2, U_{2c}) \quad (45b)$$

$$R_{2pa} + R_{1c} + R'_{1c} + R_{2c} \leq I(Y_2; U_{1c}, U_{2c}, X_2) + I(X_2, U_{2c}; U_{1c}) \quad (45c)$$

$$R_{2pa} + R_{1c} + R'_{1c} \leq I(Y_2; X_2, U_{1c}|U_{2c}) + I(X_2; U_{1c}) \quad (45d)$$

$$R_{1c} + R'_{1c} \leq I(Y_2, X_2, U_{2c}; U_{1c}) \quad (45e)$$

$$R_{2pa} \leq I(Y_2; X_2|U_{2c}, U_{1c}) + I(U_{1c}; X_2|U_{2c}) \quad (45f)$$

$$R_{1pb} + R'_{1pb} + R_{1c} + R'_{1c} + R_{2c} \leq I(Y_1; U_{1pb}, U_{1c}, U_{2c}) + I(U_{1pb}, U_{1c}; U_{2c}) \quad (45g)$$

$$R_{1c} + R_{1pb} + R'_{1c} + R'_{1pb} \leq I(Y_1, U_{2c}; U_{1pb}, U_{1c}) + I(U_{1pb}; U_{1c}) \quad (45h)$$

$$R_{1pb} + R'_{1pb} \leq I(Y_1, U_{2c}, U_{1c}; U_{1pb}) \quad (45i)$$

$$R'_{1c} = I(U_{1c}; X_2|U_{2c}) \quad (47a)$$

$$R'_{1c} + R'_{1pb} = I(X_2; U_{1c}, U_{1pb}|U_{2c}) \quad (47b)$$

$$R_{2c} + R_{1c} + R_{2pa} + R'_{1c} \leq I(Y_2; U_{2c}, U_{1c}, X_2) + I(U_{1c}; X_2|U_{2c}) \quad (47c)$$

$$R_{2pa} + R_{1c} + R'_{1c} \leq I(Y_2; U_{1c}, X_2|U_{2c}) + I(U_{1c}; X_2|U_{2c}) \quad (47d)$$

$$R_{1c} + R'_{1c} \leq I(Y_2; U_{1c}|U_{2c}, X_2) + I(U_{1c}; X_2|U_{2c}) \quad (47e)$$

$$R_{2pa} \leq I(Y_2; X_2|U_{2c}, U_{1c}) + I(U_{1c}; X_2|U_{2c}) \quad (47f)$$

$$R_{1pb} + R'_{1pb} + R_{1c} + R'_{1c} + R_{2c} \leq I(Y_1; U_{2c}, U_{1c}, U_{1pb}) \quad (47g)$$

$$R_{1c} + R_{1pb} + R'_{1c} + R'_{1pb} \leq I(Y_1; U_{1c}, U_{1pb}|U_{2c}) \quad (47h)$$

$$R_{1pb} + R'_{1pb} \leq I(Y_1; U_{1pb}|U_{2c}, U_{1c}) \quad (47i)$$

- (47g)-(47b) vs. (45g)-(45b)-(45a)

$$\begin{aligned} & ((47g) - (47b)) - ((45g) - (45b) - (45a)) \\ &= -I(X_2; U_{1c}, U_{1pb}|U_{2c}) \\ &\quad -I(U_{1pb}, U_{1c}; U_{2c}) + I(U_{1c}; U_{2c}, X_2) \\ &\quad +I(U_{1pb}; U_{2c}, X_2) \\ &= -I(U_{1pb}, U_{1c}; X_2, U_{2c}) + I(U_{1c}; U_{2c}, X_2) \\ &\quad +I(U_{1pb}; U_{2c}, X_2) \\ &= -I(U_{1pb}; X_2, U_{2c}) - I(U_{1c}; X_2, U_{2c}|U_{1pb}) \\ &\quad +I(U_{1c}; U_{2c}, X_2) + I(U_{1pb}; U_{2c}, X_2) \\ &= -I(U_{1c}; X_2, U_{2c}|U_{1pb}) + I(U_{1c}; U_{2c}, X_2) \\ &= -H(U_{1c}|U_{1pb}) + H(U_{1c}|X_2, U_{2c}, U_{1pb}) \\ &\quad +H(U_{1c}) - H(U_{1c}|X_2, U_{2c}) \\ &= I(U_{1c}; U_{1pb}) > 0 \end{aligned}$$

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

- (47h) – (47b) vs. (45h) – (45b) – (45a):

$$\begin{aligned} & ((47h) - (47b)) - ((45h) - (45b) - (45a)) \\ &= -I(X_2; U_{1c}, U_{1pb}|U_{2c}) - I(U_{2c}; U_{1c}, U_{1pb}) \\ &\quad +I(U_{1pb}; U_{2c}, X_2) - I(U_{1pb}; U_{1c}) \\ &\quad +I(U_{1c}; X_2, U_{2c}) \\ &= -I(X_2, U_{2c}; U_{1c}, U_{1pb}) + I(U_{1pb}; U_{2c}, X_2) \\ &\quad -I(U_{1pb}; U_{1c}) + I(U_{1c}; X_2, U_{2c}) \\ &= -I(X_2, U_{2c}; U_{1pb}) - I(U_{1c}; X_2, U_{2c}|U_{1pb}) \\ &\quad +I(U_{1pb}; U_{2c}, X_2) - I(U_{1pb}; U_{1c}) \\ &\quad +I(U_{1c}; X_2, U_{2c}) \\ &= -I(U_{1c}; X_2, U_{2c}, U_{1pb}) + I(U_{1c}; X_2, U_{2c}) \\ &= -I(U_{1c}; X_2, U_{2c}) - I(U_{1c}; U_{1pb}|X_2, U_{2c}) \\ &\quad +I(U_{1c}; X_2, U_{2c}) \\ &= 0 \end{aligned}$$

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

- (47i) – (47b) + (47a) vs. (45i) – (45b):

$$\begin{aligned} & ((47i) - (47b) + (47a)) - ((45i) - (45b)) \\ &= -I(U_{1pb}; X_2|U_{2c}, U_{1c}) - I(U_{1pb}; U_{2c}, U_{1c}) \\ &\quad +I(U_{1pb}; X_2, U_{2c}) \\ &= -I(U_{1pb}; X_2, U_{2c}, U_{1c}) + I(U_{1pb}; U_{2c}, X_2) \\ &= -I(U_{1pb}; U_{1c}|U_{2c}, X_2) \\ &= 0 \end{aligned}$$

H. Containment of [24, Thm. 2] in \mathcal{R}_{RTD} of Section H

The independently derived region in [12, Thm. 2] uses a similar encoding structure as that of \mathcal{R}_{RTD} with two exceptions: a) the binning is done sequentially rather than jointly as in \mathcal{R}_{RTD} leading to binning constraints (43)–(45) in [12, Thm. 2] as opposed to (10a)–(10c) in Thm.V.1. 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 2 parts in [12, Thm. 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, Thm.2], denoted as $\mathcal{R}_{CC} \subseteq \mathcal{R}_{RTD}$ in two steps:

- We first show that we may WLOG set $U_{11} = \emptyset$ in [12, Thm.2], creating a new region R'_{CC} .
- We next make a correspondence between our RVs and those of [12, Thm.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, Thm. 2] may be dropped WLOG. Consider the region \mathcal{R}_{CC} of [12, Thm. 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:

$$R_1 \leq I(Y_1; V_{11}, U_{11}, V_{20}, U_{10}) \quad (48)$$

$$R_2 \leq I(Y_2; V_{20}, V_{22}|U_{10}) - I(V_{22}, V_{20}; U_{11}|U_{10}) \quad (49)$$

$$\begin{aligned} R_1 + R_2 &\leq I(Y_1; V_{11}, U_{11}|V_{20}, U_{10}) + I(Y_2; V_{22}, V_{20}, U_{10}) \\ &\quad - I(V_{22}; U_{11}, V_{11}|V_{20}, U_{10}) \end{aligned} \quad (50)$$

$$\begin{aligned} R_1 + R_2 &\leq I(Y_1; V_{11}, U_{11}, V_{20}, U_{10}) + I(Y_2; V_{22}|V_{20}, U_{10}) \\ &\quad - I(V_{22}; U_{11}, V_{11}|V_{20}, U_{10}) \end{aligned} \quad (51)$$

$$\begin{aligned} 2R_2 + R_1 &\leq I(Y_1; V_{11}, U_{11}, V_{20}|U_{10}) + I(Y_2; V_{22}|V_{20}, U_{10}) \\ &\quad + I(Y_2; V_{20}, V_{22}, U_{10}) - I(V_{22}, V_{20}; U_{11}|U_{10}) \\ &\quad - I(V_{22}; U_{11}, V_{11}|V_{20}, U_{10}) \end{aligned} \quad (52)$$

Now let \mathcal{R}'_{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}'_{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}_{CC} , of all rate tuples satisfying:

$$R_1 \leq I(Y_1; V_{11}, U_{11}, V_{20}, U_{10}) \quad (53)$$

$$R_2 \leq I(Y_2; V_{20}, V_{22}|U_{10}) \quad (54)$$

$$R_1 + R_2 \leq I(Y_1; V_{11}, U_{11}|V_{20}, U_{10}) + I(Y_2; V_{22}, V_{20}, U_{10}) - I(V_{22}; U_{11}, V_{11}|V_{20}, U_{10}) \quad (55)$$

$$R_1 + R_2 \leq I(Y_1; V_{11}, U_{11}, V_{20}, U_{10}) + I(Y_2; V_{22}|V_{20}, U_{10}) - I(V_{22}; U_{11}, V_{11}|V_{20}, U_{10}) \quad (56)$$

$$2R_2 + R_1 \leq I(Y_1; V_{11}, U_{11}, V_{20}|U_{10}) + I(Y_2; V_{22}|V_{20}, U_{10}) + I(Y_2; V_{20}, V_{22}, U_{10}) - I(V_{22}; U_{11}, V_{11}|V_{20}, U_{10}) \quad (57)$$

Comparing the two regions equation by equation, we see that

- (48) = (53)
- (49) < (54) as this choice of RVs sets the generally positive mutual information to 0
- (50) = (55)
- (51) = (56)
- (52) < (57) 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}_{CC} of [12, Thm. 2] without loss of generality, obtaining the region \mathcal{R}'_{CC} defined in (53) – (57). We show that \mathcal{R}'_{CC} may be obtained from the region \mathcal{R}_{RTD} with the assignment of RVs, rates and binning rates in Table VIII.

Evaluating \mathcal{R}'_{CC} defined by (53) – (57) with the above assignment, translating all RVs into the notation used here, we obtain the region:

$$R'_{1c} \geq 0 \quad (58a)$$

$$R'_{1pb} + R'_{2pb} \geq I(U_{1pb}; U_{2pb}|U_{2c}, U_{1c}) \quad (58b)$$

$$R_{2pb} + R'_{2pb} \leq I(Y_2; U_{2pb}|U_{2c}, U_{1c}) \quad (58c)$$

$$R_{2pb} + R'_{2pb} + R_{1c} + R'_{1c} \leq I(Y_2; U_{1c}, U_{2pb}|U_{2c}) \quad (58d)$$

$$R_{2pb} + R'_{2pb} + R_{1c} + R'_{1c} + R_{2c} \leq I(Y_2; U_{1c}, U_{2c}, U_{2pb}) \quad (58e)$$

$$R_{1pb} + R'_{1pb} \leq I(Y_1; U_{1pb}|U_{2c}, U_{1c}) \quad (58f)$$

$$R_{1pb} + R'_{1pb} + R_{1c} + R'_{1c} \leq I(Y_1; U_{1pb}, U_{1c}|U_{2c}) \quad (58g)$$

$$R_{1pb} + R'_{1pb} + R_{1c} + R'_{1c} + R_{2c} \leq I(Y_1; U_{1pb}, U_{1c}, U_{2c}) \quad (58h)$$

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}_{RTD} with $R_{2pa} = 0$

$$R'_{1c} \geq 0 \quad (59a)$$

$$R'_{1c} + R'_{1pb} \geq 0 \quad (59b)$$

$$R'_{1c} + R'_{1pb} + R'_{2pb} \geq I(U_{1pb}; U_{2pb}|U_{2c}, U_{1c}) \quad (59c)$$

$$R_{2pb} + R'_{2pb} \leq I(Y_2; U_{2pb}|U_{2c}, U_{1c}) \quad (59d)$$

$$R_{2pb} + R'_{2pb} + R_{1c} + R'_{1c} \leq I(Y_2; U_{1c}, U_{2pb}|U_{2c}) \quad (59e)$$

$$R_{2pb} + R'_{2pb} + R_{1c} + R'_{1c} + R_{2c} \leq I(Y_2; U_{1c}, U_{2c}, U_{2pb}) \quad (59f)$$

$$R_{1pb} + R'_{1pb} \leq I(Y_1; U_{1pb}|U_{2c}, U_{1c}) \quad (59g)$$

$$R_{1pb} + R'_{1pb} + R_{1c} + R'_{1c} \leq I(Y_1; U_{1pb}, U_{1c}|U_{2c}) \quad (59h)$$

$$R_{1pb} + R'_{1pb} + R_{1c} + R'_{1c} + R_{2c} \leq I(Y_1; U_{1pb}, U_{1c}, U_{2c}) \quad (59i)$$

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

\mathcal{R}_{RTD} may produce rates larger than \mathcal{R}_{CC} . However, in general, no strict inclusion of \mathcal{R}_{CC} in \mathcal{R}_{RTD} has been shown.

I. Containment of [13, Thm. 4.1] in \mathcal{R}_{RTD} of Section VI-C

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}_{RTD} .

Again, following the argument of [30, Appendix 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, Thm. 4.1] as:

$$R'_{22} \geq I(W_2; V_1, X_1|U_1, U_2) \quad (60a)$$

$$R'_{11} + R'_{22} \geq I(W_2; W_1, V_1, X_1|U_1, U_2) \quad (60b)$$

$$R_{11} + R'_{11} \leq I(V_1, X_1, W_1; Y_1|U_1, U_2) \quad (60c)$$

$$R_{12} + R_{11} + R'_{11} \leq I(U_1, V_1, X_1, W_1; Y_1|U_2) \quad (60d)$$

$$R_{21} + R_{11} + R'_{11} \leq I(U_2, V_1, X_1, W_1; Y_1|U_1) \quad (60e)$$

$$R_{12} + R_{21} + R_{11} + R'_{11} \leq I(U_1, V_1, X_1, W_1, U_2; Y_1) \quad (60f)$$

$$R_{22} + R'_{22} \leq I(W_2; Y_2|U_1, U_2) \quad (60g)$$

$$R_{21} + R_{22} + R'_{22} \leq I(U_2, W_2; Y_2|U_1) \quad (60h)$$

$$R_{12} + R_{22} + R'_{22} \leq I(U_1, W_2; Y_2|U_2) \quad (60i)$$

$$R_{12} + R_{21} + R_{22} + R'_{22} \leq I(U_1, U_2, W_2; Y_2). \quad (60j)$$

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

$$R'_{22} \geq I(W_2; V'_1|U_1, U_2) \quad (61a)$$

$$R'_{11} + R'_{22} \geq I(W_2; W_1, V'_1|U_1, U_2) \quad (61b)$$

$$R_{11} + R'_{11} \leq I(V'_1, W_1; Y_1|U_1, U_2) \quad (61c)$$

$$R_{12} + R_{11} + R'_{11} \leq I(U_1, V'_1, W_1; Y_1|U_2) \quad (61d)$$

$$R_{21} + R_{11} + R'_{11} \leq I(U_2, V'_1, W_1; Y_1|U_1) \quad (61e)$$

$$R_{12} + R_{21} + R_{11} + R'_{11} \leq I(U_1, V'_1, W_1, U_2; Y_1) \quad (61f)$$

$$R_{22} + R'_{22} \leq I(W_2; Y_2|U_1, U_2) \quad (61g)$$

$$R_{21} + R_{22} + R'_{22} \leq I(U_2, W_2; Y_2|U_1) \quad (61h)$$

$$R_{12} + R_{22} + R'_{22} \leq I(U_1, W_2; Y_2|U_2) \quad (61i)$$

$$R_{12} + R_{21} + R_{22} + R'_{22} \leq I(U_1, U_2, W_2; Y_2) \quad (61j)$$

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 V.1 as in Table IX.

TABLE VIII
ASSIGNMENT OF RVs OF SECTION H

RV, rate of Theorem V.1	RV, rate of [23, Thm. 1]	Comments
U_{2c}, R_{2c}	U_{10}, R_{10}	TX 2 \rightarrow RX 1, RX 2
$X_2 = U_{2c}, R_{2pa} = 0$	$U_{11} = \emptyset, R_{11} = 0$	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}	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	

TABLE IX
ASSIGNMENT OF RVs OF SECTION I

RV, rate of Theorem V.1	RV, rate of [23, Thm. 1]	Comments
U_{2c}, R_{2c}	U_1, R_{12}	TX 2 \rightarrow RX 1, RX 2
X_2, R_{2pa}	V'_1, R'_{11}	TX 2 \rightarrow RX 2
U_{1c}, R_{1c}	U_2, R_{21}	TX 1 \rightarrow RX 1, RX 2
U_{1pb}, R_{1pb}	W_2, R_{22}	TX 1 \rightarrow RX 1
$U_{2pb}, R_{2pb} = 0$	W_1	TX 1 \rightarrow RX 2
R'_{1c}	$L_{20} - R_{20}$	
R'_{1pb}	$L_{11} - R_{11}$	
R'_{2pb}	$L_{22} - R_{22}$	
X_1	X_0	
X_2	X_1	

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

$$R'_{1pb} \geq I(U_{1pb}; X_2 | U_{2c}, U_{1c}) \quad (62a)$$

$$R'_{1pb} + R'_{2pb} \geq I(U_{1pb}; U_{2pb}, X_2 | U_{2c}, U_{1c}) \quad (62b)$$

$$R_{2pa} + R'_{2pb} \leq I(X_2, U_{2pb}; Y_2 | U_{2c}, U_{1c}) \quad (62c)$$

$$R_{2c} + R_{2pa} + R'_{2pb} \leq I(U_{2c}, X_2, U_{2pb}; Y_2 | U_{1c}) \quad (62d)$$

$$R_{1c} + R_{2pa} + R'_{2pb} \leq I(U_{1c}, X_2, U_{2pb}; Y_2 | U_{2c}) \quad (62e)$$

$$R_{2c} + R_{1c} + R_{2pa} + R'_{2pb} \leq I(U_{2c}, X_2, U_{1c}, U_{1pb}; Y_2) \quad (62f)$$

$$R_{1pb} + R'_{1pb} \leq I(U_{1pb}; Y_1 | U_{2c}, U_{1c}) \quad (62g)$$

$$R_{1c} + R_{1pb} + R'_{1pb} \leq I(U_{1c}, U_{1pb}; Y_1 | U_{2c}) \quad (62h)$$

$$R_{2c} + R_{1pb} + R'_{1pb} \leq I(U_{2c}, U_{1pb}; Y_1 | U_{1c}) \quad (62i)$$

$$R_{2c} + R_{1c} + R_{1pb} + R'_{1pb} \leq I(U_{2c}, U_{1c}, U_{1pb}; Y_1) \quad (62j)$$

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 V.1 and consider the factorization of the remaining RVs as in the scheme of (62), 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 region

$$R'_{1c} = I(U_{1c}; X_2 | U_{2c}) \quad (63a)$$

$$R'_{1pb} \geq I(U_{1pb}; X_2 | U_{2c}, U_{1c}) \quad (63b)$$

$$R'_{1pb} + R'_{2pb} \geq I(U_{1pb}; X_2, U_{2pb} | U_{2c}, U_{1c}) \quad (63c)$$

$$R_{2pa} + R'_{2pb} \leq I(Y_2; X_2, U_{2pb} | U_{2c}, U_{1c}) + I(U_{1c}; X_2 | U_{2c}) \quad (63d)$$

$$R_{1c} + R_{2pa} + R'_{2pb} \leq I(Y_2; U_{1c}, X_2, U_{2pb} | U_{2c}) \quad (63e)$$

$$R_{2c} + R_{1c} + R_{2pa} + R'_{2pb} \leq I(Y_2; U_{2pb}, U_{1c}, U_{2c}, X_2) \quad (63f)$$

$$R_{1pb} + R'_{1pb} \leq I(Y_1; U_{1pb} | U_{2c}, U_{1c}) \quad (63g)$$

$$R_{1c} + R_{1pb} + R'_{1pb} \leq I(Y_1; U_{1c}, U_{1pb} | U_{2c}) \quad (63h)$$

$$R_{2c} + R_{1c} + R_{1pb} + R'_{1pb} \leq I(Y_1; U_{2c}, U_{1c}, U_{1pb}) \quad (63i)$$

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 (62) and (63) for a fixed input distribution, equation by equation:

$$(63b) = (62a)$$

$$(63c) = (62b)$$

$$(63d) = (62c)$$

$$(63e) = (62e)$$

$$(63f) = (62f)$$

$$(63g) = (62g)$$

$$(63h) = (62h)$$

$$(63i) = (62j)$$

We see that (62d) and (62i) are extra bounds that further restrict the region in [13] to be contained in the region of Theorem V.1.