On the Capacity of Interference Channel With Causal and Noncausal Generalized Feedback at the Cognitive Transmitter

Mahtab Mirmohseni, Student Member, IEEE, Bahareh Akhbari, Member, IEEE, and Mohammad Reza Aref

Abstract—In this paper, taking into account the effect of link delays, we investigate the capacity region of the cognitive interference channel (C-IFC), where cognition can be obtained from either causal or noncausal generalized feedback. For this purpose, we introduce the causal C-IFC with delay (CC-IFC-WD) in which the cognitive user’s transmission can depend on future received symbols as well as the past ones. We show that the CC-IFC-WD model is equivalent to a classical causal C-IFC (CC-IFC) with link delays. Moreover, CC-IFC-WD extends both genie-aided and causal cognitive radio channels and bridges the gap between them. First, we derive an outer bound on the capacity region for the arbitrary value of $L$ and specialize this general outer bound to the strong interference case. Then, under strong interference conditions, we tighten the outer bound. To derive the achievable rate regions, we concentrate on three special cases: 1) classical CC-IFC ($L = 0$); 2) CC-IFC without delay ($L = 1$); and 3) CC-IFC with unlimited look-ahead in which the cognitive user noncausally knows its entire received sequence. In each case, we obtain a new inner bound on the capacity region. The derived achievable rate regions under special conditions reduce to several previously known results. Moreover, we show that the coding strategy which we use to derive an achievable rate region for the classical CC-IFC achieves the capacity for the classes of degraded and semideterministic classical CC-IFC under strong interference conditions. Furthermore, we extend our achievable rate regions to the Gaussian case. Providing some numerical examples for Gaussian CC-IFC-WD, we compare the performances of the different strategies and investigate the rate gain of the cognitive link for different delay values. We show that one can achieve larger rate regions in the “without delay” and “unlimited look-ahead” cases than in the classical CC-IFC; this improvement is likely due to the fact that, in the former two cases, the cognitive user can cooperate more effectively with the primary user by knowing the current and future received symbols.

Index Terms—Causal cognitive radio, Gel’fand–Pinsker coding, generalized block Markov coding, instantaneous relaying, interference channel (IFC), noncausal decode-and-forward.

I. INTRODUCTION

INTERFERENCE management is one of the key issues in wireless networks wherein multiple source–destination pairs share same medium and interfere with each other. Interference channel (IFC) [1] is the simplest model for this scenario, with two independent transmitters sending messages to their intended receivers. However, users with cognitive radio technology may sense the medium and use the obtained data to adapt their transmissions to cooperate with other users and improve their own rates as well as the rates of others. Cognitive IFC (C-IFC) refers to a two-user IFC in which the cognitive user (secondary user) has the ability to obtain the message being transmitted by the other user (primary user), in either a noncausal or causal manner. C-IFC was first introduced in [2], where for the noncausal C-IFC an achievable rate region was derived by combining Gel’fand–Pinsker (GP) binning [3] and a well-known simultaneous superposition coding scheme (rate splitting) applied to the IFC [4], which allows the receivers to decode part of the nonintended message.

Noncausal C-IFC, also termed genie-aided C-IFC, in which the cognitive user has noncausal full or partial knowledge of the other user’s transmitted message, has been widely investigated in [5]–[16] and the studies represented in the references therein. Yet, capacity results are known only in special cases. For an overview on the capacity results of the noncausal C-IFC, see [13], which contains the strongest results for the noncausal channel model. In the causal C-IFC (CC-IFC), the cognitive user can exploit knowledge of the primary user’s message from the causally received signals (information overheard by the feedback link from the channel and not that sent back from the receivers). Due to the complex nature of the problem, although CC-IFC is a more realistic and appropriate model for practical applications than the noncausal C-IFC, CC-IFC has been far less investigated in comparison to the latter [16]. In [2], achievable rate regions for the CC-IFC that consist of noncooperative causal transmission protocols have been characterized. An improved rate region for CC-IFC employing a cooperative coding strategy based on the block Markov superposition coding (full decode-and-forward (DF) [17]) and GP coding has been derived in [18]. Also, inner and outer bounds on the capacity region of CC-IFC have been derived in [19]. However, the problem of finding the capacity region of CC-IFC remains open. A more general model in which both transmitters are causally cognitive has been proposed in [20], called IFC with generalized feedback (IFC-GF).
in contrast to the output feedback, refers to the information overheard by the transmitter(s) over the channel and not to the information sent back by the receiver(s). Different achievable rate regions for IFC-GF have been obtained in [20]–[22], combining the methods of rate splitting, block Markov superposition coding, and GP binning. Moreover, outer bounds on the capacity region of the Gaussian cognitive Z-IFC were derived in the causal case [23]. It is noteworthy that in IFC-GF, cooperation between transmitters is performed using the links which share the same band as the links in IFC. Another scenario for transmitters cooperation is the case in which the cooperative links are orthogonal to each other as well as the links in IFC, termed conferencing. Multiple-access channel (MAC) with conferencing was first studied by Willems [24], and in [7] is extended to the compound MACs with conferencing encoders.

CC-IFC reveals the characteristics of the broadcast, MAC, and relay channels. Since an arbitrarily long delay is required to achieve the capacity, link delays have no effect on the capacity of broadcast and MAC. However, relaying structure may be changed by introducing link delays, and this can change the capacity of channels with relays [25]. Consider the classical CC-IFC in Fig. 1(a) and suppose that there are delays of $L_{\text{t}}$ units on the links between the primary user and the receivers, of $L_{\text{r}_1}$ units on the links between the primary user and the receivers, and of $L_{\text{r}_2}$ units on the link between transmitters. We refer to this channel as CC-IFC with link delays. We assume that all link delays are positive integers and the cognitive user hears the primary user’s transmitted signal earlier than do the receivers, i.e., $L_{\text{t}} \leq L_{\text{1}_1}$. A simple example which satisfies this assumption is shown in Fig. 2. We use this channel to obtain an information theoretical model which extends genie-aided and causal cognitive radio channels.

In order to obtain the information theoretical limits of cognitive radios, causal and noncausal C-IFC models attempt to capture the specifications of the cognitive radio technology [26], which aims at developing communication systems with the capability of sensing the environment and then adapting to it. For this purpose, researchers focus mostly on the noncausal C-IFC models. Moreover, despite the complex nature of the CC-IFC model, it is unsuited to all scenarios. In fact, due to the cognitive user’s cognitive ability, it may hear the primary user’s transmitted signal earlier than do the receivers, and the cognitive user can utilize this extra information to cooperate in sending or to precode against the primary user’s message.

The special features discussed previously motivate us to define the causal cognitive interference channel with delay (CC-IFC-WD) as an IFC where one of the transmitters can causally overhear the channel and its transmission depends on the future (noisy) received symbols as well as the past ones. This can be seen as the equivalent of the classical CC-IFC with time units of delay on the cognitive user’s received signal (or on the link between the transmitters). To physically motivate this channel model, we show that CC-IFC-WD [see Fig. 1(b)] is equivalent to the CC-IFC with link delays shown in Fig. 1(a), where $L_{\text{t}} \leq L_{\text{1}_1}$. As can be seen in Fig. 2, this channel model may fit wireless networks where the transmitters are close to each other or there is a high-speed link between them. Since setting $L = 0$ in the CC-IFC-WD model results in a classic causal model, CC-IFC-WD extends CC-IFC. Since, instead of the primary user’s message, a noisy version of the primary user channel input is provided to the cognitive user (when the cognitive user has unlimited look ahead and noncausally knows its entire received sequence), CC-IFC-WD also extends noncausal C-IFC. Therefore, CC-IFC-WD is a middle point between the genie-aided (noncausal) C-IFC and CC-IFC. In fact, a simple strategy which allows the users to cooperate instantaneously could be beneficial and could increase the channel capacity, as does the case in the relay with delay (RWD) channel [27]. The RWD channel has been vastly investigated in [25] and [27], wherein different upper and lower bounds and some capacity results have been derived. The lower bounds are achieved based on the combination of cooperative strategies such as full or partial DF, instantaneous relaying (for $L > 0$), where the relay sends a function of its current received symbol, and noncausal DF (for the unlimited look-ahead case), in which the relay predecodes part or all of the message at the beginning of the block and transmits the message to the receiver in cooperation with the source. A new general upper bound which holds for any arbitrary amount of delay has been derived in [28] and is shown to be tighter in some cases than the previously established bounds. It has been shown that the capacity of the discrete memoryless RWD channel is strictly larger than that of the classical relay channel [25, 27].

A. Main Contributions and Organization

In this paper, we study the IFC with causal and noncausal generalized feedback at the cognitive transmitter by defining the CC-IFC-WD. We derive new results regarding the capacity region of this channel for both discrete memoryless and Gaussian
cases. Our contributions in the rest of this paper are organized as follows.

1) We introduce the general CC-IFC-WD in Section II, where we also prove the equivalence of this channel model with CC-IFC with link delays.

2) In Section III, we first derive an outer bound on the capacity region of the new channel model (CC-IFC-WD) for an arbitrary value of $L$. Based on the fact that the receivers cannot cooperate, we use the idea in [29] in providing the cognitive receiver with a side information which has the same marginal distribution as the primary receiver’s signal and an arbitrary correlation with the cognitive receiver’s signal. This idea has been utilized in [30] to establish an outer bound on the capacity region of IFC-GF. We also make use of the techniques in [28] to incorporate the amount of the delay $L$. Next, we apply the strong interference condition at the primary receiver to the general outer bound in order to derive an outer bound under this condition, which is further tightened by setting the strong interference condition at the cognitive receiver.

3) To determine the achievable rate regions, we focus on three special cases in Section IV: 1) Classical CC-IFC which corresponds to $L = 0$; 2) CC-IFC without delay ($L = 1$); and 3) CC-IFC with unlimited look ahead.

4) A new inner bound for the classical CC-IFC is presented in Section IV-A. This bound is based on the coding schemes which combine cooperative, collaborative, and interference mitigating strategies. These strategies include rate splitting at both transmitters as in the Han–Kobayashi (HK) scheme [4], GP binning at the cognitive user, and generalized block Markov coding (partial DF) [17]. Next, we compare our scheme with the previous results and show that our scheme includes the scheme in [18] for CC-IFC, and the schemes in [20]–[22] tailored to CC-IFC.

5) In Section IV-B, we consider the CC-IFC without delay ($L = 1$), where the current received symbol (at the cognitive user) could also be utilized and present a new inner bound for this channel. Our coding scheme is based on the combination of the strategies in Section IV-A with instantaneous relaying. This means that the cognitive user, having access to the current received symbol, sends a function of its current received symbol and the codeword obtained by other strategies.

6) CC-IFC with unlimited look ahead, in which the cognitive user noncausally knows its entire received sequence, is investigated in Section IV-C. To obtain the achievable rate region, we employ noncausal partial DF strategy in which the cognitive user can contribute to the rate of the primary user by encoding a part of the primary user’s message and cooperating with the primary user to transmit this decoded part of the message. We remark that using a coding scheme based on instantaneous relaying is feasible for this case. However, to compare this strategy with noncausal partial DF, we restrict the use of this scheme to $L = 1$. When the cognitive link between transmitters is ideal, CC-IFC with unlimited look ahead reduces to a noncausal C-IFC. Therefore, we compare our proposed scheme with the results in [5]–[10], [31] for noncausal C-IFC, and show that our scheme encompasses most of the previous results and all of the capacity achieving schemes in [6] for weak interference, [7] for strong interference, and [31] for a class of Z cognitive channel.

7) In Section V, we derive the capacity regions for the classes of degraded and semideterministic classical CC-IFC under strong interference conditions, where achievability proofs follow from the region in Section IV-A, and for the converse parts, we evaluate the outer bound in Section III for $L = 0$.

8) In Section VI, Gaussian CC-IFC-WD is investigated where we extend the achievable rate regions of Section IV to the Gaussian case. Providing some numerical examples for Gaussian CC-IFC-WD, we investigate the rate gain of the cognitive link for different delay values. In addition, we compare the strategies used in our coding schemes and show that instantaneous relaying and noncausal DF improve the rate region noticeably.

9) Finally, Section VII concludes this paper.

II. CHANNEL MODELS AND PRELIMINARIES

Throughout this paper, the following notations are used:

* Upper case letters, e.g., $X$, are used to denote random variables (RVs) and lower case letters, e.g., $x$, show their realizations. The probability mass function (pmf) of an RV $X$ with alphabet $\mathcal{X}$ is denoted by $p_X(x)$, where the subscript $X$ is occasionally omitted. Additionally, $|\mathcal{X}|$ denotes the cardinality of a finite discrete set $\mathcal{X}$. $A^n(x,y)$ specifies the set of $n$-strings, jointly typical sequences of length $n$ on $p(x,y)$, abbreviated by $A^n$ if it is clear from the context. The notation $X^i_j$ indicates a sequence of RVs $(X_i, X_{i+1}, \ldots, X_j)$, where we use $X^i$ instead of $X^i_j$ for the sake of brevity. $N(0, \sigma^2)$ denotes the normal distribution with zero mean and variance $\sigma^2$.

Consider the CC-IFC-WD in Fig. 3 with finite input alphabets $\mathcal{X}_1, \mathcal{X}_2$, finite output alphabets $\mathcal{Y}_2, \mathcal{Y}_3, \mathcal{Y}_4$, and a channel transition probability distribution $p(y_2, y_3, y_4|x_1, x_2)$, denoted by $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_2, y_3, y_4|x_1, x_2))$, $\mathcal{Y}_2 \times \mathcal{Y}_3 \times \mathcal{Y}_4)$, where $X_1 \in \mathcal{X}_1$ and $X_2 \in \mathcal{X}_2$ are inputs of Transmitter 1 (Tx1) and Transmitter 2 (Tx2), respectively, $Y_2 \in \mathcal{Y}_2$ is the secondary user output, and $Y_3 \in \mathcal{Y}_3$ and $Y_4 \in \mathcal{Y}_4$ are channel outputs at the Receiver 1 (Rx1) and Receiver 2 (Rx2), respectively. In $n$ channel uses, each Transmitter $u_i(Tx_i)$ sends a message $m_i$ to the Receiver $u_i(Rx_i)$ for $i \in \{1, 2\}$.

**Definition 1:** A $(2^{nR_1}, 2^{nR_2}, n, I_{e}^{(n)})$ code for CC-IFC-WD consists of 1) two message sets $\mathcal{M}_1 = \{1, \ldots, 2^{nR_1}\}$ and $\mathcal{M}_2 = \{1, \ldots, 2^{nR_2}\}$ for the primary and secondary users, respectively; 2) an encoding function at the primary user $f_1: \mathcal{M}_1 \mapsto \mathcal{X}_1^n$; 3) a set of encoding functions at the secondary user $x_2,i = f_{2,i}(m_2,y_2^{i-1})$, for $1 \leq i \leq n$ and $m_2 \in \mathcal{M}_2$; 4) two decoding functions at Rx1 and Rx2, $g_1: \mathcal{Y}_3^n \mapsto \mathcal{M}_1$ and $g_2: \mathcal{Y}_2^n \mapsto \mathcal{M}_2$. We assume that the channel is memoryless. Thus, for $m_1 \in \mathcal{M}_1$, and $m_2 \in \mathcal{M}_2$, the joint pmf of $X_1 \times X_2 \times X'_1 \times X'_2 \times Y_2 \times Y_3 \times Y_4$ is given by
The CC-IFC with link delays is defined as a channel

\[
\begin{align*}
\tilde{y}_{3,i} = y_{3,i}, & \quad \tilde{y}_{4,i} = y_{4,i}, \quad \tilde{y}_{2,i} = y_{2,i} \quad \tilde{x}_{1,i} = x_{1,i}, \quad \tilde{x}_{2,i} = x_{2,i} \\
\tilde{y}_{3,i} = \tilde{y}_{3,i+L} = y_{2,i} & \quad \tilde{x}_{1,i+L} = x_{1,i}, \quad \tilde{x}_{2,i+L} = x_{2,i} \\
\end{align*}
\]

where \( L \leq L_{1r} \). Substituting the RVs of (3) into (2), the equivalent channel model of CC-IFC-WD is obtained as

\[
\begin{align*}
\tilde{y}_{2,i} &= y_{2,i} \\
\tilde{x}_{2,i} &= \tilde{f}_{2,i}(m_2, \tilde{y}_{2,1}, \ldots, \tilde{y}_{2,i}, \tilde{y}_{2,i+L-1}).
\end{align*}
\]

Comparing (4) with (1) completes the proof.

### III. OUTER BOUNDS ON THE CAPACITY REGION OF DISCRETE MEMORYLESS CC-IFC-WD

In this section, we investigate the outer bounds on the capacity region of CC-IFC-WD. Since the receivers cannot cooperate, we give the cognitive receiver a side information with the same marginal distribution as the primary receiver’s signal but an arbitrary correlation with the cognitive receiver’s signal, as in [29] and [30]. Based on this idea, and using the techniques in [28] for defining the auxiliary RVs, we first derive an outer bound on the capacity region of general CC-IFC-WD for arbitrary values of \( L \).

**Theorem 2:** The capacity region of CC-IFC-WD with the joint pmf (1), is contained in the region

\[
R_{\text{out}} = \bigcup_{p(u,v), p(x_1 | u), p(y_1 | x_1, v), f_2(x_2, y_2)} \left\{ (R_1, R_2) : R_1 \geq 0, R_2 \geq 0 \right\}
\]

\[
R_1 \leq I(X_1; Y_2 | V), I(X_1; Y_2 | X_2, Y_2, T) \\
R_2 \leq I(V; Y_2 | X_1, T) \\
R_2 \leq I(U, V; Y_2 | X_1, T) + I(U, V; Y_4 | X_1, Y_3^2) \\
R_1 + R_2 \leq I(X_1, U, V, T; Y_3) + I(U, V, T; Y_4 | X_1, Y_3)
\]

where \( x_2 = f_2(x_2, y_2, y_3) \). Also, \( Y_3^2 \) has the same marginal distribution of \( Y_3^1 \), i.e., \( p(y_3^2 | x_1^m, x_2^m) = p(y_3^2 | x_1^m, x_2^m) \), but \( p(y_3^2, y_4^2 | x_1^m, x_2^m) \) is an arbitrary joint distribution. We remark that, the dependence on \( L \) is through the input distribution.

**Remark 1:** Nullifying \( Y_4 \) and setting \( R_2 = 0 \), \( R_{\text{out}} \) reduces to the capacity upper bound derived in [28, Th. 1] for the RWD channel.

**Proof:** See Appendix A.

Now, we impose the strong interference condition at the primary receiver under which the interfering signal at Rx1 is strong enough that both messages can be decoded; we assume that the following strong interference condition hold

\[
I(X_2; Y_4 | X_1) \leq I(X_2; Y_3 | X_1).
\]

**Theorem 3:** The capacity region of CC-IFC-WD with the joint pmf (1), satisfying (9), is contained in the region...
\[
\mathcal{R}^{\text{str}}_{\text{out}} = \bigcup_{P(u|x)p(x|t)p(t|v),f_{2}^{2}(v,u,y_{2})} \left\{(R_{1},R_{2}) : R_{1} \geq 0, R_{2} \geq 0 \right\}
\]
\[
R_{1} \leq I(X_{1};Y_{2}|V,T) + I(X_{1};Y_{3}|X_{2},Y_{2},T)
\]  \hspace{1cm} (10)
\[
R_{2} \leq \min\{I(V;Y_{4}|X_{1},T),I(U,V;Y_{3}|X_{1},T)\}
\]  \hspace{1cm} (11)
\[
R_{1} + R_{2} \leq I(X_{1},U,V,T;Y_{3})
\]  \hspace{1cm} (12)

where \(x_{2} = f_{2}^{2}(v,u,y_{2})\).

**Proof:** The strong interference condition at (9) implies that
\[
I(X_{2};Y_{4}|X_{1},Y_{3}') \leq I(X_{2};Y_{3}|X_{1},Y_{3}').
\]  \hspace{1cm} (13)

Thus, we can compute the following mutual information term as
\[
I(U,V;T;Y_{4}|X_{1},Y_{3}') \leq I(U,V;T;X_{2};Y_{4}|X_{1},Y_{3}') \leq (a) I(X_{2};Y_{4}|X_{1},Y_{3}') \leq I(X_{2};Y_{3}|X_{1},Y_{3}').
\]

where (a) follows from the memoryless property of the channel and (b) from (13). Now, by substituting \(Y_{3}' = Y_{3}\), the region in Theorem 2 (\(\mathcal{R}^{\text{str}}_{\text{out}}\)) reduces to \(\mathcal{R}^{\text{str}}_{\text{out}}\).

Next, we apply the strong interference condition at the cognitive receiver to further tighten the outer bound which we use to derive the capacity results in Section V. Assume that the following strong interference condition at Rx2 hold
\[
I(X_{1};Y_{3}) \leq I(X_{1};Y_{4}).
\]  \hspace{1cm} (14)

**Theorem 4:** The capacity region of CC-IFC-WD with the joint pmf (1), satisfying (9) and (14), is contained in the region
\[
\mathcal{R}^{\text{str}}_{\text{out}} = \bigcup_{P(u|x)p(x|t)p(t|v),f_{2}^{2}(v,u,y_{2})} \left\{(R_{1},R_{2}) : R_{1} \geq 0, R_{2} \geq 0 \right\}
\]
\[
R_{1} \leq I(X_{1};Y_{2}|V,T) + I(X_{1};Y_{3}|X_{2},Y_{2},T)
\]  \hspace{1cm} (15)
\[
R_{2} \leq \min\{I(V;Y_{4}|X_{1},T),I(U,V;Y_{3}|X_{1},T)\}
\]  \hspace{1cm} (16)
\[
R_{1} + R_{2} \leq \min\{I(X_{1},U,V,T;Y_{3}),I(X_{1},U,V,T;Y_{4})\}
\]  \hspace{1cm} (17)

where \(x_{2} = f_{2}^{2}(v,u,y_{2})\).

**Proof:** See Appendix A.

**IV. INNER BOUNDS ON THE CAPACITY REGION OF DISCRETE MEMORYLESS CC-IFC-WD**

In this section, we consider the discrete memoryless CC-IFC-WD introduced in Section II and concentrate on three special cases: 1) classical CC-IFC, which corresponds to \(L = 0\); 2) CC-IFC without delay (\(L = 1\)), where the current received symbol (at the cognitive user) can also be utilized; and 3) CC-IFC with unlimited look ahead, in which the cognitive user knows its entire received sequence noncausally. For all setups, the inner bounds on the capacity region for the general discrete memoryless case are derived. For the first case, we utilize a coding scheme based on the combination of generalized block Markov superposition coding, rate splitting, and GP binning against part of the interference. In addition to the strategies used in the first case, we apply instantaneous relaying in the second setup due to the knowledge of the current received symbol at the cognitive user. Furthermore, we employ noncausal partial DF instead of generalized block Markov coding in the last case.

The outline of the proofs is presented. Also, we compare our proposed schemes with the results in [4]–[10], [18], [20]–[22], [27], and [31] and show that some previously known rate regions are included in our achievable rate regions.

**A. Classical CC-IFC (L = 0)**

We present a new achievable rate region for this setup. In our coding scheme, we employ the following strategies:

1) **Generalized block Markov coding (partial DF [17]):** In order to boost the rate of the primary user, the cognitive user can cooperate in sending the message of the primary user sent in the previous block.

2) **Rate splitting at both transmitters:** This allows the improvement of both rates through interference cancellation at both receivers as in the HK scheme [4]. Also, the cognitive user can partially decode the primary user’s message due to the splitting.

The message of the primary user (\(m_{1}\)) is split into four parts, i.e., \(m_{1} = (m_{1,\text{d}},m_{1,\text{c}},m_{1,\text{pl}},m_{1,\text{pm}})\). The private parts (\(m_{1,\text{pl}},m_{1,\text{pm}}\)) can be decoded only at the intended receiver (Rx1), while the common parts (\(m_{1,\text{dc}},m_{1,\text{cm}}\)) can be decoded at the nonintended receiver (Rx2) as well, allowing interference cancellation at Rx2. Note that, subscripts \(c\) (or \(p\)) refers to the common (or private) part of the message. Moreover, as Tx2 attempts to decode the primary user’s message via overhearing the channel (\(Y_{2}\)), we further consider two parts for partial decoding at the cognitive user (Tx2), where subscript \(d\) (or \(n\)) refers to the part of the primary user’s message which can (or cannot) be decoded by the cognitive user. Therefore, \((m_{1,\text{dc}},m_{1,\text{pl}})\) can be decoded at Tx2, and we refer to them as cooperative messages, while \((m_{1,\text{cm}},m_{1,\text{pm}})\) cannot be decoded at Tx2, and we refer to them as noncooperative messages.

The cognitive user splits its message (\(m_{2}\)) into two parts, i.e., \(m_{2} = (m_{2,\text{d}},m_{2,\text{p}})\), for interference cancellation at Rx1, where \(m_{2,\text{d}}\) and \(m_{2,\text{p}}\) are the common and private messages, as in the HK scheme [4].

3) **GP binning at the cognitive user:** The cognitive user predecodes its message against the part of the primary user’s message which was sent in the previous block and decoded by the cognitive user. This approach improves the rate of the cognitive user by correlated codebooks (using block Markov coding). Moreover, since the common message should be decoded in both receivers, binning against \(m_{2,\text{d}}\)
provides no improvement. Therefore, Tx2 generates codewords for $m_2c$ and $m_{2p}$, superimposing on $m_{1cd}$ in order to support its transmission, and bins its codewords against $m_{1pd}$ to precancel this part of the interference. Previous results generally focus on two binning techniques: in the first technique, two independent binning steps are applied for GP coding, as in [10] for noncausal C-IFC, while in the second technique, the second codeword is superimposed on the first binned one prior to the second binning step as in [9] for noncausal C-IFC. Instead, we use joint binning, which brings potential improvements.

Consider auxiliary RVs $T_c, T_p, U_{1c}, U_{1p}, V_{1c}, V_{1p}, U_{2c}, U_{2p}$, and a time-sharing RV $Q$ defined on arbitrary finite sets $T_c, T_p, U_{1c}, U_{1p}, V_{1c}, V_{1p}, U_{2c}, U_{2p}, X_1, X_2, Y_2, Y_3, Y_4$, and $\mathcal{P}_2$ denote the set of all joint pmfs $p(.)$ on $Z_1$ that can be factored in the form of

$$p(z_1) = p(y)p(t_c|y)p(t_p|t_c, y)p(u_{1c}|t_c, y)p(u_{1p}|u_{1c}, t_p, t_c, y)$$

$$p(v_{1c}|t_c, y)p(v_{1p}|v_{1c}, t_p, t_c, y)p(x_1|v_{1p}, v_{1c}, u_{1c}, t_p, t_c, y)$$

$$p(u_{2c}, t_p, t_c, y)p(x_2|u_{2c}, t_p, t_c, y)p(y_2, y_3, y_4|x_1, x_2).$$

(18)

Let $\mathcal{R}_1(Z_1)$ denote the set of all nonnegative rate pairs $(R_{1c}, R_2)$ where $R_{1c} = R_{1cd} + R_{1cm} + R_{1pd} + R_{1pm}$ and $R_2 = R_{2c} + R_{2p}$, such that there exist nonnegative $(L_{2c}, L_{2p})$ satisfying (19)–(39), shown at the bottom of the page.

**Theorem 5:** For any $p(.) \in \mathcal{P}_2$, the region $\mathcal{R}_1(Z_1)$ is an achievable rate region for the discrete memoryless classical CC-IFC (CC-IFC-WD with $L = 0$), i.e., $\bigcup_{Z_1 \in \mathcal{P}_2} \mathcal{R}_1(Z_1) \subseteq \mathcal{C}_0$.

**Outline of the Proof:** We propose the following random coding scheme, which contains regular generalized block Markov superposition coding, rate splitting, and GP coding in the encoding part. For decoding at the receivers, we utilize backward decoding. As mentioned earlier, messages of the primary and cognitive users are split into four and two parts, respectively, i.e., $m_1 = (m_{1ch}, m_{1cm}, m_{1pd}, m_{1pm})$ and $m_2 = (m_{2c}, m_{2p})$. Tx1 uses generalized block Markov superposition coding technique and creates $m_{1c}$, $m_{1p}$ codewords for cooperative messages of the previous block $(m_{1ch}, m_{1cm})$, $m_{2c}$, $m_{2p}$, for cooperative messages of the current block $(m_{1pd}, m_{1pm})$, and $v_{1c}$, $v_{1p}$, for noncooperative messages of the current block $(m_{1cm}, m_{1pm})$, where c in the subscript refers to a codeword related to the common part of the message (to be decoded at both receivers) and p refers to a codeword related to the private part of the message (to be decoded at the intended receiver only). At Tx1, all codewords related to the private messages are superimposed on the codewords related to the common messages. Note that, the cognitive user can decode $u_{2c}$, $u_{2p}$ using $t_{c} \cdot y_c$, $t_{p} \cdot y_p$, where $v_{1c}$, $v_{1p}$ are decoded at Rx2 and all of the aforementioned codewords are decoded at Rx1. Additionally, Tx2 encodes its split message.
with two codewords: joint binning against \( t_p^n \) conditioned on \( t_p^n \) is used to create \( u_{1,p}^n, u_{2,p}^n \) for \( m_{21}, m_{22} \), respectively. We remark that, in order to establish a cooperative strategy, all codewords are correlated due to block Markov scheme. The encoding scheme and relation between RVs are graphically shown in Fig. 4. Now, consider a block Markov encoding scheme with \( B \) blocks of transmission, each of \( n \) symbols.

**Codebook Generation:** Let \( q^n \) be a random sequence of \( Q^n \) according to the probability \( \prod_{i=1}^{n} p(q_i) \) and fix a joint pmf as (18).

**Primary User:**

1. Generate \( 2^{nR_{col}} \) independent and identically distributed (i.i.d) \( t_p^n \) sequences, each with probability \( \prod_{i=1}^{n} p(t_{c,i}, q_i) \).

   Index them as \( t_p^n(m_{1cd}^n) \) where \( m_{1cd}^n \in [1, 2^{nR_{col}}] \).

2. For each \( t_p^n(m_{1cd}^n) \), generate \( 2^{nR_{col}} \) i.i.d. \( t_p^n \) sequences, according to \( \prod_{i=1}^{n} p(t_{c,i}, q_i) \). Index them as \( u_{1,c}^n(m_{1cd}^n) \).

3. For each \( u_{1,c}^n(m_{1cd}^n) \), generate \( 2^{nR_{col}} \) i.i.d. \( u_{1,c}^n \) sequences, according to \( \prod_{i=1}^{n} p(u_{c,i}, t_{c,i}, q_i) \). Index them as \( u_{1,p}^n(m_{1cd}^n, m_{1cd}^n) \).

4. For each \( u_{1,c}^n(m_{1cd}^n) \), generate \( 2^{nR_{col}} \) i.i.d. \( u_{1,c}^n \) sequences, each with probability \( \prod_{i=1}^{n} p(u_{1,c,i}, t_{c,i}, q_i) \). Index them as \( u_{1,c}^n(m_{1cd}^n, m_{1cd}^n) \).

5. For each \( u_{1,c}^n(m_{1cd}^n, m_{1cd}^n) \), generate \( 2^{nR_{col}} \) i.i.d. \( u_{1,c}^n \) sequences, according to \( \prod_{i=1}^{n} p(u_{1,c,i}, t_{c,i}, q_i) \). Index them as \( u_{1,p}^n(m_{1cd}^n, m_{1cd}^n, m_{1cd}^n, m_{1cd}^n) \).

6. For each \( u_{1,c}^n(m_{1cd}^n, m_{1cd}^n, m_{1cd}^n, m_{1cd}^n) \), generate \( 2^{nR_{col}} \) i.i.d. \( u_{1,c}^n \) sequences, according to \( \prod_{i=1}^{n} p(u_{1,c,i}, t_{c,i}, q_i) \). Index them as \( u_{1,p}^n(m_{1cd}^n, m_{1cd}^n, m_{1cd}^n, m_{1cd}^n) \).

**Secondary User:**

1. For each \( u_{1,p}^n(m_{1cd}^n, m_{1cd}^n) \), generate \( 2^{n(R_{2c}+L_{2c})} \) i.i.d. \( u_{2,p}^n \) sequences, each with probability \( \prod_{i=1}^{n} p(u_{2,c,i}, t_{c,i}, q_i) \). Index them as \( u_{2,c}^n(m_{21}, m_{22}, m_{1cd}^n) \), where \( m_{2c} \in [1, 2^{nR_{2c}}] \) and \( L_{2c} \in [1, 2^{nL_{2c}}] \).

2. For each \( u_{2,c}^n(m_{1cd}^n) \), generate \( 2^{n(R_{2c}+L_{2c})} \) i.i.d. \( u_{2,c}^n \) sequences, according to \( \prod_{i=1}^{n} p(u_{2,c,i}, t_{c,i}, q_i) \). Index them as \( u_{2,c}^n(m_{21}, m_{22}, m_{1cd}^n) \), where \( m_{2c} \in [1, 2^{nR_{2c}}] \) and \( L_{2c} \in [1, 2^{nL_{2c}}] \).

**Encoding:** (At the Beginning of Block b):

**Primary User (Transmitter 1):** In order to transmit the message \( m_{1,b} = (m_{1ad,b}, m_{1cd,b}, m_{1cd,b}, m_{1cd,b}) \), encoder 1 picks codewords \( v_{1,c}^n(m_{1ad,b}, m_{1cd,b}, m_{1cd,b}, m_{1cd,b}) \), \( v_{1,c}^n(m_{1cd,b}, m_{1cd,b}, m_{1cd,b}, m_{1cd,b}) \), \( v_{1,c}^n(m_{1cd,b}, m_{1cd,b}, m_{1cd,b}, m_{1cd,b}) \), \( v_{1,c}^n(m_{1cd,b}, m_{1cd,b}, m_{1cd,b}, m_{1cd,b}) \).\( v_{1,c}^n(m_{1cd,b}, m_{1cd,b}, m_{1cd,b}, m_{1cd,b}) \).

We assume that in the first block, cooperative information is \( m_{1ad,b} = (m_{1ad,b}, m_{1cd,b}) = (0, 0) \), and in the last block, a previously known message \( m_{1ad,b} = (m_{1ad,b}, m_{1cd,b}) = (1, 1) \) is transmitted.

**Cognitive User (Transmitter 2):** Tx2 at the beginning of block b knows \( \tilde{m}_{1cd,b-1} \) and \( \tilde{m}_{1cd,b-1} \), which are estimates of the parts of the common and private messages sent by Tx1 in the previous block and can be decoded by the cognitive user. In order to send \( m_{2,b} = (m_{2ad,b}, m_{2cd,b}) \), encoder 2, knowing the codewords \( v_{1,c}^n(\tilde{m}_{1cd,b-1}, \tilde{m}_{1cd,b-1}) \) and \( v_{1,c}^n(\tilde{m}_{1cd,b-1}) \), seeks an index pair \( (\tilde{L}_{2b}, \tilde{L}_{2b}) \) such that

\[
(u_{2,c}^n([m_{2ad,b}, L_{2c}], m_{1cd,b-1}), v_{2,c}^n([m_{2ad,b}, L_{2c}], m_{1cd,b-1}))
\]

\[
(\tilde{m}_{1cd,b-1}, \tilde{m}_{1cd,b-1}, m_{1cd,b-1})
\]

\[
\in A_{c}^{n}(U_{2,c}, U_{2,c}, T_{c}, T_{c}, Q)
\]
sufficiently high probability, if \( n \) is sufficiently large and (19)–(21) hold. Then, Tx2 sends \( x_2^B \) generated according to \( \prod_{j=1}^{B} p(x_2^j|u_{2,j}^B, {\mathbf i}_{2,j^*}, t_{2,j^*}, q_j) \). The codewords at Tx1 and \( j^* \) Tx2 used in transmission are listed in Table I.

### Decoding:

**Cognitive User (Transmitter 2):** Tx2 at the end of block \( b \) wants to correctly recover \( m_{1,3dB,b} \) and \( m_{3,cd,b} \). Hence, it looks for a unique pair \((\hat{m}_{1,3dB,b}, \hat{m}_{3,cd,b})\), such that

\[
\begin{align*}
\{y_2^B(b), u_0^B(m_{1,3dB,b}, \hat{m}_{3,cd,b} , m_{1,3dB,b,1}, m_{3,cd,b,1})
\}
\end{align*}
\]

\[
\begin{align*}
\{\hat{y}_2^B, u_0^B(m_{1,3dB,b,1}, \hat{m}_{3,cd,b,1} , m_{1,3dB,b,1,1}, m_{3,cd,b,1,1})
\}
\end{align*}
\]

\[
\begin{align*}
\{u_2^B(m_{2,3dB,1}, l_{2,3dB,1}), t_2^B(m_{2,3dB,1,1}, l_{2,3dB,1,1}), q_2^B(m_{2,3dB,1,1}, q_2^B) \in A_0^B (Y_2, U_{2,1}, U_{2,2}, U_{1,1} , T_{2,1}, T_{2,2}, Q)
\}
\end{align*}
\]

This step can be accomplished with small enough probability of error, i.e., \((\hat{m}_{1,3dB,b}, \hat{m}_{3,cd,b}) = (m_{1,3dB,b}, m_{3,cd,b})\), for sufficiently large \( n \) if (38)–(39) hold.

Backward decoding is used at Rx1 and Rx2; hence, they begin to decode after all \( B \) blocks are received.

**Receiver 1:** In block \( b \), Rx1 looks for a unique quadruple \((m_{1,3dB,b}, m_{1,cd,b}, m_{1,3dB,b,1}, m_{3,cd,b,1})\) and some pair \((m_{2,3dB,1}, l_{2,3dB,1})\) such that

\[
\begin{align*}
\{y_1^B(b), u_0^B(m_{1,3dB,b}, m_{1,cd,b}, m_{1,3dB,b,1}, m_{3,cd,b,1})
\}
\end{align*}
\]

\[
\begin{align*}
\{u_1^B(m_{1,3dB,b,1}, m_{1,cd,b,1}, m_{1,3dB,b,1,1}, m_{3,cd,b,1,1})
\}
\end{align*}
\]

\[
\begin{align*}
\{t_1^B(m_{2,3dB,1,1}, l_{2,3dB,1,1}), q_1^B(m_{2,3dB,1,1}, q_1^B) \in A_0^B (Y_1, U_{1,1}, U_{1,2}, U_{2,1}, T_{1,1}, T_{1,2}, Q)
\}
\end{align*}
\]

where \((m_{1,3dB,b}, m_{1,cd,b})\) were decoded in the previous step of backward decoding (i.e., block \( b + 1 \)). Here, for large enough \( n \), the probability of error can be made sufficiently small if (22)–(30) hold.

**Receiver 2:** In block \( b \), Rx2 finds a unique triple \((m_{2,3dB,1}, m_{3,cd,b,1})\) and some triple \((l_{2,3dB,1}, l_{2,3dB,1}, m_{1,cd,b})\) such that

\[
\begin{align*}
\{y_2^B(b), u_0^B(m_{2,3dB,1}, l_{2,3dB,1}, m_{3,cd,b,1})
\}
\end{align*}
\]

\[
\begin{align*}
\{u_2^B(m_{3,cd,b,1}, l_{2,3dB,1,1}, m_{3,cd,b,1,1})
\}
\end{align*}
\]

\[
\begin{align*}
\{t_2^B(m_{2,3dB,1,1}, l_{2,3dB,1,1}), q_2^B(m_{2,3dB,1,1}, q_2^B) \in A_0^B (Y_2, U_{2,1}, U_{2,2}, U_{1,1} , T_{2,1}, T_{2,2}, Q)
\}
\end{align*}
\]

where \(m_{1,cd,b}\) was decoded in the previous step of backward decoding (i.e., block \( b + 1 \)). Note that, since \(m_{1,cd}\) plays a fundamental role in the backward decoding, it is necessary for Rx2 to correctly decode \( m_{3,cd,b,1} \). However, this causes no additional constraint on the rate region. With an arbitrarily high probability, no error occurs in the second receiver if \( n \) is sufficiently large and (31)–(37) hold. In Appendix B, we provide the complete error analysis.

**Remark 2 (Comparison With Existing Results):** Now, we compare the scheme of Theorem 5 with the known results for CC-IFC and special cases of this channel and show that Theorem 5 includes the rate regions of the following schemes:

1. **The HK Region [4]:** Consider the case where the cognitive user cannot overhear the channel, i.e., \( Y_2 = \emptyset \). If we set \( T_c = T_p = U_{1,1} = U_{1,2} = \emptyset \) and \( L_{2,1} = L_{2,2} = R_{1,cd} = R_{1,3dB} = 0 \), rename \( V_1^B = X_1 \), and define \( X_2 \) as a deterministic function of \( U_{2,1} \) and \( U_{2,2} \), then the derived rate region reduces to the HK region.

2. **The Relay Channel:** If we omit Rx2, i.e., \( Y_4 = \emptyset \), and the cognitive user has no message to transmit, i.e., \( R_2 = 0 \), then the model reduces to the relay channel. By setting \( T_c = T_p = U_{1,1} = U_{1,2} = \emptyset \) and \( L_{2,1} = L_{2,2} = R_{1,cd} = R_{1,3dB} = 0 \), and redefining \( U_{2,1} = X_2 \), the rate region reduces to the partial DF rate for the relay channel [17], which includes the capacity regions of the degraded [17] and semideterministic relay channels [33]. Note that (26), (28), and (35)–(37) can be dropped, because these bounds correspond to the decoding of the common message from the nonintended transmitter. Hence, these events cause no error unless another intended message is incorrectly decoded.

3. **The Region in [18] for CC-IFC (\( R_{SJXW} \)):** Scheme in [18] to achieve \( R_{SJXW} \) differs from our scheme to achieve \( R_1 \) in the following:

   a) The message of the primary user in \( R_{SJXW} \) is fully decoded by the cognitive user; therefore, \( m_1 \) is split into two parts. While in \( R_1 \), we use partial DF and split \( m_1 \) into four parts, in which we can achieve the scheme of \( R_{SJXW} \) by nullifying extra parts. By introducing two extra parts that are sent directly to the receivers, we aim to achieve a reasonable rate region (no less than IFC) even when the condition of the cognitive link is poor.

   b) In \( R_{SJXW} \), the codewords conveying the private and common messages are generated independently. However, we use superposition encoding on the codewords related to the private messages by using codewords related to the common messages as cloud centers. Thus, we derive a potentially larger achievable rate region with a simpler description.

### Table I

| \( b \) | \( m_{1,3dB,b} \) | \( m_{1,cd,b} \) | \( m_{1,3dB,b,1} \) | \( m_{3,cd,b,1} \) |
|---|---|---|---|---|
| block 1 | block 2, \( b = 2, \ldots, B - 1 \) | block \( B \) |
c) The codewords of Tx2 in \( R_{S,JXW} \) are generated independently and binned against all codewords of Tx1. However, in \( R_1 \) we generate the codewords of Tx2 \((U_{2x}, U_{2p})\) by superimposing them on the common cooperative codeword of Tx1 \((U_{1c})\) and then binning them against the private cooperative codeword of Tx1 \((U_{1p})\) conditioned on \( U_{1c} \). Thus, \( R_3 \) can be reduced to \( R_{S,JXW} \) if \( U_{2x} \), \( U_{2p} \), and \( U_{1c} \) are generated independently. Note that, since common message should be decoded by both receivers, binning against the common message provides no improvement. A similar result has been concluded in [31] for the cognitive Z-IFC.

By setting \( V_{1c} = V_{1p} = 0 \) and \( R_{1cn} = R_{1pn} = 0 \), \( R_1 \) reduces to \( R'_1 \subseteq R_1 \). Note that, in this scenario (28) and (30) can be dropped, since they correspond to the incorrect decoding of the common message from the nonintended transmitter. Now, in the scheme of \( R_{S,JXW} \), generate \( U_1 \) and \( X_{11} \) conditioned on \( U_2 \) and \( X_{12} \). Then, bounds (3), (9), and (12) in [18] can be dropped and \( R_{S,JXW} \) is enlarged to a region \( R'_{S,JXW} \) as a result of removing these rate constraints \( R'_{S,JXW} \subseteq R_{S,JXW} \). Redefining \( T_x = U_1 \), \( U_{1p} = X_{11}, \ T_p = U_2, \ U_{1c} = X_{12}, \ U_{2c} = V_1, \ U_{2p} = V_2 \) in \( R'_1 \), one gets \( R'_{S,JXW} \subseteq R'_1 \). Therefore, \( R_{S,JXW} \subseteq R_1 \).

4) The Region in [21] Tailored to CC-IFC (\( R_{CC} \)): The region in [21] has been derived for IFC-GF and can be reduced to a region for CC-IFC. In order to perform this reduction, assume that Tx1 is the cognitive user and set \( Y_1 = G_1 = H_1 = W_1 = 0 \) and \( R_{13} = 0 \) in the region in [21] to obtain \( R_{CC} \). Note that, indices 1 and 2 are switched, due to the positions of the primary and cognitive users being switched in this model. \( R_{CC} \) is different from \( R_3 \) in that

1) in \( R_{CC} \), the primary user splits its message into three parts.

In fact, the cooperative message is private and is not decoded at the cognitive user’s receiver. This means that the cognitive user cannot decode the common message of the primary user;

2) the scheme in [21] is based on the irregular encoding/successive decoding technique, while we use the regular encoding/backward decoding [34]. The latter results in fewer RVs and a simpler scheme;

3) the binning in \( R_{CC} \) is done sequentially, in contrast to the joint binning technique employed in \( R_1 \), which brings potential improvement.

By setting \( U_{1c} = T_c = 0 \) and \( R_{1dt} = 0 \), and redefining \( U_{2c} = N_1, \ U_{2p} = M_1, \ T_p = S_2, \ U_{1p} = W_2, \ U_{1c} = U_{2c} \), and \( V_{1p} = V_2 \), \( R_1 \) reduces to a region which includes \( R_{CC} \) as a subset.

5) The Region in [20] Tailored to CC-IFC (\( R_T \)): Similar to the previous case, we reduce the region in [20] to CC-IFC, which has been originally derived for IFC-GF. We assume that Tx2 is the cognitive user and set \( V_2 = V_0 \) and \( R_{2x} = 0 \) in the region of [20] to obtain the reduced region \( R_T \) for CC-IFC. \( R_1 \) includes \( R_T \) as a special case, because

1) the message of the primary user is split into three parts in \( R_T \), i.e., the cognitive user only decodes a part of the common message and there is no cooperative private message;

2) there is no binning in the \( R_T \) scheme and the cognitive user acts simply as a relay for the primary user’s message.

Applying the assignments, \( R_{1pt} = L_{2c} = L_{2p} = 0, \ T_p = U_{1p} = 0, \ T_c = V_0, \ U_{1c} = V_1, \ U_{1p} = X_1, \ U_{2c} = U_2, \) and \( U_{2p} = X_2, \ R_1 \) reduces to \( R_T \).

6) The Region in [22] Tailored to CC-IFC (\( R_{Y-T} \)): Considering Tx2 as the cognitive user, we reduce the region in [22] to CC-IFC by setting \( V_2 = S_2 = Z_2 = 0 \) and \( R_{2x} = R_{2y} = 0 \) in the region of [22] to obtain the reduced region \( R_{Y-T} \) for CC-IFC. Moreover, by nullifying \( S_2 \), one can set \( R_{11c} = 0 \) since the first binning step in [22] can be omitted. The scheme of \( R_{Y-T} \) is different from \( R_3 \) in the following aspects.

1) In \( R_{Y-T} \), binning is done sequentially and conditionally, while \( R_3 \) utilizes joint binning technique. Therefore, our scheme achieves a potentially larger rate region compared to \( R_{Y-T} \).

2) We use joint backward decoding at the receivers, while two-step decoding is used for \( R_{Y-T} \). Joint decoding cannot have worse performance than the sequential ones.

By setting \( T_c = Q, T_p = S_1, \ U_{1c} = V_1, \ U_{1p} = Z_1, V_{1c} = U_1, V_{1p} = T_1, \ U_{2c} = U_2, \) and \( U_{2p} = T_2, \ R_3 \) reduces to a region which includes \( R_{Y-T} \) as a result of the earlier differences.

Now, in order to understand the shape of the achievable rate region, we give a compact expression for \( R_{3}(Z_1) \) which is easier to compute.

**Corollary 1**: The region \( R_3(Z_1) \), after Fourier–Motzkin elimination [32], can be expressed as

\[
R_1 \leq \min \left( \min (I_{21} + I_4 + I_{16}, I_5) - I_1, \ I_{31} + \min (I_4 + I_{17} - I_2, I_3) \right)
\]

\[
R_2 \leq \min \left( \min (I_{19}, I_{41} + \min (I_{10}, I_{13})) - I_1' \right)
\]

\[
R_1 + R_2 \leq \min \left( I_{14} + I_5, I_{15} + \min (I_7, I_8 - I_1), \ I_{21} + I_{17} + \min (I_{30}, I_4 + I_{13}), \ I_4 + \min (I_2 + I_8 + I_3 + I_15), \ I_{21} + I_{14} + \min (I_{12}, I_4 + I_{10}, I_{17} - I_2) - I_1' \right)
\]

\[
2R_1 + R_2 \leq \min \left( I_4 + I_{15} + \min (I_{60}, I_{11} - I_1), \ I_{21} + 2I_4 + I_{17} + I_{16}, \ I_4 + I_{17} + \min (I_{21} + I_{12}, I_5), \ I_{21} - I_1 \right)
\]

\[
R_1 + 2R_2 \leq \min \left( I_{21} + I_{10} + I_{14} + \min (I_{14} + I_{16}, I_{18}), \ I_{14} + I_{15} + \min (I_{20} + I_{10}, I_8) - 2I_1' \right)
\]

\[
2R_1 + 2R_2 \leq \min \left( I_4 + \min (I_{14} + I_{11}, I_{17} + I_8), \ I_{20} + I_{14} + \min (I_{60}, I_{11} - I_1) \right) + I_{21} + I_{15} - 2I_1' \]

\[
2R_1 + 3R_2 \leq I_{21} + I_{10} + 2I_{14} + I_{11} + I_{15} - 3I_1' \]

\[
3R_1 + 2R_2 \leq 2I_{21} + 2I_4 + I_{11} + I_{17} + I_{15} - 2I_1' \]

subject to \( I_1 \leq I_{16} \) and \( I_2 \leq I_{17} \),

where \( \{I_i, i = 1, \ldots, 21\} \) are defined in (19)–(39), and \( I_1' \leq \max (I_1 + I_2, I_3) \).

**B. CC-IFC Without Delay (\( L = 1 \))**

In this case, the cognitive user can utilize the current received symbol as well as the past ones in order to cooperate with the
primary user or reduce the interference effect. Note that the derived inner bound in Theorem 5 is an inner bound on the capacity region for the CC-IFC without delay. However, knowledge of the present received symbol may lead to the expectation of achieving higher rates using this additional information. Instantaneous relaying is a cooperative scheme which exploits only the current received symbol. In general, the cognitive user may need to utilize both the current and the past received symbols to obtain an optimal coding scheme for the CC-IFC without delay. Hence, we establish an achievable rate region based on a scheme which involves the superposition of the scheme used in Theorem 5 with instantaneous relaying. In fact, the cognitive user, knowing the current received symbol, sends a function of the codeword obtained by the scheme of Theorem 5 and its corresponding part in the desired channel. This method can improve the rate region by allowing the primary and cognitive users to cooperate instantaneously.

Consider auxiliary RVs $T_c$, $T_p$, $U_{1c}$, $U_{1p}$, $V_{1c}$, $V_{1p}$, $U_{2c}$, $U_{2p}$, $V_{2c}$, $V_{2p}$, and a time-sharing RV $Q$ defined on arbitrary finite sets $T_c$, $T_p$, $U_{1c}$, $U_{1p}$, $V_{1c}$, $V_{1p}$, $U_{2c}$, $U_{2p}$, and $Q$, respectively. Let $Z_2 = (Q, T_c, T_p, U_{1c}, U_{1p}, V_{1c}, V_{1p}, U_{2c}, U_{2p}, V_{2c}, V_{2p}, X_1, X_2, Y_2, Y_3, Y_4)$, and $P_2$ be the set of all joint prefix $p(.)$ on $Z_2$ that can be factored in the form of

$$p(z_2) = p(q)p(t_c|q)p(t_p|t_c, q)p(u_{1c}|t_c, q)p(u_{1p}|u_{1c}, t_p, t_c, q)$$

$$p(u_{2c}|t_c, q)p(u_{2p}|u_{1c}, t_p, t_c, q)p(x_1|u_{1p}, u_{2c}, u_{1c}, t_p, t_c, q)$$

$$p(u_{2p}|t_p, t_c, q)p(u_{2c}|u_{2p}, t_p, t_c, q)p(x_2|u_{2p}, y_2, q).$$

Remark 3: The cognitive user can now transmit the decoded part. This is possible only when the $y_2$ is collected at the beginning of the codebook generation.

Theorem 6: For any $p(.) \in P_2$, the region $R_2(Z_2)$ is an achievable rate region for the discrete memoryless CC-IFC without delay (CC-IFC-WD with $L = 1$), i.e.,

$$\bigcup_{Z_2 \in P_2} R_2(Z_2) \subseteq C_1.$$

Proof: The achievability proof follows by combining the scheme used in Theorem 5 and instantaneous relaying. Encoding and decoding follow the same lines as in Theorem 5. Hence, we only highlight the differences for the sake of brevity. The main difference is that during the codebook generation at the cognitive user (TX2), $x_2^\Pi$ (instead of $x_2^n$ in Theorem 5) is generated according to

$$\prod_{i=1}^{\Pi} p(y_2^i|y_2^i, \tilde{y}_2^i, \tilde{p}_i, \tilde{t}_i, q_i).$$

and in the encoding session, Transmitter 2 at time $i$ and upon receiving $y_2^i$, sends a deterministic function of $y_2^i$ and $y_2^i$, i.e., $x_2^i = f_2^i(y_2^i, y_2^i, q_i)$ where $f_2^i(y_2^i, y_2^i, q_i)$ has been fixed at the beginning of the codebook generation.

Remark 4: Here, TX2 sends a deterministic function of $y_2$. Therefore, a function (not necessarily deterministic) of $X_1$ is transmitted by TX2, which interferes at RX2. This scheme can boost $R_2$ as it allows RX2 to decode the unwanted message and to cancel the interference. Hence, we consider linear mapping for the Gaussian CC-IFC without delay in Section VI.

Remark 5: This scheme is feasible for any $L \geq 1$. Moreover, $f_2^i(y_2^i, y_2^i, q_i)$ can be extended to $x_2^i = f_2^i(y_2^i, y_2^i, \ldots, y_2^i, L-1; q_i)$.

Remark 6: Nullifying $Y_4$, $T_c$, $T_p$, $U_{1p}$, $V_{1p}$, and $U_{2p}$, and setting $R_2 = L_2 = L_{2p} = R_{1p} = R_{1pd} = 0$ and $U_{2c} = V_2$, the model and the achievable rate based on partial DF and instantaneous relaying for the RWD channel [27, Th. 2.5], which achieves all known capacity results for discrete memoryless RWD [25], [27]. We remark that, as discussed for the relay channel in Remark 2, (26), (28), (35)–(37) can be dropped in Theorem 6 for this scenario.

C. CC-IFC With Unlimited Look Ahead

Now, we investigate the CC-IFC with unlimited look ahead, defined in Section II. This means that the cognitive user noncausally knows its entire received sequence. We derive an achievable rate region using a coding scheme based on combining noncausal partial DF, rate splitting, and GP binning against part of the interference.

1) Noncausal partial DF: The cognitive user can contribute to the rate of the primary user by encoding a part of the primary user’s message and cooperating with the primary user to transmit the decoded part. This is possible only when the cognitive user noncausally has knowledge of the entire received sequence, as in the unlimited look-ahead case.

2) Rate splitting: Similar to the scheme used in Theorem 5, rate splitting is employed at both transmitters and the messages of the primary and cognitive users are split into four and two parts, respectively, i.e., $m_0 = (m_0ch, m_0cn, m_0pd, m_0pn)$ and $m_2 = (m_2c, m_2p)$. In fact, common (subscript $c$) and private (subscript $p$) parts are used for interference cancellation at the nonintended receivers as in the HK scheme [4], and cooperative (subscript $d$) and noncooperative (subscript $n$) parts account for noncausal partial DF strategy. Moreover, TX2 jointly bins its codewords against the cooperative private part of $m_2$, (i.e., $m_2pd$) to precancel this part of the interference at RX2.

3) GP binning at the cognitive user: The cognitive user can partially decode the primary user’s message in a noncausal manner, and its rate is improved by precoding against the (partially) known interference.

Consider auxiliary RVs $U_{1c}$, $U_{1p}$, $V_{1c}$, $V_{1p}$, $U_{2c}$, $U_{2p}$, and a time-sharing RV $Q$ defined on arbitrary finite sets $U_{1c}$, $U_{1p}$, $V_{1c}$, $V_{1p}$, $U_{2c}$, $U_{2p}$, and $Q$, respectively. Let
$Z_3 = (Q, U_{1c}, U_{2c}, V_3, V_{2p}, U_{2p}, U_{2q}, X_1, X_2, Y_2, Y_3, Y_4)$, and $\mathcal{P}_3$ denote the set of all joint pmfs $p(.)$ on $Z_3$ that can be factored in the form of (18) with $(I_{mp}, I_{mc}) = (u_{3p}, u_{1c})$. Let $\mathcal{R}_3(Z_3)$ be the set of all nonnegative rate pairs $(R_1, R_2)$ where $R_1 = R_{1cd} + R_{1cn} + R_{1pm} + R_{1cm}$ and $R_2 = R_{2c} + R_{2p}$, such that there exist nonnegative $(I_{2c}, I_{2p})$ which satisfy (19)–(37) with $(T_{mp}, T_{mc}) = (U_{1p}, U_{1q})$ and

$$R_{1cd} \leq I(U_{1p}, Y_2 | U_{1c}, Q) \quad (45)$$
$$R_{1cn} + R_{1pm} \leq I(U_{1q}, Y_1 | U_{1c}, Q). \quad (46)$$

**Theorem 7:** For any $p(.) \in \mathcal{P}_3$, the region $\mathcal{R}_3(Z_3)$ is an achievable rate region for the discrete memoryless CC-IFC with unlimited look ahead, i.e., $\bigcup_{Z_3 \in \mathcal{P}_3} \mathcal{R}_3(Z_3) \subseteq \mathcal{C}_N$.

**Remark 7:** For the unlimited look-ahead case, using a coding scheme based on instantaneous relaying is feasible. However, to compare this strategy with noncausal partial DF, we restrict the use of this scheme to $L = 1$. In fact, applying the strategy of Theorem 6 to the CC-IFC with unlimited look ahead without using noncausal partial DF will achieve $\mathcal{R}_2$ and in order to utilize the extra information in this case we must employ a noncausal strategy. Moreover, a strategy based on noncausal partial DF and instantaneous relaying achieves a region which encompasses the ones for other values of $L$, wherein eliminating the noncausal partial DF will result in $\mathcal{R}_2$ and deleting instantaneous relaying will result in $\mathcal{R}_3$. Thus, to compare the noncausal partial DF and instantaneous relaying strategies, we must consider $\mathcal{R}_2$ and $\mathcal{R}_3$. Therefore, to reduce complexity, we prefer to exclude the instantaneous relaying in the scheme of Theorem 7.

**Proof:** The proof of Theorem 7 is similar to that of Theorem 5, with the exception that there is no dependence on the previous block messages and the transmitters noncausally cooperate using correlated codewords. For this reason, Tx1 uses superposition coding with four codewords: $u^n_{1c}, u^n_{2p}$ for cooperative messages $(m_{1ab}, m_{1bp}),$ and $u^n_{1c}, u^n_{2p}$ for $m_{1bn}, m_{1pm}$, where all codewords related to the private messages are superimposed on the codewords related to the common messages and codewords conveying noncooperative information are superimposed on the cooperative codewords. Using joint binning against $u^n_{1p}$, Tx2 creates $u^n_{2c}, u^n_{2p}$, for its own messages, while in order to relay $m_{1cd}$, these codewords are generated conditioned on $u^n_{1c}$. Due to the noncausal cooperative scheme, simultaneous joint decoding is used instead of backward decoding at the receivers. Thus, the proof follows the same lines as in Theorem 5 and is omitted here for the sake of brevity.

**Remark 8:** As mentioned earlier, CC-IFC with unlimited look ahead generalizes the noncausal C-IFC and can reduce to this channel model when $p(y_2|x_2)$ is ideal, i.e., the cognitive link between the transmitters is noise free. To obtain an achievable rate region for this case, we use the region $\mathcal{R}_3$ of Theorem 7 and assume that the cognitive user can fully decode the message of the primary user $(m_1)$. Therefore, in $\mathcal{R}_3$, we set $V_3 = V_{1p} = \emptyset$ and $R_{1cn} = R_{1pm} = 0$, drop (38) and (39) due to the elimination of the cognitive link, and drop (28) and (30) because they correspond to the incorrect decoding of the common message from the nonintended transmitter, and derive $\mathcal{R}_{NC}$ for noncausal C-IFC. Now, we compare $\mathcal{R}_{NC}$ with the known results for noncausal C-IFC:

1) The Region in [9] ($\mathcal{R}_{MGKS}$):
- In $\mathcal{R}_{MGKS}$, the binning is done sequentially and conditionally in two steps, while we utilize the joint binning technique in $\mathcal{R}_{NC}$ with potential improvement.
- In $\mathcal{R}_{MGKS}$, the message of the primary user is split into two parts; however, the nonintended receiver decodes none of these parts.

Noting that the positions of the primary and cognitive users are switched in $\mathcal{R}_{MGKS}$, setting $R_{1cd} = 0$ and $U_{1c} = \emptyset$, redefining $U_{1p} = (X_{2a}, X_{2b}), U_{2c} = U_{1c},$ and $U_{2p} = U_{1a}$, and considering the aforementioned discussion reduces $\mathcal{R}_{NC}$ to a region which includes $\mathcal{R}_{MGKS}$.

2) The Region in [10] ($\mathcal{R}_{IJ}$):
   a) There is no rate splitting for the message of the primary user in $\mathcal{R}_{IJ}$.
   b) In $\mathcal{R}_{IJ}$, the binning is done independently in contrast to our joint binning technique in $\mathcal{R}_{NC}$.

   Thus, if we set $R_{1cd} = 0$ and $U_{1c} = \emptyset$, and redefine $U_{1p} = W, U_{2c} = U$, and $U_{2p} = V$ in $\mathcal{R}_{NC}$, our region is reduced to one which includes $\mathcal{R}_{IJ}$ as a subset.

3) Weak Interference in [6, Proposition 3.1] ($\mathcal{R}_{WVA}$): By switching the position of the primary and cognitive users in [6], we assume that the second transmitter is cognitive. Now, set $R_{1cd} = R_{2c} = L_{2c} = 0$ and $U_{1c} = U_{2c} = \emptyset$, and redefine $U_{1p} = (X_1, U)$ and $U_{2p} = V$ in $\mathcal{R}_{NC}$. Since there is no common message to be decoded at Rx1, drop (34). Applying these assignments, $\mathcal{R}_{NC}$ reduces to $\mathcal{R}_{WVA}$.

4) Strong Interference in [7, Th. 5] ($\mathcal{R}_{MYK}$): By setting $R_{1cd} = R_{2p} = L_{2c} = L_{2p} = 0$ and $U_{1p} = U_{2p} = \emptyset$, redefining $U_{1c} = X_1$ and $U_{2c} = X_2$ and dropping (26), (29), and (35) which are due to the incorrect decoding of the common message at the nonintended receivers, $\mathcal{R}_{NC}$ reduces to the capacity region of noncausal C-IFC with strong interference, also referred to as strong IFC with unidirectional cooperation, derived in [7, Th. 5].

5) The Region in [31] ($\mathcal{R}_{LMGS}$): Noting that the first transmitter is cognitive in [31], set $R_{2p} = 0$ and $U_{2p} = \emptyset$; redefine $U_{1c} = V$, $U_{2c} = U$, and $U_{1p} = X_2$ in $\mathcal{R}_{NC}$; and drop (35). Then, it can be easily shown that our region reduces to $\mathcal{R}_{LMGS}$ which achieves the capacity for a class of the cognitive Z-IFCs.

6) The Regions in [12] and [13] ($\mathcal{R}_{RTD}$): The region in [13] is the largest known achievable rate region for the noncausal C-IFC, which has some differences in the binning technique with the one in [12]. Our scheme does not include these regions. The reason is as follows: In $\mathcal{R}_{RTD}$, a part of the primary user’s message is sent only by the cognitive user based on using Marton coding [38]. In fact, this scheme is possible because the cognitive user knows the primary user’s message by a genie. However, in our proposed model, i.e., the CC-IFC with unlimited look ahead, the cognitive user must decode the message of the primary user in a noncausal manner. Therefore, the entire message must be sent by the primary user and our scheme cannot include the method of $\mathcal{R}_{RTD}$.
7) The Region in [14, Th. 14.1] for Noncausal C-IFC ($\mathcal{R}_{\text{JMC}}$): The broadcast channel with two cognitive relays is considered in [14], which is reduced to noncausal C-IFC by removing one of the relays. Our scheme and the one used to achieve $\mathcal{R}_{\text{JMC}}$ differ in the binning technique in the cognitive user. In $\mathcal{R}_{\text{JMC}}$, Marton coding is used for sending the private parts of the primary and cognitive user’s messages. However, we use GP binning for the common and private parts of the cognitive user’s message against the private message of the primary user. It appears that no subset relation can be established between $\mathcal{R}_{\text{JMC}}$ and $\mathcal{R}_{\text{NC}}$.

8) The Regions in [15] for Noncausal Cognitive Z-IFC: In [15], simple and easily computable rate regions have been derived for noncausal cognitive Z-IFC, which are also achievable for noncausal C-IFC. The region in [15, Proposition 3.1] is based on [14, Th. 14.1], which was discussed previously. By setting $R_{2p} = 0$ and $U_{2p} = \emptyset$, $\mathcal{R}_{\text{NC}}$ includes the regions in [15, Corollary 3.2] and [15, Proposition 3.2].

Remark 9: The region $\mathcal{R}_3$ of Theorem 7 achieves the capacity region of the partially cognitive IFC under strong interference conditions characterized in [8, Th. 5]. Setting $(m_{1cb}, m_{1pd}) = W_0$, $(m_{1cr}, m_{2pm}) = W_1$, and $m_{2} = W_2$ in the scheme of CC-IFC with unlimited look ahead results in the model of the partially cognitive IFC, also referred to as IFC with partial unidirectional cooperation. In order to derive the region of [8, Th. 5], set $R_{1pd} = R_{1pm} = R_{2p} = I_{2c} = L_{2c} = 0$ and $U_{1p} = V_{1p} = U_{2p} = \emptyset$; rename $R_{1cd} = R_0, R_{1cm} = R_{1}, R_{2c} = R_{2c}, U_{2c} = U, V_{2c} = X_{1},$ and $U_{2c} = X_{2}$; and drop (26), (28), and (35) in $\mathcal{R}_3$. Note that the events corresponding to these bounds cause no error in this case.

V. CAPACITY RESULTS FOR TWO SPECIAL CASES OF THE CLASSICAL CC-IFC

In this section, we investigate the classical CC-IFC (CC-IFC-WD with $L = 0$) with joint pmf $p^*$, given by (1) with $L = 0$. We find the capacity regions for the classes of degraded and semidegenerate classical CC-IFC under strong interference conditions, where we use the achievable rate region in Theorem 5 for the achievability of these regions and the outer bound of Theorem 4 for the converse parts.

A. Degraded Classical CC-IFC

We define degraded classical CC-IFC as a classical CC-IFC (CC-IFC-WD with $L = 0$) where the degradedness condition for the Tx1–Rx1 pair with the cognitive user as a relay holds for every $p^*$. More precisely

$$p(y_1|x_1, x_2, y_2) = p(y_1|x_2, y_2)$$

i.e., $X_1 \rightarrow (X_2, Y_2) \rightarrow Y_3$ forms a Markov chain. Next, we assume that the strong interference conditions (9) and (14) at Rx1 and Rx2 hold for every distribution $p^*$, where under these conditions the interfering signals are strong enough to decode both messages at both receivers.

Theorem 8: The capacity region of the degraded classical CC-IFC with the joint pmf $p^*$, satisfying conditions (9) and (14), is given by

$$\mathcal{C}_d^* = \bigcup_{p(t)p(x_1)p(y_2)} \left\{ (R_1, R_2) : R_1 \geq 0, R_2 \geq 0 \\
R_1 \leq (X_1; Y_2|X_2, T) \\
R_2 \leq I(X_2; Y_2|X_1, T) \\
R_1 + R_2 \leq \min\{I(X_1, X_2; Y_3), I(X_1, X_2; Y_4)\} \right\},$$

(50)

Remark 10: The message of the cognitive user ($m_2$) can be decoded at Rx1 under condition (9). Hence, the bound in (48) and the first bound in (50) give the capacity region of the degraded relay channel in (47)[17] with a private message $m_2$ from the relay to the receiver. Note that, due to the degradedness condition, the cognitive user is able to decode the message of the primary user ($m_1$). Moreover, $m_1$ can be decoded at Rx2 under condition (14). Therefore, we have an MAC with common information at Rx2, where $R_1$ is the common rate, $R_2$ is the private rate for the second transmitter, and the private rate for the first transmitter is zero. The bound in (49) and the second bound of (50) give the capacity region for this MAC [39].

Proof:

Achievability: For this part, we use the region $\mathcal{R}_3$ in Theorem 5 (or Corollary 1) and ignore the time-sharing RV $Q$. Let, $T = U_{1p} = V_{1p} = U_{2p} = \emptyset$ and $R_{2p} = R_{2pm} = R_{1pd} = 0$, which negates the private parts of both messages, making the messages common to both receivers. Furthermore, assume that the cognitive user (Tx2) fully decodes the message of the primary user ($m_1$). Consequently, it is necessary to set $U_{2c} = 0$ and $V_{2c} = \emptyset$. In order to omit the GP coding, we set $L_{2c} = L_{2p} = 0$ as well. Note that (26), (28)–(30), and (35) can be dropped, because these bounds correspond to the decoding of the common message from the unintended transmitter. Re-defining $T_c = T, U_{2c} = X_2, U_{2c} = X_1$, and $U_{2c} = X_1$ completes the proof for the achievability.

Converse: To prove the converse part, we evaluate $\mathcal{R}_{\text{sep}}$ of Theorem 4 for $L = 0$ (classical CC-IFC). Considering (128)

$$I_i = Y_{2i+L-1}^1|_{L=0} = \emptyset.$$ (51)

Moreover, in this case, Definition 1 provides

$$X_{2i} = f_{2i}(M_2, y_{2i-1}^iL) = f_{2i}(M_2, y_{2i}^i) = f_{2i}(V_i),$$ (52)

Therefore, the pmf in Theorem 4 reduces to the one in Theorem 8. Based on (15)

$$R_1 \leq I(X_1; Y_2|V, T) + I(X_1; Y_3|X_2, Y_2, T)$$

(a) $I(X_1; Y_2|V, T, X_2) + I(X_1; Y_3|X_2, Y_2, T)$

(b) $\leq H(Y_2|T, X_2) - H(Y_2|X_1, V, T, X_2)$

$+ I(X_1; Y_3|X_2, Y_2, T)$

(c) $I(X_1; Y_2|X_2, T) + I(X_1; Y_3|X_2, Y_2, T)$ (53)

where (a) is obtained using (52), (b) is due to the fact that conditioning does not increase the entropy, and (c) follows from the joint pmf $p^*$, given by (1) with $L = 0$. Subsequently, applying
condition (47) to (53) results in (48). Similarly, we utilize the first bound in (16) to obtain (49) as follows:

$$R_2 \leq I(V; Y_4|X_1, T)$$

(a) $$I(V; X_2; Y_4|X_1, T)$$

$$= H(Y_4|X_1, T) - H(Y_4|X_1, T, V, X_2)$$

(b) $$I(X_2; Y_4|X_1, T)$$

(54)

where for (a) we use (52) and (b) is obtained from the joint pmf $p^\ast$. In a similar manner, we derive the first bound in (50) by using the first bound of (17)

$$R_1 + R_2 \leq I(X_1, V; T; Y_3)$$

(a) $$I(X_1, V, T, X_2; Y_3)$$

$$= H(Y_3) - H(Y_3|X_1, V, T, X_2)$$

(b) $$I(X_1, X_2; Y_3)$$

(55)

where (a) and (b) are obtained with the same reasons as that used in (54). Finally, similar to (55), the second bound in (50) can be easily obtained from the second bound of (17). This completes the converse proof.

If we consider the following condition

$$I(X_1, X_2; Y_3) \leq I(X_1, X_2; Y_4)$$

(56)

instead of (14), the capacity region is given by the following corollary.

**Corollary 2**: The capacity region of the degraded classical CC-IFC with the joint pmf $p^\ast$, satisfying conditions (9) and (56), is given by

$$C_{\ast} = \bigcup_{p(t)p(x_1|t)p(x_2|t)} \left\{ (R_1, R_2) : R_1 \geq 0, R_2 \geq 0 \right\}$$

$$R_1 \leq I(X_1; Y_2|X_2, T)$$

$$R_2 \leq I(X_2; Y_4|X_1, T)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y_3)$$

(57)

(58)

(59)

**Remark 11**: If we assume that the cognitive link between the transmitters is ideal, then the cognitive user can decode the message of the primary user without any rate constraint and the bound in (57) will be dropped. In this case, by setting $T = \emptyset$, $C_{\ast}$ coincides with the capacity region of the strong IFC with unidirectional cooperation (or noncausal C-IFC), satisfying (9) and (56), which has been characterized in [7, Th. 5], [40].

**Remark 12 (Comparison of Two Sets of Conditions)**: We can write (56) as

$$I(X_1; Y_3) + \left[ I(X_2; Y_3|X_1) - I(X_2; Y_4|X_1) \right] \leq I(X_1; Y_4).$$

Considering (9), it can be seen that $I_{\text{diff}} \geq 0$. Hence, the conditions of Corollary 2 imply those of Theorem 8. Therefore, the strong interference conditions of Theorem 8 are weaker compared to the conditions obtained in [7] and [40].

### B. Semideterministic Classical CC-IFC

Here, we consider classical CC-IFC (CC-IFC-WD with $L = 0$) with the deterministic component for the channel output of the cognitive transmitter, i.e., the received signal at the cognitive user (Tx2) is a deterministic function of the primary user’s input signal

$$Y_2 = h_2(X_1).$$

(60)

Assume that for every distribution $p^\ast$, this semideterministic classical CC-IFC satisfies (9), (14) and the following additional condition:

$$I(X_1; Y_3|Y_2, X_2) \leq I(X_1; Y_4|Y_2, X_2).$$

(61)

**Theorem 9**: The capacity region of the semideterministic classical CC-IFC, defined by (60) with the joint pmf $p^\ast$, satisfying conditions (9), (14), and (61), is given by

$$C_{\ast} = \bigcup_{p(t)p(x_1|t)p(x_2|t)} \left\{ (R_1, R_2) : R_1 \geq 0, R_2 \geq 0 \right\}$$

$$R_1 \leq H(Y_2|X_2, T) + I(X_1; Y_3|Y_2, X_2, T)$$

$$R_2 \leq I(X_2; Y_4|X_1, T)$$

$$R_1 + R_2 \leq \min\{I(X_1, X_2; Y_3), I(X_1, X_2; Y_4)\}$$

(62)

(63)

(64)

**Remark 13**: Similar to Remark 10, the aforementioned channel model can be seen as a semideterministic relay channel of (60)[33] with a private message $m_2$ from the relay to the receiver and an MAC with common information at Rx2 [39].

**Proof**:

**Achievability**: Similar to Theorem 8, we specialize the region $R_1$ in Theorem 5 with $Q = \emptyset$. In order to cancel the private parts of the messages, let $T_p = U_{1p} = V_{1p} = U_{2p} = \emptyset$ and $R_{2p} = R_{1p} = R_{1pd} = 0$. Moreover, ignore GP coding by setting $L_{2p} = L_{2pd} = 0$. Also, redefine $U_{2c} = X_2, V_{1c} = X_1, U_{3c} = U, T_c = T$. Thus, $R_1$ reduces to

$$R_1 \leq I(U; Y_2|X_2, T) + I(X_1; Y_3|U, X_2, T)$$

$$R_1 \leq I(U; Y_2|X_2, T) + I(X_1; Y_4|U, X_2, T)$$

$$R_2 \leq I(X_2; Y_4|X_1, T)$$

$$R_2 \leq I(X_2; Y_4|X_1, T)$$

$$R_1 + R_2 \leq \min\{I(X_1, X_2; Y_3), I(X_1, X_2; Y_4)\}$$

$$R_1 + R_2 \leq \min\{I(X_1, X_2; Y_3|U, T), I(X_1, X_2; Y_4|U, T)\}$$

(65)

(66)

(67)

(68)

(69)

(70)

where $P_1$ in (18) becomes

$$p(t)p(x_1, u|t)p(x_2|t).$$

(71)

Due to condition (9), the bound in (67) is redundant. Now, in the aforementioned region, let $U = Y_2$, which is feasible because the primary user knows $Y_2 = h_2(X_1)$. Then, due to condition (61), the bound in (66) becomes redundant and (65), (68), and (69) reduce to (62)–(64), respectively. Moreover, pmf in (71) becomes


\begin{align}
p(t)p(x_1|t)p(x_2|t).
\end{align}

Hence, due to the conditional independence of \(X_2\) and \(X_1\) given \(T\), the following equations are obtained:

\begin{align}
I(X_2;Y_3|X_1,T) &= I(X_2;Y_3|X_1,Y_2,T) = I(X_2;Y_3|Y_2,T) \quad (73) \\
I(X_2;Y_4|X_1,T) &= I(X_2;Y_4|X_1,Y_2,T) = I(X_2;Y_4|Y_2,T). \quad (74)
\end{align}

Combining (73) and (9), the first bound in (70) becomes redundant. In a similar manner, (74) and (61) make the second bound in (70) redundant. This completes the proof for achievability.

Converse: For this part, we use the bounds derived in the converse proof of Theorem 8. Bounds in (63) and (64) are obtained directly from (54) and (55). For (62), we use (53) to obtain

\begin{align}
R_1 &\leq I(X_1;Y_2|X_2,T) + I(X_1;Y_3|X_2,Y_2,T) \\
&= H(Y_2|X_2,T) + I(X_1;Y_3|X_2,Y_2,T) \quad (75)
\end{align}

where (60) has been used for (75).

VI. GAUSSIAN CC-IFC-WD

In this section, we consider Gaussian CC-IFC-WD and extend the achievable rate regions \(R_1(Z_1), R_2(Z_2),\) and \(R_3(Z_3)\) derived for the discrete memoryless classical CC-IFC (\(L = 0\)), the discrete memoryless CC-IFC without delay (\(L = 1\)), and the discrete memoryless CC-IFC with unlimited look ahead, respectively, to the Gaussian case. Moreover, we present some numerical examples in order to investigate the effects of the delay and the rate gain of the cognitive link in this channel. Thus, we compare the strategies which are used for achieving the aforementioned rate regions.

A. Channel Model for the Gaussian CC-IFC-WD

Gaussian CC-IFC-WD, as depicted in Fig. 5, at time \(i = 1, \ldots, n\) can be modeled mathematically as

\begin{align}
\bar{Y}_{2,i} &= h_{21}X_{1,i} + Z_{2,i} \\
\bar{Y}_{3,i} &= h_{31}X_{1,i} + h_{32}X_{2,i} + Z_{3,i} \\
\bar{Y}_{4,i} &= h_{41}X_{1,i} + h_{42}X_{2,i} + Z_{4,i}
\end{align}

(76)

where \(h_{21}, h_{31}, h_{32}, h_{41},\) and \(h_{42}\) are known channel gains. Additionally, \(X_1\) and \(X_2\) are input signals with average power constraints

\begin{align}
\frac{1}{n} \sum_{i=1}^{n} (\bar{x}_{u,i})^2 \leq P_u \quad (77)
\end{align}

for \(u \in \{1, 2\}\). Also, \(Z_{2,i}, Z_{3,i},\) and \(Z_{4,i}\) are i.i.d and independent zero mean Gaussian noise components with powers \(N_2, N_3,\) and \(N_4\), respectively. Note that, at the cognitive user, we have a set of encoding functions \(f_{2,i} = f_{2,i}(m_2,y_{2,i}^{2,1+L})\) for \(i = 1, \ldots, n\) and \(m_2 \in M_2\).

B. Achievable Rate Regions for the Gaussian CC-IFC-WD

To simplify notation, we define

\begin{align}
\theta(x) &= \frac{1}{2} \log(1+x). \quad (78)
\end{align}

First, we consider the Gaussian classical CC-IFC (\(L = 0\)). For certain \(\{0 \leq \beta_s \leq 1, s \in \{1, 2, 3, 4\}\}\), \(\{0 \leq \gamma_t \leq 1, t \in \{1, 2, 3\}\}\) with \(\beta_1 + \beta_2 + \beta_3 + \beta_4 \leq 1\) and \(\gamma_2 + \gamma_3 \leq 1\), we define \(I'_t, i = 1, \ldots, 21\) as (79)–(99), shown at the bottom of page 2828, where

\begin{align}
\alpha_1 &\leq h_{42}\gamma_2 P_2 + D + (h_{41}\sqrt{\beta_3 P_1} + h_{42}\sqrt{\gamma_3 P_2})^2 \\
\alpha_2 &\geq D + (h_{41}\sqrt{\beta_3 P_1} + h_{42}\sqrt{\gamma_3 P_2})^2 \\
A &\geq \gamma_1 P_2 + \alpha_1^2 h_{41}^2 \beta_1 \beta_3 P_1 \\
B &\geq \alpha_2^2 h_{41}^2 \beta_1 \beta_3 P_1 + (\gamma_2 + h_{42}^2 \gamma_2) P_2 \\
C &\geq h_{42}^2 \gamma_2 \beta_1 \beta_3 P_2 P_2 - (1 - \alpha_2 h_3) h_{42}^2 \gamma_2 \beta_1 \beta_3 P_2 P_2 \\
D &\geq h_{41}^2 (\beta_1 + \beta_2 + \beta_3 + \beta_4) P_1 + h_{42}^2 \gamma_2 P_1 \\
F &\geq (h_{41}^2 \beta_3 \alpha_1 P_1 + 2h_{42}^2 \gamma_2 P_1) h_{42}^2 \gamma_2 \alpha_1 P_1 P_2 \\
&+ C \theta_2 (\alpha_2^2 h_{42}^2 \gamma_2 P_2 + \alpha_1^2 h_{41}^2 \beta_1 \beta_3 P_1 + 2h_{42}^2 \gamma_2 P_2) \\
&+ (A \gamma_2 + \gamma_1^2 P_2) h_{42}^2 \gamma_2 \beta_1 \beta_3 P_2 P_2. \quad (106)
\end{align}

Now, replacing each term in (19)–(39) with the corresponding term from (79)–(99) (replacing \(I_i\) with \(I'_i\) for \(i = 1, \ldots, 21\)), we obtain the Gaussian counterpart of \(R_1\), namely \(R_1^*\).

Theorem 10: For the Gaussian classical CC-IFC (CC-IFC-WD with \(L = 0\)), defined in Section VI-A, the convex closure of the region

\begin{align}
\bigcup_{\{\beta_s, \gamma_t: s \in \{1, 2, 3, 4\}\}, \gamma_2 + \gamma_3 \leq 1 \atop \gamma_2 \leq 1} R_1^*
\end{align}

where \(r \in \{1, 2, 3, 4\}, s \in \{1, 2\}\) and \(t \in \{1, 2, 3\}\), is an achievable rate region.

Proof: The achievable rate region \(R_1\) in Theorem 5 (or Corollary 1) can be extended to the discrete-time Gaussian memoryless case with continuous alphabets by standard arguments [41]. Hence, it is sufficient to evaluate (19)–(39) with an appropriate choice of input distribution to reach (79)–(99). We constrain all the inputs to be Gaussian and set the time-sharing RV \(Q = \emptyset\).

For certain \(\{0 \leq \beta_s \leq 1, s \in \{1, 2, 3, 4\}\}, \{0 \leq \beta_s' \leq 1, s \in \{1, 2\}\}, \) and \(\{0 \leq \gamma_t \leq 1, t \in \{1, 2, 3\}\}\), consider the following mapping \((MA1)\) for the codebook generated in Theorem 5 with respect to the pmf (18), which contains the Gaussian version of the generalized block Markov superposition coding, rate splitting, and GP coding:

\begin{align}
T_c &\sim N(0, \beta_4 P_1) \quad (107) \\
T_p = T_p' + T_c \\
\text{where } T_p' &\sim N(0, \beta_3 P) \quad (108)
\end{align}
\[ U_{1c} = U_{1c}' + T_c \quad \text{where} \quad U_{1c}' \sim \mathcal{N}(0, \beta_2 P_1) \]  
(109)

\[ U_{1p} = U_{1p}'+U_{1c}'+T_p'+T_c \quad \text{where} \quad U_{1p}' \sim \mathcal{N}(0, \beta_2 P_1) \]  
(110)

\[ V_{1c} = V_{1c}' + T_c \quad \text{where} \quad V_{1c}' \sim \mathcal{N}(0, \beta_2 P_1) \]  
(111)

\[ V_{1p} = V_{1p}'+V_{1c}'+T_p'+T_c \quad \text{where} \quad V_{1p}' \sim \mathcal{N}(0, \beta_2 P_1) \]  
(112)

\[ X_1 = V_{1p} + V_{1c} + U_{1p} + U_{1c} + T_p + T_c \]  
(113)

\[ U_{2c} = U_{2c}' + \alpha_1 S_1 \quad \text{where} \quad U_{2c}' \sim \mathcal{N}(0, \gamma_2 P_2) \]  
(114)

\[ U_{2p} = U_{2p}' + \alpha_2 S_2 \quad \text{where} \quad U_{2p}' \sim \mathcal{N}(0, \gamma_2 P_2) \]  
(115)

\[ X_2 = U_{2p} + U_{2c} + \sqrt{\frac{\beta_3 P_2}{\beta_4 P_1}} T_c \]  
(116)

where \( \alpha_1 \) and \( \alpha_2 \) are defined in (100) and (101), respectively, and

\[ S_1 = h_{41} T_p \]  
(117)

\[ S_2 = h_{41} T_p + h_{42} U_{2c}' \]  
(118)

Parameters \( \beta_4 \) and \( \beta_3 \) determine the amounts of \( P_1 \) which are dedicated for constructing the basis of cooperation for sending common and private messages, respectively. Parameter \( \beta_2 \) specifies the amount of \( P_1 \) which is allocated for relaying through the cognitive user to perform GP decoding. The remaining parts of \( P_1 \), distinguished with parameters \( \beta_4 \) and \( \beta_3 \), are sent directly to Rx1. Parameters \( \gamma_1 \) and \( \gamma_2 \) determine the amounts of \( P_2 \) which are dedicated for sending the common message, the private message, and relaying, respectively. To execute GP coding, parameters \( \alpha_1 \) and \( \alpha_2 \) are utilized. In fact, optimal values for \( \alpha_1 \), \( \alpha_2 \), \( S_1 \), and \( S_2 \) can be determined by optimizing the rate region of these parameters. However, this method is cumbersome, so we use the modified version of Costa’s dirty paper coding (DPC) results [42].

Applying the power constraints in (77) to MAP yields

\[ \beta_4 + \beta_1 + \beta_2 + \beta_2 + \beta_3 + \beta_4 \leq 1 \]

\[ \gamma_1 + \gamma_2 + \beta_3 \leq 1 \]

Using the aforementioned mapping (MAP3) with the channel model in (76), the remainder of the proof is straightforward.

Next, we investigate the Gaussian CC-IFC without delay \( (L = 1) \). First, we modify \( P^*_1, i = 1, \ldots, 21 \) in (79)–(99), by replacing \( h_{u1} \) with \( h_{u1}^* \), \( h_{u2} \) with \( h_{u2}^* \), and \( N_u \) with \( N_u' \) for \( u \in \{3, 4\} \), and refer to them as \( P_{1i}^* \) for \( i = 1, \ldots, 21 \).

Consider the channel model in Fig. 5 and (76) with \( L = 1 \), i.e., \( X_2 = f_2(m_2, Y_2^2) \). In order to obtain the Gaussian counterpart of \( \mathcal{R}_2 \), namely \( \mathcal{R}_2^* \), we replace each term \( \{I_i, i = 1, \ldots, 21\} \) in (19)–(39) with its corresponding term \( \{I_i^*, i = 1, \ldots, 21\} \), for certain \( \{0 \leq \beta_i \leq 1, r \in \{1, 2, 3, 4\}\} \), \( \{0 \leq \beta_{i}' \leq 1, s \in \{1, 2\}\} \), and

\[ h_{u1} = h_{u1} + h_{u2} h_{u2} h_{u2} \]  
(120)

\[ h_{u2}' = h(1 - \beta) h_{u2} \]  
(121)

\[ N_{u}' = N_u + h^2\beta_2^2 h_{u2}^2 N_2 \]  
(122)

for \( u \in \{3, 4\} \), where \( 0 \leq \beta \leq 1 \), \( h \) is a normalizing parameter and the following inequalities hold:

\[ \begin{align*}
\frac{\beta_1}{h^2} + \frac{\beta_2}{h^2} + \frac{\beta_2}{h^2} + \frac{\beta_3}{h^2} + \beta_4 & \leq 1 \\
\frac{\beta_1}{h^2} + \beta_1 + \beta_2 + \beta_2 + \beta_3 + \beta_4 & \leq 1 \\
(h_{21} \beta \sqrt{\frac{\beta_1 P_1 + (1 - \beta) \sqrt{\frac{\beta_2 P_2}{\beta_1 P_1}}}{h^2}} + \beta_2 N_2 \quad + (1 - \beta) \gamma_2 P_2) & \quad \leq 1
\end{align*} \]

\[ (119) \]

Theorem 11: For the Gaussian CC-IFC without delay (CC-IFC-WD with \( L = 1 \)), in Section VI-A, the convex closure of the region

\[ \bigcup_{\beta_1, \beta_2, \beta_3, \beta_4} \mathcal{R}_2^* \quad \text{where} \quad \begin{cases} 
\beta_1 + \beta_2 + \beta_2 + \beta_3 + \beta_4 & \leq 1 \\
\beta_1 + \beta_2 + \beta_2 + \beta_3 + \beta_4 & \leq 1 \\
0 \leq \beta_i & \leq 1, \quad r \in \{1, 2, 3, 4\}
\end{cases} \]

\[ \{0 \leq \beta_{i}' \leq 1, \quad s \in \{1, 2\}, \quad \beta_i & \leq 1, \quad i \in \{1, 2\} \}, \quad \text{is an achievable rate region.} \]

Proof: The proof of Theorem 11 is similar to that of Theorem 10. Considering Theorem 6, \( V_2 \) is generated according to

\[ \prod_{i=1}^n q(v_{2i}, v_{2p}, u_{2s}, t_{2s}, t_{2s}, u_{2s}, q_i) \quad \text{and} \quad x_{2i} = f_2(m_{2i}, Y_{2p}^2) \]  
(123)

for Gaussian inputs and \( Q = \emptyset \), appropriate mapping (MAP3) for the codebook generated in Theorem 6, with respect to the pmf \( P_2 \) defined in (44), consists of (107)–(115), (117) and (118), and

\[ V_2 = U_{2p} + U_{2c} + \sqrt{\frac{\beta_3 P_2}{\beta_4 P_1}} T_c \]  
(120)

\[ X_2 = h(\beta Y_2 + (1 - \beta) V_2) \]  
(121)

where \( 0 \leq \beta \leq 1 \) and \( h \) is a normalizing parameter. In fact, the cognitive user sends a linear function of its received symbol.
\[ I_1^* = \theta \left( \frac{\alpha_1^2 h_{31}^2 \beta_3 P_1}{\gamma_2 P_2} \right) \]

\[ I_2^* = \theta \left( \frac{\alpha_1^2 h_{31}^2 \beta_3 P_1}{(\gamma_2 + \alpha_1^2 h_{32} \gamma_2) P_2} \right) \]

\[ I_3^* = I_1^* + \theta \left( \frac{\alpha_1^2 h_{31}^2 \beta_3 (1 - \alpha_1 h_{42})^2 \gamma_1 P_1 P_2}{A \gamma_2 P_2} \right) + \theta \left( \frac{\alpha_1^2 (\alpha_1 h_{31}^2 \beta_3 P_1 + h_{42} \gamma_1 P_2)^2}{A \gamma_2 P_2 + C \alpha_2^2} \right) \]

\[ I_4^* = \theta \left( \frac{h_{32}^2 \beta_3 P_1}{h_{32}^2 \gamma_2 P_2 + N_3} \right) \]

\[ I_5^* = I_1^* + \theta \left( \frac{h_{32}^2 \beta_3 P_1 + h_{32}^2 \gamma_1 P_2}{h_{32}^2 \gamma_2 P_2 + N_3} \right) \]

\[ I_6^* = I_1^* + \theta \left( P_1 \frac{A h_{31}^2 (\beta_1 + \beta_2 + \beta_3) + \alpha_1^2 h_{31}^2 \beta_3 P_1 + h_{32}^2 \gamma_1 P_2}{A \gamma_2 P_2 - h_{31}^2 \beta_3 P_1} - 2 \alpha_1 h_{31} h_{42} \beta_4 \gamma_1 P_2 \right) \]

\[ I_7^* = I_1^* + \theta \left( \frac{h_{32}^2 P_1}{h_{32}^2 \gamma_2 P_2 + N_3} \right) \]

\[ I_8^* = I_1^* + \theta \left( \frac{h_{32}^2 (\beta_1 + \beta_2 + \beta_3) P_1 + h_{32}^2 \gamma_1 P_2}{h_{32}^2 \gamma_2 P_2 + N_3} \right) \]

\[ I_9^* = I_1^* + \theta \left( \frac{h_{32}^2 (\beta_1 + \beta_2 + \beta_3) P_1 + h_{32}^2 \gamma_1 P_2}{h_{32}^2 \gamma_2 P_2 + N_3} \right) \]

\[ I_{10}^* = I_3^* + \theta \left( \frac{C_{31} h_{32}^2 \beta_3 P_1 + 2 \alpha_1 h_{42} \gamma_2 (\alpha_1^2 + \gamma_2) P_2}{B (N_4 + h_{31}^2 (\beta_1 + \beta_2) P_1)(A \gamma_2 P_2 + C \alpha_2^2) + C(1 - \alpha_1 h_{42})^2 \gamma_2 P_2} \right) \]

\[ I_{11}^* = I_3^* + \theta \left( \frac{C_{31} h_{32}^2 \beta_3 P_1 + 2 \alpha_1 h_{42} \gamma_2 (\alpha_1^2 + \gamma_2) P_2}{B (N_4 + h_{31}^2 (\beta_1 + \beta_2) P_1)(A \gamma_2 P_2 + C \alpha_2^2) + C(1 - \alpha_1 h_{42})^2 \gamma_2 P_2} \right) \]

\[ I_{12}^* = I_3^* + \theta \left( \frac{C_{31} h_{32}^2 \beta_3 P_1 + 2 \alpha_1 h_{42} \gamma_2 (\alpha_1^2 + \gamma_2) P_2}{B (N_4 + h_{31}^2 (\beta_1 + \beta_2) P_1)(A \gamma_2 P_2 + C \alpha_2^2) + C(1 - \alpha_1 h_{42})^2 \gamma_2 P_2} \right) \]

\[ I_{13}^* = I_3^* + \theta \left( \frac{C_{31} h_{32}^2 \beta_3 P_1 + 2 \alpha_1 h_{42} \gamma_2 (\alpha_1^2 + \gamma_2) P_2}{B (N_4 + h_{31}^2 (\beta_1 + \beta_2) P_1)(A \gamma_2 P_2 + C \alpha_2^2) + C(1 - \alpha_1 h_{42})^2 \gamma_2 P_2} \right) \]

\[ I_{14}^* = I_3^* + \theta \left( \frac{C_{31} h_{32}^2 \beta_3 P_1 + 2 \alpha_1 h_{42} \gamma_2 (\alpha_1^2 + \gamma_2) P_2}{B (N_4 + h_{31}^2 (\beta_1 + \beta_2) P_1)(A \gamma_2 P_2 + C \alpha_2^2) + C(1 - \alpha_1 h_{42})^2 \gamma_2 P_2} \right) \]

\[ I_{15}^* = \theta \left( \frac{C_{31} h_{32}^2 \beta_3 P_1 + 2 \alpha_1 h_{42} \gamma_2 (\alpha_1^2 + \gamma_2) P_2}{B (N_4 + h_{31}^2 (\beta_1 + \beta_2) P_1)(A \gamma_2 P_2 + C \alpha_2^2) + C(1 - \alpha_1 h_{42})^2 \gamma_2 P_2} \right) \]

\[ I_{16}^* = I_{13}^* + \theta \left( \frac{A h_{42} \gamma_2 (\alpha_1^2 + \gamma_2) P_2}{B (N_4 + h_{31}^2 (\beta_1 + \beta_2) P_1)(A \gamma_2 P_2 + C \alpha_2^2) + C(1 - \alpha_1 h_{42})^2 \gamma_2 P_2} \right) \]

\[ I_{17}^* = I_{14}^* + \theta \left( \frac{A h_{42} \gamma_2 (\alpha_1^2 + \gamma_2) P_2}{B (N_4 + h_{31}^2 (\beta_1 + \beta_2) P_1)(A \gamma_2 P_2 + C \alpha_2^2) + C(1 - \alpha_1 h_{42})^2 \gamma_2 P_2} \right) \]

\[ I_{18}^* = \theta \left( \frac{h_{32}^2 \beta_3 P_1}{h_{32}^2 \gamma_2 P_2 + N_3} \right) \]

\[ I_{19}^* = \theta \left( \frac{h_{32}^2 (\beta_1 + \beta_2) P_1}{h_{32}^2 \gamma_2 P_2 + N_3} \right) \]

\[ I_{20}^* = \theta \left( \frac{h_{32}^2 (\beta_1 + \beta_2) P_1}{h_{32}^2 \gamma_2 P_2 + N_3} \right) \]

\[ I_{21}^* = \theta \left( \frac{h_{32}^2 (\beta_1 + \beta_2) P_1}{h_{32}^2 \gamma_2 P_2 + N_3} \right) \]
and the codeword $V_2$, where $h_\beta$ determines the amount of $P_2$ which is dedicated for instantaneous relaying by the cognitive user. Also, (119) is obtained by applying the power constraints in (77) to $\text{MAP}_2$.

Evaluating $\{I_i, i = 1, \ldots, 21\}$ in (19)–(39), using the aforementioned mapping (MAP$_2$) and (76), results in $\{I^*_i, i = 1, \ldots, 21\}$. Considering Theorem 6, the remainder of the proof is straightforward.

Finally, we derive the rate region for the Gaussian CC-IFC with unlimited look ahead. Let $I^*_i = 1, \ldots, 19 and$

$$I^*_{20} = \epsilon \left( \frac{h_{21}^2 (\beta_2^* + \beta_1^*) P_1 + N_2}{\beta_2^* (\beta_3 + \beta_4^*) P_1 + N_2} \right).$$

$$I^*_{21} = \epsilon \left( \frac{h_{21}^2 (\beta_3 + \beta_4^*) P_1}{\beta_3 (\beta_3 + \beta_4^*) P_1 + N_2} \right).$$

Now, replacing each term in (19)–(37), (45), and (46) with its corresponding term from (79)–(97), (122), and (123), i.e., replacing $I_i$ with $I^*_i$ for $i = 1, \ldots, 21$, yields the Gaussian counterpart of $\mathcal{R}_3$, to which we refer as $\mathcal{R}^*_3$.

**Theorem 12:** For the Gaussian CC-IFC with unlimited look ahead, defined in Section VI-A, the convex closure of the region

$$\bigcup_{\{\beta_0, \beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*\} \in [0, 1]} \mathcal{R}^*_3$$

where $r \in \{1, 3, 4\}$ and $t \in \{1, 2, 3\}$, is an achievable rate region.

**Proof:** The proof follows the same lines as that of Theorem 10, except that according to Theorem 7 there is no dependence on the previous block messages. Therefore, it is possible to set $T_c = U_{1c}$ and $T_p = U_{1p}$, or equivalently $\beta_2^* = 0$ and $\beta_2 = 0$ in $\text{MAP}_1$ to obtain $\text{MAP}_3$.

**C. Numerical Examples for the Gaussian CC-IFC-WD**

Here, we provide some numerical examples of the rate regions $\mathcal{R}^*_1, \mathcal{R}^*_2, \mathcal{R}^*_3$, and $\mathcal{R}^*_4$ in Theorem 10, $\mathcal{R}^*_2$ in Theorem 11, and $\mathcal{R}^*_3$ in Theorem 12. Comparing the strategies used to achieve the aforementioned rate regions, we investigate the effects of the delay, cooperation, and interference cancellation in this channel. First, we consider the rate gain of the cognitive link for different strategies.

Fig. 6 compares the regions $\mathcal{R}^*_1, \mathcal{R}^*_2$, and $\mathcal{R}^*_3$ with the HK region in [4], where the overheard information is neglected, for $P_1 = P_2 = 6, h_{31} = h_{32} = 1, h_{41} = \sqrt{0.55},$ and $N_2 = N_3 = N_4 = 1.$ Moreover, an outer bound on the capacity region of CC-IFC-WD is provided by intersecting the capacity region of the Gaussian MIMO broadcast channel (MIMO-BC) [43] with the rate of the TX2-Rx2 interference-free channel, i.e., $R_2 \leq \theta \left( \frac{P_2}{N_2} \right).$ These regions are shown in Fig. 7 for $P_2 = 6$ and $P_2 = 1.5$. Due to the cooperative strategies, $\mathcal{R}^*_1, \mathcal{R}^*_2, \mathcal{R}^*_3$ outperform the HK region. Especially when the cognitive link is sufficiently strong, i.e., $h_{21} = 4$, $\mathcal{R}^*_2$ and $\mathcal{R}^*_3$ achieve rates close to the outer bound for a small $P_2$, because the cognitive user can decode and cooperate more effectively and can allocate more power for simultaneous cooperation. Due to instantaneous relaying and noncausal DF schemes, larger regions are obtained for $L = 1$ and unlimited look ahead ($L = \infty$) cases than for $L = 0$.

To compare the noncausal DF ($\mathcal{R}^*_3$) and instantaneous relaying ($\mathcal{R}^*_2$) based on Figs. 6 and 7, one must consider the condition of the cognitive link. For a strong cognitive link ($h_{21} = 4$), the performance of the noncausal DF scheme is better (especially when the cognitive user sends at higher rates), and allowing sufficient time for the cognitive user to decode increases...
the rates that can be achieved. However, when \( h_{21} = 1 \), instantaneous relaying outperforms DF for small \( R_2 \). In fact, when poor conditions exist for the cognitive link, instantaneous relaying is the only scheme that can outperform the HK scheme for the primary user \( (R_1) \) when the cognitive user sends at lower rates. We remark that, since an instantaneous relaying scheme is feasible for every \( L \geq 1 \), the convex hull of the regions \( \mathcal{R}_2^0 \) and \( \mathcal{R}_3^0 \) is achievable for CC-IFC with unlimited look ahead \( (L = \infty) \) using a coding scheme based on a combination of instantaneous relaying with noncausal DF strategies.

Fig. 8 portrays the impacts of partial DF relaying \( (\gamma_3) \), instantaneous relaying \( (\beta) \), and interference cancellation by DPC \( (\alpha_1 \) and \( \alpha_2) \) for \( P_1 = P_2 = 6, h_{31} = h_{42} = 1, h_{32} = h_{41} = \sqrt{0.55} \), and \( N_2 = N_3 = N_4 = 1 \). Considering \( \mathcal{R}_1^0 \) \( (L = 0) \) and \( \mathcal{R}_3^0 \) \( (L = \infty) \), we see that when \( R_2 \) is large, setting \( \gamma_3 = 0 \) (no DF relaying) performs better. This more efficient performance means that in this case interference cancellation by DPC is a better strategy. However, when the cognitive user sends at lower rates and can allocate more power for relaying, DPC provides less improvement.

It is worth noting that the region related to \( \alpha_1 = \alpha_2 = 0 \) can also be obtained by the general scheme if the rate region is optimized for these parameters instead of using (100) and (101). A similar argument can be made for \( \mathcal{R}_3^0 \) \( (L = 1) \). However, the performance improvement in the latter case is due mostly to the instantaneous relaying, especially when \( R_2 \) is small.

Fig. 9 compares the regions \( \mathcal{R}_1^0, \mathcal{R}_2^0, \) and \( \mathcal{R}_3^0 \), with the HK region for \( P_1 = P_2 = 6, h_{31} = h_{42} = 1, N_2 = N_3 = N_4 = 1 \), and different values of \( h_{32} \) and \( h_{41} \), where results similar to those depicted in Fig. 6 can be concluded at the strong interference \( (h_{32} = h_{41} = \sqrt{1.5}) \) and the mixed interference \( (h_{32} = \sqrt{0.55}, h_{41} = \sqrt{1.5}) \) regimes.

In Fig. 10, in order to investigate the effect of the noise in the channel between the transmitters (the cognitive link), we compare the region \( \mathcal{R}_3^0 \) for the unlimited look-ahead case \( (L = \infty) \) with the noncausal scheme in [9] for \( P_1 = P_2 = 6, h_{31} = h_{42} = 1, h_{32} = \sqrt{2}, h_{41} = \sqrt{0.3}, N_2 = N_3 = N_4 = 1 \), and different values of \( N_2 \). We see that, when poor conditions exist for the cognitive link, i.e., \( N_2 = 100 \), one cannot gain very much using the strategy of \( \mathcal{R}_3^0 \) in comparison with the HK scheme. As \( N_2 \) decreases, the performance approaches the rates achieved in the noncausal scheme of [9] as well as the outer bound. For \( N_2 = 0 \), our rate region outperforms that in [9] in agreement with the discussion in part 1 of Remark 8.

VII. CONCLUSION

We introduced the CC-IFC-WD and investigated its capacity region. We derived a general outer bound on the capacity region for arbitrary value of \( L \) and specialized it to the strong interference case. We tightened the outer bound under strong interference conditions. We also obtained achievable rate regions for three special cases: 1) Classical CC-IFC; 2) CC-IFC without delay; and 3) CC-IFC with unlimited look ahead. Coding schemes were based on the generalized block Markov superposition coding, rate splitting, and GP binning. Moreover, instantaneous relaying and noncausal partial DF were employed in the second and third cases, respectively. Furthermore, using the derived inner and outer bounds, we characterized the capacity regions for the classes of the degraded and semideterministic classical CC-IFC under strong interference conditions. We showed that these channel models can be seen as a combination of the degraded or semideterministic relay channel with private message from the relay to the receiver and the MAC with common information.

Also, we investigated Gaussian CC-IFC-WD by extending our achievable rate regions to the Gaussian case and providing some numerical examples in order to examine the rate gain of the cognitive link. We compared different strategies which we have used in the coding schemes and showed that instantaneous
relaying and noncausal DF improve the rate region noticeably and achieve rates close to the outer bound for a strong cognitive link, especially when the rate of the cognitive user is small. In addition, comparing the partial (causal or noncausal) DF, instantaneous relaying and DPC (GP binning) strategies, we attempted to identify the cases wherein each strategy is dominant. The results showed that when the cognitive user sends at higher rates, interference cancellation by DPC is a better strategy. However, when the cognitive user sends at lower rates and can dedicate more power to cooperating with the primary user, DPC provides less improvement.

APPENDIX A
PROOF OF THE OUTER BOUNDS

Proof of Theorem 2: Consider a \((2^{nR_1}, 2^{nR_2}, n, P_{c}^{(f)})\) code with an average error probability \(P_e^{(f)} \to 0\), which implies that \(P_{c1}^{(f)} \to 0\) and \(P_{c2}^{(f)} \to 0\). Applying Fano’s inequality [41] results in
Now, using Fano’s inequality, we derive the bounds in Theorem 2. First, we provide some useful lemmas which we need in the proof of this theorem.

Lemma 1: \((M_1, Y_u^{(i-1)}) \rightarrow (X_{1,i}, V_i, U_i) \rightarrow (Y_{u,i}, Y_{2,i})\) forms a Markov chain, where \(u \in \{3, 4\}\).

Proof: Noting (1), consider \(p(m_1, m_2, x_{1,i}, y_u, y_{2,i + 1}, y_{2,i})\) which can be written as
\[
\begin{align*}
\text{(a)} & = p(m_1, x_{1,i}, y_u, v_i, u_i, y_{2,i}) \\
& = p(m_1, y_u^{i-1}, x_{1,i}, v_i, u_i) p(y_u, y_{2,i} | m_1, y_u^{i-1}, x_{1,i}, v_i, u_i) \\
& \quad \times p(y_{2,i} | m_1, y_u^{i-1}, x_{1,i}, v_i, u_i, y_{2,i}) \\
\text{(b)} & = p(y_{2,i} | m_1, y_u^{i-1}, x_{1,i}, v_i, u_i, y_{2,i}) p(y_{2,i} | x_{1,i}, v_i, u_i) \\
& \quad \times p(y_{2,i} | m_1, y_u^{i-1}, x_{1,i}, v_i, u_i, y_{2,i}) \\
& \quad \times p(y_{2,i} | m_1, y_u^{i-1}, x_{1,i}, v_i, u_i, y_{2,i}) \\
\end{align*}
\]
where we use (127) and (128) for (a), (b)–(d) follow from the joint pmf (1) and the fact that \(X_{2,i}\) is a deterministic function of \(V_i, U_i\), and \(Y_{2,i}\).

Lemma 2: For \(u \in \{3, 4\}\), \(X_{2,i} \rightarrow (X_{1,i}, V_i, U_i) \rightarrow (Y_{u,i}, Y_{2,i})\) forms a Markov chain.

Proof: Note the joint pmf in (1) and consider \(p(m_2, x_{1,i}, x_{2,i}, y_{2,i + 1}, y_{2,i}, y_{h,i})\) which can be written as
\[
\begin{align*}
\text{(a)} & = p(x_{1,i}, x_{2,i}, v_i, u_i, y_{h,i}, y_{2,i}) \\
& = p(x_{2,i}, x_{1,i}, v_i, u_i) p(y_{h,i}, y_{2,i} | x_{1,i}, x_{2,i}, v_i, u_i) \\
& \quad \times p(y_{2,i} | x_{1,i}, x_{2,i}, v_i, u_i, y_{2,i}) \\
\text{(b)} & = p(x_{2,i}, x_{1,i}, v_i, u_i) p(y_{2,i} | x_{1,i}, v_i, u_i) \\
& \quad \times p(y_{2,i} | x_{1,i}, x_{2,i}, v_i, u_i, y_{2,i}) \\
\text{(c)} & = p(y_{2,i} | x_{1,i}, v_i, u_i) p(y_{h,i} | x_{1,i}, v_i, u_i, y_{2,i}) \\
& \quad \times p(y_{h,i} | x_{1,i}, x_{2,i}, v_i, u_i, y_{2,i}) \\
\end{align*}
\]
where we use (127) and (128) for (a), (b) is due to the joint pmf given by (1), and (c) follows from the fact that \(X_{2,i}\) is a deterministic function of \(V_i, U_i\), and \(Y_{2,i}\).

Now, using Fano’s inequality, we derive the bounds in Theorem 2. For the first bound
\[
\begin{align*}
nR_1 & = H(M_1) \overset{(a)}{=} H(M_1 | M_2) \\
& = I(M_1; Y_3^n | M_2) + H(M_1 | Y_3^n, M_2) \overset{(b)}{=} I(M_1; Y_3^n | M_2) + n\delta_{3n}
\end{align*}
\]
where (a) follows since $M_1$ and $M_2$ are independent and (b) holds due to (124) and the fact that conditioning does not increase the entropy. Hence, we obtain

$$nR_1 - n\delta_{1n} \leq I(M_1; Y_3^n | M_2) \overset{(a)}{=} I(M_1; Y_3^n, Y_2^n | M_2)$$

$$\overset{(b)}{=} I(M_1; Y_2^n | M_2) + I(M_1; Y_3^n | M_2, Y_2^n, X_2^n)$$

$$\overset{(c)}{=} \sum_{i=1}^{n} \left\{ I(M_1; Y_3|Y_2^{i-1}, M_2) + I(M_1, X_{1,i}; Y_3|Y_2^{i-1}, M_2, Y_2^n, X_2^n) \right\}$$

$$\overset{(d)}{=} \sum_{i=1}^{n} \left\{ I(M_1, X_{1,i}; Y_3|Y_2^{i-1}, M_2) + I(M_1, X_{1,i}; Y_3, X_{1,i}|Y_2^{i-1}, M_2, Y_2^n, X_2^n) \right\}$$

$$\overset{(e)}{=} \sum_{i=1}^{n} \left\{ H(Y_2|V_i, T_i) - H(Y_2|V_i, T_i, M_1, X_{1,i}) \right\}$$

$$+ \sum_{i=1}^{n} \left\{ H(Y_3|X_{1,i}, Y_2^{i-1}, T_i) - H(Y_3|X_{1,i}, Y_3^{i-1}, V_i, T_i, Y_2^n, X_2^n) \right\}$$

$$\overset{(f)}{=} \sum_{i=1}^{n} \left\{ I(X_1^{i-1}; Y_2^{i-1}|V_i, T_i) + I(X_1^{i-1}; Y_3|X_{1,i}, Y_2^{i-1}, T_i) \right\}$$

$$\overset{(g)}{=} n \left\{ I(X_1^{i-1}; Y_2^{i-1}, V_i^{i-1}, X_{1,i}) + I(X_1^{i-1}; Y_3|X_{1,i}, Y_2^{i-1}, V_i^{i-1}, X_{1,i}) \right\}$$

$$= n \left\{ I(X_1^{i-1}; Y_2^{i-1}, V_i^{i-1}, X_{1,i}) + I(X_1^{i-1}; Y_3|X_{1,i}, Y_2^{i-1}, V_i^{i-1}, X_{1,i}) \right\}$$

$$\overset{(h)}{=} n \left\{ I(X_1^{i-1}; Y_2^{i-1}, V_i^{i-1}, X_{1,i}) + I(X_1^{i-1}; Y_3|X_{1,i}, Y_2^{i-1}, V_i^{i-1}, X_{1,i}) \right\}$$

$$\overset{(i)}{=} n \left\{ I(X_1^{i-1}; Y_2^{i-1}, V_i^{i-1}, X_{1,i}) + I(X_1^{i-1}; Y_3|X_{1,i}, Y_2^{i-1}, V_i^{i-1}, X_{1,i}) \right\}$$

$$\overset{(j)}{=} n \left\{ I(X_1^{i-1}; Y_2^{i-1}, V_i^{i-1}, X_{1,i}) + I(X_1^{i-1}; Y_3|X_{1,i}, Y_2^{i-1}, V_i^{i-1}, X_{1,i}) \right\}$$

$$\overset{(k)}{=} n \left\{ I(X_1^{i-1}; Y_2^{i-1}, V_i^{i-1}, X_{1,i}) + I(X_1^{i-1}; Y_3|X_{1,i}, Y_2^{i-1}, V_i^{i-1}, X_{1,i}) \right\}$$

$$\overset{(l)}{=} \sum_{i=1}^{n} I(M_2^{i-1}; Y_2^{i-1}, M_1, Y_3^{i-1}, X_{1,i})$$

$$\overset{(m)}{=} \sum_{i=1}^{n} I(V_i; Y_2^{i-1}, M_1, Y_3^{i-1}, X_{1,i})$$

$$\overset{(n)}{=} nI(V_Q; Y_3|X_{1,i}, Q_T, Q) = nI(V; Y_3|X_{1,i}, Q_T, Q)$$

where (a) and (d) are obtained from the nonnegativity of mutual information, (b) is based on the chain rule, (c) is obtained since $X_{1,i}$ is a deterministic function of $M_1$, (e) holds since the channel is memoryless with the joint pmf (1), (f) follows from (126) and (127) and the fact that conditioning does not increase the entropy, (f) follows from the fact that the channel is memoryless with the joint pmf (1), and (g) is obtained by using a standard time-sharing argument, where $Q$ is a time-sharing RV, independent of all other RVs and uniformly distributed over $\{1, 2, \ldots, n\}$, and we define $X_1 Q = X_1, X_2 Q = X_2, Y_2 Q = Y_2, Y_3 Q = Y_3, V_2 = V$, and $(Q_T, Q) = T$.}

Now, as a result of applying Fano’s inequality in (125) and the independence of the messages, we can bound $R_2$ as

$$nR_2 - n\delta_{2n} \leq I(M_2; Y_3^n | M_1)$$

$$\overset{(a)}{=} I(M_2; Y_3^n, Y_2^n | M_1)$$

$$\overset{(b)}{=} \sum_{i=1}^{n} I(M_2; Y_3^n, Y_2^n | Y_2^{i-1}, Y_2^{i-1}, M_1)$$

$$\overset{(c)}{=} \sum_{i=1}^{n} I(M_2; Y_3^n, Y_2^n | Y_2^{i-1}, Y_2^{i-1}, M_1, X_{1,i})$$

$$\overset{(d)}{=} \sum_{i=1}^{n} \{ I(M_2; Y_2^{i-1}, Y_3^n, Y_2^n | Y_2^{i-1}, Y_2^{i-1}, M_1, X_{1,i}) + I(M_2; Y_2^{i-1}, Y_3^n, Y_2^n | Y_2^{i-1}, Y_2^{i-1}, M_1, X_{1,i}) \}$$

where (a) and (d) are obtained from the nonnegativity of mutual information, (b) is based on the chain rule, (c) is obtained since $X_{1,i}$ is a deterministic function of $M_1$, (e) holds since the channel is memoryless with the joint pmf (1), (f) follows from (126) and (127), and Lemma 1 for $u = 4$, and for (g) we use the time-sharing argument of (130)-(g) and $Y_Q = Y_2$.

Now, let $Y_Q$ be any RV with the same marginal distribution of $Y_3$, i.e., $p(y_Q^n | x_Q^n, x_2^n) = p(y_3^n, y_Q^n | x_Q^n, x_2^n)$, but with an arbitrary joint distribution $p(y_Q^n, y_3^n | x_Q^n, x_2^n)$. Subsequently, the second bound on $R_2$ can be derived as

$$nR_2 - n\delta_{2n} \leq I(M_2; Y_3^n | M_1)$$

$$\overset{(a)}{=} I(M_2; Y_3^n, Y_2^n | M_1, X_{1,i})$$

$$\overset{(b)}{=} \sum_{i=1}^{n} I(M_2; Y_3^n, Y_2^n | M_1, X_{1,i})$$

$$\overset{(c)}{=} \sum_{i=1}^{n} \left\{ I(U_i; V_i, T_i) + I(U_i, V_i, T_i, Y_4^n | Y_3^n, X_{1,i}) \right\}$$

where (a) is based on the facts that $X_Q$ is a deterministic function of $M_1$ and mutual information is nonnegative, (b) is obtained from the chain rule and the memoryless property of the channel with the joint pmf (1), (c) is true due to the memoryless property of the channel, the definition of $Y_Q$, and the fact that conditioning does not increase the entropy, (d) follows from (126)-(128) and Lemma 2 for $u = 3$, and for (e) we use the defined time-sharing argument and $Y_3 Q = Y_3, U_Q = U$.

Next, we bound $R_1 + R_2$ as

$$nR_1 + R_2 - n(\delta_{1n} + \delta_{2n}) \leq I(M_1; Y_3^n) + I(M_2; Y_3^n | M_1)$$

$$\overset{(a)}{=} I(M_1; Y_3^n) + I(M_2; Y_3^n | M_1)$$

$$\overset{(b)}{=} H(Y_3^n) - H(Y_3^n | M_1) + H(Y_3^n | M_1)$$

$$\overset{(c)}{=} H(Y_3^n | M_1) + H(Y_3^n | M_1)$$

$$\overset{(d)}{=} H(Y_3^n | M_1) + H(Y_3^n | M_1)$$

where (a) and (d) are obtained from the nonnegativity of mutual information, (b) is based on the chain rule, (c) is obtained since $X_{1,i}$ is a deterministic function of $M_1$, (e) holds since the channel is memoryless with the joint pmf (1), (f) follows from (126) and (127), and Lemma 1 for $u = 4$, and for (g) we use the time-sharing argument of (130)-(g) and $Y_Q = Y_2$.}
where (a) follows from the definition of $Y_3'$ and the fact that conditioning does not increase the entropy, (b) is due to the memoryless property of the channel and the definition of $Y_3'$, (c) follows from the steps (b)–(d) in (132) and the fact that the channel is memoryless, and (d) follows from Lemma 2 for $u = 3$ and the fact that mutual information is nonnegative. This completes the proof.

Proof of Theorem 4: The bounds in (15) and (16) and the first bound in (17) follow from (10)–(12). Therefore, we need to prove the second sum-rate bound in (17). Consider a code with the properties of that in the proof of Theorem 2. First, we state the following lemma.

**Lemma 3:** If (14) holds, then

$$I(X_1^T; Y_{3'}^n) \leq I(X_1^T; Y_4^n).$$

**Proof:** The proof relies on the result in [44, Proposition 1] and follows the same lines as in [7, Lemma 5] and [45, Lemma].

Before proceeding to bound the sum rate, we need to state the following inequalities:

$$I(M_1; Y_3^n) \leq I(M_1, X_1^T; Y_3^n)$$

$$= I(X_1^T; Y_3^n) + I(M_1; Y_3^n | X_1^T)$$

$$\leq I(X_1^T; Y_3^n) + H(M_1 | Y_3^n | X_1^T)$$

$$= I(X_1^T; Y_3^n) + H(M_1 | X_1^T, Y_3^n)$$

$$\leq I(X_1^T; Y_3^n) + H(M_1 | X_1^T, Y_3^n)$$

$$\leq I(X_1^T; Y_3^n) + H(M_1 | X_1^T, Y_3^n)$$

where (a) and (c) follow from the deterministic relation between $X_1^T$ and $M_1$, (b) is due to the chain rule, and (d) holds due to the nonnegativity of the entropy.

$$I(M_2; Y_4^n | M_1) \leq I(M_2, X_2^T; Y_4^n | M_1, X_1^T)$$

$$= H(Y_4^n | M_1, X_1^T) - H(Y_4^n | M_1, X_1^T, M_2, X_2^T)$$

$$\leq H(Y_4^n | X_1^T) - H(Y_4^n | M_1, X_1^T, X_2^T) = I(X_1^T; Y_4^n | X_1^T)$$

where (a) is based on the facts that conditioning does not increase the entropy and $(M_1, M_2) \rightarrow (X_1, X_2) \rightarrow Y_4$ forms a Markov chain.

Now, the second bound in (17) can be obtained as

$$n(R_1 + R_2) - n(\delta_{1n} + \delta_{2n}) \leq I(M_1; Y_3^n) + I(M_2; Y_4^n | M_1)$$

and

$$\leq I(X_1^T; Y_3^n) + I(X_2^T; Y_4^n | X_1^T)$$

$$\leq \sum_{i=1}^n I(X_1, i; Y_3) + I(U_i, V_i, T_i; Y_4; Y_3, Y_4')$$

$$\leq nI(X_1, U_i, V_i, T_i; Y_3) + nI(U_i, V_i, T_i; Y_4; Y_3, Y_4')$$

for (a) follows from (135) and (136), (b) from condition (134), and (c) from Lemma 2 for $u = 4$. This completes the proof.

**APPENDIX B**

**ANALYSIS OF THE PROBABILITY OF ERROR FOR THEOREM 5**

Due to the symmetry of the random codebook generation, the probability of error is independent of the specific messages. Hence, without loss of generality, we assume that the message tuples $m_{1,b} = (m_{1b}, m_{2b}, m_{3b}, m_{4b}) = (1, 1, 1, 1)$ and $m_{2,b} = (m_{2b}, m_{3b}, m_{4b}) = (1, 1, 1)$ are encoded and transmitted in each block $b$. Recall that, in the first block, the cooperative information is defined as $(m_{1b}, m_{1b}, m_{1b}) = (1, 1, 1)$. In the last block, a previously known message $(m_{1b}, m_{1b}, m_{1b}, B) = (1, 1, 1)$. Furthermore, backward decoding is utilized at Rx1 and Rx2. Consider the events (138)–(145), shown at the bottom of the next page.

Moreover, we define $\mathcal{F}_{b-1}$ to be the event in which no errors have occurred up to block $b$. Note that, in Rx1 and Rx2, up to block $b$ means blocks $b + 1, \ldots, B$, due to backward decoding. We can write the overall probability of error as

$$P_e = Pr \left[ \bigcup_{b=1}^{B} \bigcup_{m^n \in [1, 2^{nT}-1]} \bigcup_{n' \in [1, 2^{nL/2}]} E_{\text{enc}2, b, m^n, n'} \right]$$

$$\cup \bigcup_{b=1}^{B-1} \left( E_{\text{dec}2,b, 1, 1, 1} \cup \bigcup_{(i', j', n') \neq (1, 1)} E_{\text{dec}2, b, i', j', n'} \right)$$

$$\bigcup_{b=1}^{B-1} \left( E_{\text{dec}3,b, 1, 1, 1, 1} \cup \bigcup_{(i, j, k, l, m, n) \neq (1, 1, 1, 1)} E_{\text{dec}3, i, j, k, l, m, n} \right)$$

$$\bigcup_{b=1}^{B-1} \left( E_{\text{dec}4,b, 1, 1, 1, 1} \cup \bigcup_{(i', j', k', l', m', n') \neq (1, 1, 1, 1)} E_{\text{dec}4, i', j', k', l', m', n'} \right)$$

$$\leq \sum_{b=1}^{B} Pr (E_{\text{enc}2, b} | \mathcal{F}_{b-1}) + \sum_{b=1}^{B-1} (E_{\text{dec}2,b} | E_{\text{enc}2, b} | \mathcal{F}_{b-1})$$

$$+ \sum_{b=1}^{B} (E_{\text{dec}3,b} | E_{\text{enc}2, b} | \mathcal{F}_{b-1}) + \sum_{b=1}^{B} (E_{\text{dec}4,b} | E_{\text{enc}2, b} | \mathcal{F}_{b-1})$$

where we define

$$E_{\text{enc}2, b} = \bigcup_{m^n \in [1, 2^{nT}-1]} E_{\text{enc}2, b, m^n, n'}$$

for $u = 4$. This completes the proof.
and $E^c$ denotes the complement of the event $E$.

Hence, assuming that no errors have occurred up to block $b$, bounding the probability of encoding or decoding error in block $b$ for each user is sufficient for bounding the overall probability of error.

First, we bound the probability of encoding error for the cognitive user (Tx2) at the beginning of block $b$, defined as $P_{e,enc2b}$,

$$P_{e,enc2b} = Pr\left( E_{enc2b} \right) = Pr\left( \bigcup_{(m',n') \neq (1,1)} E_{enc2b,m',n'} \right).$$

Using mutual covering lemma [32, 46], $P_{e,enc2b} \rightarrow 0$ if $n \rightarrow \infty$ and (19)–(21) hold.

Next, we bound the probability of decoding error for the cognitive user (Tx2) at the end of block $b$, defined as $P_{e,dec2b}$,

$$P_{e,dec2b} = Pr\left( E_{dec2b} E_{enc2b}^c \right) = Pr\left( \bigcup_{(m',n') \neq (1,1)} E_{dec2b,m',n'} \bigg| E_{enc2b}^c \right) \leq Pr\left( E_{dec2b,m',n'} \bigg| E_{enc2b}^c \right) + \sum_{m' \neq 1, n' \neq 1} Pr\left( E_{dec2b,m',n'} \bigg| E_{enc2b}^c \right),$$

and (152)

Due to the asymptotic equipartition property [41] and considering the codebook generation of Theorem 5, $Pr\left( E_{dec2b,m',n'} \bigg| E_{enc2b}^c \right) \rightarrow 0$ as $n \rightarrow \infty$. Utilizing [41, Th. 15.2.3] for the other terms in (152) and (153), we have

$$P_{e,dec2b} \leq 2 \epsilon + 2^{nR_{1b} - n} (U_1 U_2 Y_2 U_2 T_2 P_2 T_2 Q) - \epsilon$$

$$+ 2^{nR_{1b} - n} (U_1 Y_3 U_2 U_2 T_2 P_2 T_2 Q) - \epsilon$$

and $P_{e,dec2b} \rightarrow 0$ as $n \rightarrow \infty$. Note that the second term in the right-hand side of (154) imposes no constraint on $R_{1b}$, because the events of the second terms in the right-hand side of (152) and (153) share the same pmf.

In a similar manner, the probability of the decoding error for Rx1 at the end of block $b$ (defined as $P_{e,dec3b}$) can be bounded as

$$P_{e,dec3b} \leq Pr\left( E_{dec3b} E_{enc3b}^c \right) \leq Pr\left( E_{dec3b,m',n'} \bigg| E_{enc3b}^c \right) + \sum_{m' \neq 1, n' \neq 1} Pr\left( E_{dec3b,m',n'} \bigg| E_{enc3b}^c \right),$$

(157)

For the second term in the right-hand side of (157), there are sixty cases that cause an error. However, some of these cases share the same pmf and so there are only nine distinct cases.
Now, using the packing lemma [32] (or [41, Th. 15.2.3]), we bound the probability of these events (conditioning on $E_{c}^{(m_{2})}$ suppressed). Note that in the following, when the value of an index is unspecified, e.g., $i$, that index can take any value from its set, e.g., $i = 1$ or $i \neq 1$. First, consider

$$
Pr(E_{dec3,h_{1},1,j\neq 1,k\neq 1,l_{2c},b_{1},m_{1}}) \leq 2^{-n(I(V_{1}p_{1}V_{1}p_{1}T_{2}Y_{2}U_{2c},|V_{1}T_{c}|) - 6e)}
$$

(160)

**ACKNOWLEDGMENT**

The authors would like to thank all the anonymous reviewers and the associate editor for their constructive comments and suggestions on the paper. The authors also wish to thank M. H. Yassaei and R. Bayat for their helpful comments.

REFERENCES

[1] A. B. Carleial, “Interference channels,” *IEEE Trans. Inf. Theory*, vol. 24, no. 1, pp. 60–70, Jan. 1978.

[2] D. N. C. 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.

[3] S. Gelfand and M. Pinsker, “Coding for channels with random parameters,” *Prob. Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[4] T. S. Han and K. Kobayashi, “A new achievable rate region for the interference channel,” *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 49–60, Jan. 1981.

[5] A. Jovicic and P. Viswanath, “Cognitive radio: An information-theoretic perspective,” *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 3945–3958, Sep. 2009.

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

[7] I. Maric, R. D. Yates, and G. Kramer, “Capacity of interference channels with partial transmitter cooperation,” *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3536–3548, Oct. 2007.

[8] I. Maric, A. Goldsmith, G. Kramer, and S. Shitz, “On the capacity of interference channels with a partially-cognitive transmitter,” in *Proc. IEEE Int. Symp. Inf. Theory*, 2007, pp. 2156–2160.

[9] I. Maric, A. Goldsmith, G. Kramer, and S. S. Shitz, “On the capacity of interference channels with one cooperating transmitter,” *Eur. Trans. Telecommun.*, vol. 19, pp. 405–420, Apr. 2008.

[10] J. Jiang and Y. Xin, “On the achievable rate regions for interference channels with degraded message sets,” *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4707–4712, Oct. 2008.

[11] M. Nikulin, V. Aref, and S. Aref, “Partial cognitive relay channel,” in *Proc. IEEE Inf. Theory Workshop*, 2009, pp. 341–345.

[12] S. Rini, D. Tutinetti, and N. Devroye, “State of the cognitive interference channel: A new unified inner bound, and capacity to within 1.87 bits,” in *Proc. 2010 Int. Zurich Semin. Commun.*, Mar. 2010, pp. 1–4.
MIRMOHSENI et al.: CAUSAL AND NONCAUSAL GENERALIZED FEEDBACK AT THE COGNITIVE TRANSMITTER

[13] S. Rini, D. Tuninetti, and N. Devroye, “New inner and outer bounds for the discrete memoryless noiseless interference channel and some capacity results,” IEEE Trans. Inf. Theory, vol. 57, no. 7, pp. 4087–4109, Oct. 2011.

[14] J. Jiang, I. Marin, A. Goldsmith, and S. Cui, “Achievable rate regions for broadcast channels with cognitive relays,” in Proc. IEEE Inf. Theory Workshop, Taormina, Italy, Oct. 2009, pp. 500–504.

[15] J. Jiang, I. Marin, A. Goldsmith, S. S. Shitz, and S. Cui, On the capacity of a class of cognitive Z-interference channels Jul. 2010 [Online]. Available: http://arxiv.org/abs/1007.1811v1

[16] Y. Cao and B. Chen, “Interference channel with one cognitive transmitter,” presented at the presented at the IEEE Asilomar Conf. Signals, Syst. Comput., Pacific Grove, CA, Oct. 2008.

[17] D. Tuninetti, “On interference channel with generalized feedback (IFC-GF),” in Proc. IEEE Int. Symp. Inf. Theory, 2007, pp. 2861–2865.

[18] Y. Cao and B. Chen, “An achievable rate region for interference channels with conferencing,” in Proc. IEEE Int. Symp. Inf. Theory, 2007, pp. 1251–1255.

[19] Y. Yang and D. Tuninetti, “A new achievable region for interference channel with generalized feedback,” in Proc. Annu. Conf. Inf. Sci. Syst., Mar. 2008, pp. 803–808.

[20] Y. Cao and B. Chen, “Capacity outer bounds for the cognitive Z channel,” in Proc. IEEE Conf. Global Commun., Dec. 2008, pp. 1–6.

[21] F. M. J. Willems, “The discrete memoryless multiple access channel with partially cooperating encoders,” IEEE Trans. Inf. Theory, vol. IT-29, no. 3, pp. 441–445, May 1983.

[22] A. E. Gamal, N. Hassanpour, and J. Mammen, “Relay networks with delays,” IEEE Trans. Inf. Theory, vol. 53, no. 10, pp. 3413–3431, Oct. 2007.

[23] J. Mitola, Cognitive Radio Architecture. New York: Wiley, 1991.

[24] N. Hassanpour, “Relay without delay,” Ph.D. dissertation, Dept. Electr. Eng., Stanford University, Stanford, CA, 2006.

[25] A. Salimi, M. Mirmohseni, and M. R. Aref, “A new capacity upper bound for “relay-with-delay” channel,” in Proc. IEEE Int. Symp. Inf. Theory, 2009, pp. 26–30.

[26] H. Sato, “An outer bound to the capacity region of broadcast channels,” IEEE Trans. Inf. Theory, vol. IT-24, no. 3, pp. 374–377, May 1978.

[27] D. Tuninetti, “An outer bound region for interference channels with generalized forwarding,” in Proc. IEEE Inf. Theory Appl. Workshop, San Diego, CA, Feb. 2010, pp. 1–5.

[28] N. Liu, I. Marin, A. Goldsmith, and S. S. Shitz, “Bounds and capacity results for the cognitive z-interference channel,” in Proc. IEEE Int. Symp. Inf. Theory, 2009, pp. 2422–2426.

[29] A. E. Gamal and Y.-H. Kim, Lecture Notes on Network Information Theory 2010 [Online]. Available: http://arxiv.org/abs/1001.3404

[30] A. E. Gamal and M. Aref, “The capacity of the semi-deterministic relay channel,” IEEE Trans. Inf. Theory, vol. IT-28, no. 3, pp. 536–536, May 1982.

[31] G. Kramer, M. Gastpar, and P. Gupta, “Cooperative strategies and capacity theorems for relay networks,” IEEE Trans. Inf. Theory, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[32] C. E. Shannon, “Channels with side information at the transmitter,” IBM J. Res. Develop., vol. 2, pp. 289–293, 1958.

[33] M. Mirmohseni, B. Akhbari, and M. R. Aref, “Compress-and-forward strategy for the relay channel with causal state information,” in Proc. IEEE Inf. Theory Workshop, Taormina, Italy, Oct. 2009, pp. 426–430.

[34] R. Dabora, I. Marin, and A. Goldsmith, “Relay strategies for interference forwarding,” in Proc. IEEE Inf. Theory Workshop, Porto, Portugal, May 2008, pp. 46–50.

[35] K. Marton, “A coding theorem for the discrete memoryless broadcast channel,” IEEE Trans. Inf. Theory, vol. IT-25, no. 3, pp. 306–311, May 1979.

[36] D. Spelmon and J. K. Wolf, “A coding theorem for multiple access channels with correlated sources,” Bell Syst. Tech. J., vol. 52, pp. 1037–1076, 1973.

[37] I. Marin, R. Yates, and G. Kramer, “The strong interference channel with unidirectional cooperation,” presented at the presented at the UCLSD Workshop Inf. Theory Appl., San Diego, CA, Feb. 2006.

[38] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd ed. New York: Wiley, 2006, Wiley Series in Telecommunications.

[39] M. H. M. Costa, “Writing on dirty paper,” IEEE Trans. Inf. Theory, vol. IT-29, no. 3, pp. 439–441, May 1983.

[40] H. Weingarten, Y. Steinberg, and S. Shamai, “The capacity region of the Gaussian multiple-input multiplescanner output broadcast channel,” IEEE Trans. Inf. Theory, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[41] J. Körner and K. Marton, Comparison of Two Noisy Channels, I. Csiz’ar at and P. Elias, Eds. Amsterdam, The Netherlands: North Holland, 1977.

[42] M. H. M. Costa and A. E. Gamal, “The capacity region of the discrete memoryless interference channel with strong interference,” IEEE Trans. Inf. Theory, vol. 33, no. 5, pp. 710–711, Sep. 1987.

[43] A. E. Gamal and E. C. van der Meulen, “A proof of Marton’s coding theorem for the discrete memoryless broadcast channel,” IEEE Trans. Inf. Theory, vol. IT-27, no. 1, pp. 120–122, Jan. 1981.

Mahtab Mirmohseni (S’06) received the B.Sc. and M.Sc. degrees in Electrical Engineering from Sharif University of Technology (SUT), Tehran, Iran, in 2004 and 2007, respectively. She received the Ph.D. degree in Communication Systems-Electrical Engineering from Sharif University of Technology, Tehran, Iran in January 2012. Her research interests include areas of communication theory and multiuser information theory with emphasis on cognitive networks.

Bahareh Akhbari (S’09–M’12) received the B.Sc. degree in 2003, the M.Sc. degree in 2005 and the Ph.D. degree in November 2011, all in Electrical Engineering from Sharif University of Technology (SUT), Tehran, Iran. She was also a visiting PhD student at the University of Minnesota from October 2010 to September 2011. Her research interests include network information theory and communication theory.

Mohammad Reza Aref was born in city of Yazd in Iran in 1951. He received his B.Sc. in 1975 from University of Tehran, his M.Sc. and Ph.D. in 1976 and 1980, respectively, from Stanford University, all in Electrical Engineering. He returned to Iran in 1980 and was actively engaged in academic and political affairs. He was a Faculty member of Isfahan University of Technology from 1982 to 1995. He has been a Professor of Electrical Engineering at Sharif University of Technology since 1995 and has published more than 260 technical papers in communication and information theory and cryptography in international journals and conferences proceedings. His current research interests include areas of communication theory, information theory and cryptography with special emphasis on network information theory and security for multiuser wireless communications. At the same time, during his academic activities, he has been involved in different political positions. First Vice President of I. R. Iran, Vice President of I. R. Iran and Head of Management and Planning Organization, Minister of ICT of I. R. Iran and Chancellor of University of Tehran, are the most recent ones.