IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Channels with Cooperation Links that May Be Absent

Wasim Huleihel and Yossef Steinberg

CCIT Report #894
January 2016

Electronics
Computers
Communications

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

CCIT Report #894 January 2016

Channels with Cooperation Links that May Be Absent

Wasim Huleihel and Yossef Steinberg

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel

***Abstract*—It is well known that cooperation between users in a communication network can lead to significant performance gains. A common assumption in past works is that all the users are aware of the resources available for cooperation, and know exactly to what extent these resources can be used. In this work a family of models is suggested where the cooperation links may or may not be present. Coding schemes are devised that exploit the cooperation links if they are present, and can still operate (although at reduced rates) if cooperation is not possible.**

***Index Terms*—Broadcast channel, conferencing decoders, cooperation, cribbing, multiple access channel.**

I. Introduction

Communication techniques that employ cooperation between users in a network have been an extensive area of research in recent years. The interest in such schemes stems from the potential increase in the network performance. The employment of cooperative schemes require the use of system resources - bandwidth, time slots, energy, etc - that should be allocated for the cooperation to take place. Due to the dynamic nature of modern, wireless ad-hoc communication systems, the availability of these resources is not guaranteed a priori, and the coding schemes are required to work also in the absence of the cooperation links, although possibly achieving lower communication rates.

In this work we study channels with cooperation links that may or may not be present. We focus on two cases - the physically degraded broadcast channel (BC) with conferencing decoders, and the multiple access channel (MAC) with cribbing encoders. The BC with conferencing decoders was first studied by Dabora and Servetto [2], [3], and independently by Liang and Veeravalli [6], [7], who studied also the more general setting of relay-broadcast channels (RBC). In the model of Dabora and Servetto, a two-users BC is considered, where the decoders can exchange information via noiseless communication links of limited capacities $C_{1,2}$ and $C_{2,1}$. When the broadcast channel is physically degraded, information sent from the weaker (degraded) user to the stronger is redundant, and only the capacity of the link from the stronger user to the weaker (say $C_{1,2}$) increases the communication rates. For this case, Dabora and Servetto characterized the capacity region. Their result coincides with the results of Liang and Veeravalli when the relay link of [6] is replaced with a constant rate bit pipe.

The MAC with cribbing encoders was introduced by Willems and Van Der Meulen in [9]. Here there is no dedicated communication link that can be used explicitly for cooperation. Instead, one of the encoders can crib, or listen, to the channel input of the other user. This model describes a situation in which users in a cellular system are located physically close to each other, enabling part of them to listen to the transmission of the others with high reliability - i.e., the channel between the transmitters that are located in close vicinity is almost noiseless. Willems and Van Der Meulen considered in [9] all consistent scenarios of cribbing (strictly causal, causal, non-causal, and symmetric or asymmetric), and characterized the capacity region of these models.

In the next sections, we propose and study extensions of the two models described above, when the cooperation links ($C_{1,2}$ of the physically degraded BC, and the cribbing link of the MAC) may or may not be present. For the MAC models, we first propose achievable rate regions which are based on the combination of superposition coding and block-Markov coding. Here, we consider the unreliable strictly causal, causal, and non-causal cribbing. Then, we propose a general outer bound, which is tight for some interesting special cases where some constraints on the rates of the users are added. For the physically degraded BC, the results are conclusive.

It should be noted that multi-user communication systems with uncertainty in part of the network links have been extensively studied in the literature - see, e.g., [8] and [5], and references therein. The models suggested here, of the BC and MAC with uncertainty in the cooperation links, have not been studied before.

The outline of the rest of the paper is as follows. In Section III, we consider the physically degraded BC with cooperating decoders. In Section IV, we consider the MAC with cribbing encoders,, and in Section V, we provide proofs for all our results.

II. Notation Conventions

We use $H(\cdot)$ to denote the entropy of a discrete random variable (RV), and $I(\cdot;\cdot)$ to denote the mutual information

This work was supported by the ISRAEL SCIENCE FOUNDATION (ISF) (grant no. 684/11).

between two discrete RVs. Calligraphic letters denote (discrete and finite) sets, e.g., $\mathcal{X}$, the complement of $\mathcal{X}$ is denoted by $\mathcal{X}^c$, while $|\mathcal{X}|$ stands for its cardinality. The n-fold Cartesian product of $\mathcal{X}$ is denoted by $\mathcal{X}^n$. An element of $\mathcal{X}^n$ is denoted by $x^n = (x_1, x_2, \dots, x_n)$; whenever the dimension n is clear from the context, vectors (or sequences) are denoted by boldface letters, e.g., $\boldsymbol{x}$. We denote RVs with capital letters- X, etc. We denote by $T_\epsilon^n(X)$ the weakly typical set for the (possibly vector) RV X, see [1] for the definition of this set. Finally, we denote the probability distribution of the RV X over $\mathcal{X}$ with P_X and the conditional distribution of Y given X with $P_{Y|X}$.

III. The Physically Degraded Broadcast Channel with Cooperating Decoders

Let $\mathcal{X}$, $\mathcal{Y}_1$, $\mathcal{Y}_2$ be finite sets. A broadcast channel (BC) $(\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2, P_{Y_1,Y_2|X})$ is a channel with input alphabet $\mathcal{X}$, two output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, and a transition probability $P_{Y_1,Y_2|X}$ from $\mathcal{X}$ to $\mathcal{Y}_1 \times \mathcal{Y}_2$. The BC is said to be physically degraded if for any input distribution P_X, the Markov chain $X \multimap Y_1 \multimap Y_2$ holds, i.e.,

$$P_{X,Y_1,Y_2} = P_X P_{Y_1,Y_2|X} = P_X P_{Y_1|X} P_{Y_2|Y_1}. \tag{1}$$

We will refer to Y_1 (resp. Y_2) as the stronger (resp. weaker, or degraded) user. We assume throughout that the channel is memoryless and that no feedback is present, implying that the transition probability of n-sequences is given by

$$P_{Y_1,Y_2|X}(y_1^n, y_2^n|x^n) = \prod_{i=1}^{n} P_{Y_1,Y_2|X}(y_{1,i}, y_{2,i}|x_i). \tag{2}$$

Fix the transmission length, n, and an integer $\nu_{1,2}$. Let $\mathcal{N}_{1,2} = \{1, 2, \dots, \nu_{1,2}\}$ be the index set of the conference message. Denote the sets of messages by $\mathcal{N}_k = \{1, 2, \dots, \nu_k\}$, $k = 1, 2$, and $\mathcal{N}_2' = \{1, 2, \dots, \nu_2'\}$ where ν_1, ν_2 and ν_2' are integers. A code for the BC with unreliable conference link, that may or may not be present, operates as follows. Three messages M_1, M_2, and M_2' are drawn uniformly and independently from the sets $\mathcal{N}_1$, $\mathcal{N}_2$, and $\mathcal{N}_2'$, respectively. The encoder maps this triplet to a channel input sequence, $\boldsymbol{x}(M_1, M_2, M_2')$. At the channel output, Decoder k has the output sequence Y_k^n, $k = 1, 2$, at hand. Decoder 1 (resp. Decoder 2) is required to decode the message M_1 (resp. M_2), whether or not the conference link is present. If the conference link is present, Decoder 1 sends a message $c \in \mathcal{N}_{1,2}$ to Decoder 2, based on the output sequence Y_1^n. I.e., $c = c(Y_1^n)$. Finally, Decoder 2 decodes M_2' based on his output Y_2^n and the conference message $c(Y_1^n)$. The setting of the problem is depicted in Fig. 1.

Observe that only Decoder 2 benefits when the conference link is present. Indeed, since there is only a link from Decoder 1 to Decoder 2, whatever Decoder 1 can do with the link, he can also do without it. Therefore the rate to User 1 is independent of whether the link is present or not. Only User 2 can benefit from its existence, and thus there are two sets of messages intended to User 2 - $\mathcal{N}_2$ and $\mathcal{N}_2'$.

In the following, we give a more formal description of the above described structure.

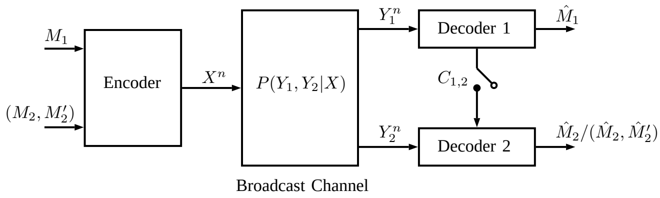

Fig. 1. Broadcast channel with unreliable cooperating decoders.

Definition 1: An $(n, \nu_1, \nu_2, \nu_2', \nu_{1,2}, \epsilon)$ code for the BC $P_{Y_1,Y_2|X}$ with an unreliable conference link is an encoder mapping

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

a conference mapping

$$h : \mathcal{Y}_1^n \to \mathcal{N}_{1,2},$$

and three decoding maps:

$$g_1 : \mathcal{Y}_1^n \to \mathcal{N}_1, \tag{3a}$$
$$g_2 : \mathcal{Y}_2^n \to \mathcal{N}_2, \tag{3b}$$
$$g_2' : \mathcal{Y}_2^n \times \mathcal{N}_{1,2} \to \mathcal{N}_2', \tag{3c}$$

such that the average probabilities of error P_e and P_e' do not exceed ϵ. Here,

$$P_e = \frac{1}{\nu_1 \nu_2 \nu_2'} \sum_{m_1, m_2, m_2'} P_{Y_1,Y_2|X}(S_e | f(m_1, m_2, m_2')) \tag{4a}$$
$$P_e' = \frac{1}{\nu_1 \nu_2 \nu_2'} \sum_{m_1, m_2, m_2'} P_{Y_1,Y_2|X}(S_e' | f(m_1, m_2, m_2')) \tag{4b}$$

where the sets S_e and S_e' are defined as

$$\begin{aligned} S_e(m_1, m_2) &= \{(\boldsymbol{y}_1, \boldsymbol{y}_2) : g_1(\boldsymbol{y}_1) \neq m_1 \text{ or } g_2(\boldsymbol{y}_2) \neq m_2\} \\ S_e'(m_1, m_2, m_2') &= S_e(m_1, m_2) \cup \\ &\qquad \{(\boldsymbol{y}_1, \boldsymbol{y}_2) : g_2'(\boldsymbol{y}_2, h(\boldsymbol{y}_1)) \neq m_2'\}, \end{aligned} \tag{5}$$

and for notational convenience, the dependence of S_e and S_e' on the messages is dropped in (4).

The conference rate $C_{1,2}$ and the communications rates (R_1, R_2, R_2') are defined as usual:

$$C_{1,2} = \frac{\log \nu_{1,2}}{n}, \quad R_k = \frac{\log \nu_k}{n}, k = 1, 2, \quad R_2' = \frac{\log \nu_2'}{n}.$$

The interpretation of the rates is as follows: $C_{1,2}$ is the conference rate in case that it is present. The rate R_k is intended to User k, $k = 1, 2$, to be decoded whether or not the conference is present. The rate R_2' is intended to User 2 and is the extra rate gained if the conference link is present.

A rate quadruple $(R_1, R_2, R_2', C_{1,2})$ is said to be achievable with unreliable conference if for any $\epsilon > 0$, $\gamma > 0$, and sufficiently large n there exists an $(n, e^{n(R_1 - \gamma)}, e^{n(R_2 - \gamma)}, e^{n(R_2' - \gamma)}, e^{n(C_{1,2} + \gamma)}, \epsilon)$ code for the BC with unreliable conference link. The capacity region is the closure of the set of all achievable quadruples $(R_1, R_2, R_2', C_{1,2})$ and is denoted by $\mathcal{C}$. For a given conference rate $C_{1,2}$, $\mathcal{C}(C_{1,2})$ stands for the section of $\mathcal{C}$ at $C_{1,2}$. Our interest is to characterize $\mathcal{C}(C_{1,2})$.

Let $\mathcal{R}(C_{1,2})$ be the convex hull of all rate triples (R_1, R_2, R_2') satisfying

$$R_2 \leq I(U;Y_2), \tag{6a}$$
$$R_2' \leq \min\left\{I(V;Y_2|U) + C_{1,2}, I(V;Y_1|U)\right\}, \tag{6b}$$
$$R_1 \leq I(X;Y_1|U,V), \tag{6c}$$

for some joint distribution of the form

$$P_{U,V,X,Y_1,Y_2} = P_{U,V} P_{X|U,V} P_{Y_1,Y_2|X} \tag{6d}$$

where $|\mathcal{U}| \leq |\mathcal{X}|+3$, and $|\mathcal{V}| \leq (|\mathcal{X}|+2)(|\mathcal{X}|+3)$. Our main result on the physical degraded BC with unreliable conference is the following

Theorem 1: For any physically degraded BC with unreliable conference of rate $C_{1,2}$,

$$\mathcal{C}(C_{1,2}) = \mathcal{R}(C_{1,2}).$$

The proof is given in Section V. Given the last result, we note to the following observations:

- Let us examine the region $\mathcal{R}(C_{1,2})$ in the following interesting extreme case. Assume that $C_{1,2} = 0$, that is, the case where even if the conference link is present, its rate is 0, and so there is no benefit from the conference link. Due to (6d) the Markov condition $(U,V) \multimap Y_1 \multimap Y_2$ holds, implying, of course, also that $V \multimap (U,Y_1) \multimap Y_2$ holds. Therefore, when $C_{1,2} = 0$, it is readily seen that the bounds in (6) reduce to

$$R_2 \leq I(U;Y_2), \tag{7a}$$
$$R_2' \leq I(V;Y_2|U), \tag{7b}$$
$$R_1 \leq I(X;Y_1|U,V). \tag{7c}$$

The total rate to User 2 is $R_2 + R_2'$. Now, it is easy to verify that after optimization over (U,V), the rates guaranteed by (7) coincide with the capacity region of the degraded BC, as one should expect. Indeed, we have:

$$R_2' + R_2 \leq I(U,V;Y_2), \tag{8a}$$
$$R_1 \leq I(X;Y_1|U,V), \tag{8b}$$

and so, by letting $\tilde{U} \triangleq (U,V)$ where $P_{\tilde{U},X,Y_1,Y_2} = P_{\tilde{U}} P_{X|\tilde{U}} P_{Y_1,Y_2|X}$, we obtain the capacity region of the degraded BC.

- Another case of interest is when $R_2 = 0$. Here, User 2 will not get any rate if the conference link is absent. Choosing U to be a null RV, the region of rates (R_1, R_2') guaranteed by (6) reduces to

$$R_2' \leq \min\left\{I(V;Y_2) + C_{1,2}, I(V;Y_1)\right\}, \tag{9a}$$
$$R_1 \leq I(X;Y_1|V), \tag{9b}$$

which coincides with the result in [3, Theorem 1].

- Theorem 1 can be easily generalized to encounter cases in which there is an input constraint of the form $\mathbb{E}\left[\sum_{i=1}^n \Gamma(X_i)\right] \leq nP$. In this case the achievable region is given by Theorem 1 where the additional constraint $\mathbb{E}\left[X_1\right] \leq P$ is needed. The achievability proof of the Theorem 1 with the input constraint is the same since by the law of large numbers the constraints are satisfied with high probability. The Converse is also similar just adding a step of introducing a "time sharing" RV Q, uniformly distributed over the set $\{1,2,\ldots,n\}$, and independent of the other RVs. Finally, define $U = (Q, U_Q)$ and $V = (Q, U_Q)$, and everything goes along without any problem.

IV. The Multiple Access Channel with Cribbing Encoders

A multiple access channel (MAC) is a quadruple $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}, P_{Y|X_1,X_2})$, where $\mathcal{X}_k$ is the input alphabet of User k, $k = 1,2$, $\mathcal{Y}$ is the output alphabet, and $P_{Y|X_1,X_2}$ is the transition probability matrix from $\mathcal{X}_1 \times \mathcal{X}_2$ to $\mathcal{Y}$. The channel is memoryless without feedback.

In this section we present achievable rates for the MAC with an unreliable cribbing - that may or may not be present - from Encoder 1 to Encoder 2. The basic assumptions are as follows. Since Encoder 2 listens to Encoder 1, he knows whether the cribbing link is present. Similarly, the decoder knows it since Encoder 2 can convey to him this message, as it is only one bit of information to transmit. Encoder 1, on the other hand, does not know whether the cribbing link is present, since he cannot be informed about it. He is only aware that cribbing could occur. Let $\mathcal{N}_1' = \{1,2,\ldots,\nu_1'\}$ and $\mathcal{N}_2'' = \{1,2,\ldots,\nu_2''\}$ be two message sets. A coding scheme operates as follows. Four messages M_1, M_1', M_2, and M_2'' are drawn uniformly and independently from the sets $\mathcal{N}_1$, $\mathcal{N}_1'$, $\mathcal{N}_2$, $\mathcal{N}_2''$, respectively. Encoder 1 maps the pair (M_1, M_1') to an input sequence $\boldsymbol{x}_1 = \boldsymbol{x}_1(M_1, M_1')$. If the cribbing link is absent, Encoder 2 maps the message M_2 to to an input sequence $\boldsymbol{x}_2 = \boldsymbol{x}_2(M_2)$. If the cribbing link is present, Encoder 2 knows $\boldsymbol{x}_1$ strictly causally, thus maps the pair $(M_2'', \boldsymbol{x}_1)$ to an input sequence $\boldsymbol{x_2}''$, in a strictly causal manner:

$$\begin{aligned}\boldsymbol{x}_2''(m_2'', \boldsymbol{x}_1) = (&x_{2,1}''(m_2''), x_{2,2}''(m_2'', x_{1,1}),\\ &\ldots, x_{2,n}''(m_2'', x_1^{n-1})).\end{aligned} \tag{10}$$

At the output, the decoder decodes (M_1, M_2) if cribbing is absent, and (M_1, M_1', M_2'') if cribbing is present.

Note that there is a slight difference in the interpretation of the message sets, compared to the message sets of the BC model studied in Section III. The pair (M_1, M_1') is encoded by User 1, where M_1 is always decoded, and M_1' is decoded only if cribbing is present. For User 2, if cribbing is absent, M_2 is encoded, whereas if cribbing is present, M_2'' is encoded. Therefore User 2 can reduce his rate in case of cribbing, in favor of increasing the rate of User 1. Due to this structure, the joint distribution of M_2 and M_2'' is immaterial, as they never appear together in the coding scheme. The setting of the problem is depicted in Fig. 2.

Following is a formal definition of the scheme described above.

Definition 2: An $(n, \nu_1, \nu_1', \nu_2, \nu_2'', \epsilon)$ code for the MAC $P_{Y|X_1,X_2}$ with unreliable strictly causal cribbing link consist

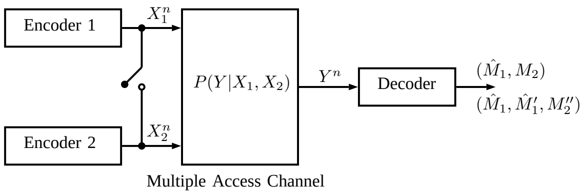

Fig. 2. MAC with unreliable cribbing encoders.

of $n+2$ encoding maps

$$f_1 : \mathcal{N}_1 \times \mathcal{N}_1' \to \mathcal{X}_1^n, \tag{11a}$$
$$f_2 : \mathcal{N}_2 \to \mathcal{X}_2^n, \tag{11b}$$
$$f''_{2,i} : \mathcal{N}''_2 \times \mathcal{X}_1^{i-1} \to \mathcal{X}_{2,i}, \quad i = 1, 2, \ldots, n, \tag{11c}$$

and a pair of decoding maps

$$g : \mathcal{Y}^n \to \mathcal{N}_1 \times \mathcal{N}_2, \tag{12a}$$
$$g' : \mathcal{Y}^n \to \mathcal{N}_1 \times \mathcal{N}_1' \times \mathcal{N}_2'', \tag{12b}$$

such that the average probabilities of error P_e and P_e' do not exceed ϵ. Here

$$P_e = \frac{1}{\nu_1 \nu_1' \nu_2} \sum_{m_1, m_1', m_2} P_{Y|X_1,X_2}(\mathcal{Q}_e | f_1(m_1, m_1'), f_2(m_2)) \tag{13a}$$

$$P_e' = \frac{1}{\nu_1 \nu_1' \nu_2''} \sum_{m_1, m_1', m_2''} P_{Y|X_1,X_2}(\mathcal{Q}_e' | f_1(m_1, m_1'), \boldsymbol{f}_2''(m_2'', f_1(m_1, m_1'))) \tag{13b}$$

where $\boldsymbol{f}_2''(m_2'', f_1(m_1, m_1'))$ is the sequence of maps $f''_{2,i}$ in (11c), the sets $\mathcal{Q}_e$ and $\mathcal{Q}_e'$ are defined as

$$\mathcal{Q}_e(m_1, m_2) = \{\boldsymbol{y} : g(\boldsymbol{y}) \neq (m_1, m_2)\} \tag{14a}$$
$$\mathcal{Q}_e'(m_1, m_1', m_2'') = \{\boldsymbol{y} : g'(\boldsymbol{y}) \neq (m_1, m_1', m_2'')\} \tag{14b}$$

and the dependence of the sets $\mathcal{Q}_e$, $\mathcal{Q}_e'$ on the messages is dropped in (13), for notational convenience.

The rates (R_1, R_1', R_2, R_2''), and achievability of a given quadruple, are defined as usual. The capacity region of the MAC with unreliable strictly causal cribbing is the closure of the collection of all achievable quadruples (R_1, R_1', R_2, R_2''), and is denoted by $\mathcal{C}_{\text{mac}}^{\text{strict}}$. Our interest is in characterizing $\mathcal{C}_{\text{mac}}^{\text{strict}}$.

Let $\mathcal{U}$ and $\mathcal{V}$, be finite sets, and let $\mathcal{P}^{\text{strict}}$ be the collection of all joint distributions $P_{U,V,X_1,X_2,X_2'',Y,Y''}$ of the form

$$P_U P_V P_{X_1|U,V} P_{X_2} P_{Y|X_1,X_2} P_{X_2''|U} P_{Y''|X_1,X_2''} \tag{15}$$

where $P_{Y''|X_1,X_2''}$ is our MAC with X_2'' at the input of Encoder 2. Let $\mathcal{I}_{\text{mac}}^{\text{strict}}$ be the collection of all quadruples (R_1, R_1', R_2, R_2'') satisfying

$$R_1 \leq I(V; Y|X_2), \tag{16a}$$
$$R_2 \leq I(X_2; Y|V), \tag{16b}$$
$$R_1 + R_2 \leq I(V, X_2; Y), \tag{16c}$$
$$R_1' \leq H(X_1|U, V), \tag{16d}$$
$$R_2'' \leq I(X_2''; Y''|U, V, X_1), \tag{16e}$$
$$R_1' + R_2'' \leq I(X_1, X_2''; Y''|V), \tag{16f}$$
$$R_1 + R_1' + R_2'' \leq I(X_1, X_2''; Y''), \tag{16g}$$

for some $P_{U,V,X_1,X_2,X_2'',Y,Y''} \in \mathcal{P}^{\text{strict}}$ where

$$|\mathcal{U}| \leq \min\{|\mathcal{X}_1| \cdot |\mathcal{X}_2| + 1, |\mathcal{Y}| + 2\} \tag{17}$$
$$|\mathcal{V}| \leq \min\{|\mathcal{X}_1| \cdot |\mathcal{X}_2| + 4, |\mathcal{Y}| + 5\}. \tag{18}$$

We start with the following result, which is proved in Subsection V-B.

Theorem 2 (Inner bound - strictly causal case): For any MAC with unreliable strictly causal cribbing

$$\mathcal{I}_{\text{mac}}^{\text{strict}} \subseteq \mathcal{C}_{\text{mac}}^{\text{strict}}.$$

Next, consider the case where causal cribbing, for the second user, is allowed, that is,

$$\boldsymbol{x}_2''(m_2'', \boldsymbol{x}_1) = (x''_{2,1}(m_2'', x_{1,1}), \ldots, x''_{2,n}(m_2'', x_1^n)), \tag{19}$$

or, equivalently, replace (11c) with:

$$f''_{2,i} : \mathcal{N}_2'' \times \mathcal{X}_1^i \to \mathcal{X}_{2,i}, \quad i = 1, 2, \ldots, n. \tag{20}$$

The capacity $\mathcal{C}_{\text{mac}}$ of the MAC with unreliable causal cribbing is defined similarly to the strictly causal case, but with (19) and (20), replacing (10) and (11c), respectively.

Let $\mathcal{P}$ be the collection of all joint distributions $P_{V,X_1,X_2,X_2'',Y,Y''}$ of the form

$$P_{V,X_1} P_{X_2} P_{Y|X_1,X_2} P_{X_2''|X_1} P_{Y''|X_1,X_2''}. \tag{21}$$

The interpretation of this joint distribution is as follows. The pair (V, X_1) are the coding RVs of User 1. These are fixed, regardless of whether cribbing is present or not. The input X_2 is the coding variable of User 2 if cribbing is absent, therefore it is independent of (V, X_1), and Y is the MAC output due to inputs X_1, X_2. When cribbing is present, User 2 encodes with X_2'' which can depend on X_1. The output of the channel due to inputs X_1 and X_2'' is denoted by Y''.

Let $\mathcal{I}_{\text{mac}}$ be the collection of all quadruples (R_1, R_1', R_2, R_2'') satisfying

$$R_1 \leq I(V; Y|X_2), \tag{22a}$$
$$R_2 \leq I(X_2; Y|V), \tag{22b}$$
$$R_1 + R_2 \leq I(V, X_2; Y), \tag{22c}$$
$$R_1' \leq H(X_1|V), \tag{22d}$$
$$R_2'' \leq I(X_2''; Y''|V, X_1), \tag{22e}$$
$$R_1' + R_2'' \leq I(X_1, X_2''; Y''|V), \tag{22f}$$
$$R_1 + R_1' + R_2'' \leq I(X_1, X_2''; Y''), \tag{22g}$$

for some $P_{V,X_1,X_2,X_2'',Y,Y''} \in \mathcal{P}$ where

$$|\mathcal{V}| \leq \min\{|\mathcal{X}_1| \cdot |\mathcal{X}_2| + 4, |\mathcal{Y}| + 5\}. \tag{23}$$

We have the following result, proved in Subsection V-C.

Theorem 3 (Inner bound - causal cribbing): For any MAC with unreliable causal cribbing

$$\mathcal{I}_{\text{mac}} \subseteq \mathcal{C}_{\text{mac}}.$$

We shall make several remarks on this result.

- The bounds on the cardinalities of U, and V, are derived in a similar manner as in [9, Appendix B], and is based on Fenchel-Eggleston-Carathéodory Theorem.

• The proof of Theorem 2 is based on the combination of superposition coding and block-Markov coding. The transmission is always performed in B sub-blocks, of length n each. In each sub-block, the messages of User 1 are encoded in two layers. First, the "resolution" information of User 1 are encoded with U, which depend on both messages M_1 and M_1'. Then, the fresh information of message M_1 is encoded with V, and finally, the fresh information of M_1 is encoded with X_1, using superposition coding around the cloud centers V and U. If the cribbing link is absent, Encoder 2 encodes his messages independently of Encoder 1. The decoder can then decode only the messages of V, that is, M_1, and X_2. If the cribbing link is present, block Markov coding is employed, similarly to the scheme used in [9] for one sided causal cribbing.

• Note that the main important observation in the achievability, is that User 1 must employ a universal encoding scheme, in the sense of being independent of the cribbing. User 2 and the decoder, however, can employ different encoding and decoding schemes, in accordance to existence or absence of the cribbing.

• When cribbing is absent, the rates R_1' and R_2'' are not decoded. Thus, setting $V = X_1$ in the region $\mathcal{I}_{\text{mac}}$ yields the capacity region of the MAC without cribbing, as expected.

• The r.h.s. of (16e) is smaller than that of (22e). Indeed,

$$\begin{aligned} I(X_2'';Y''|U,V,X_1) &= H(Y''|U,V,X_1) \\ &\quad - H(Y''|U,V,X_1,X_2'') \\ &\le H(Y''|V,X_1) \\ &\quad - H(Y''|V,X_1,X_2'') \\ &= I(X_2'';Y''|V,X_1) \end{aligned} \tag{24}$$

where the inequality follows from the fact that conditioning reduce entropy, and the Markov chain $(U,V) \multimap (X_1,X_2'') \multimap Y''$.

Unfortunately, we were not able to show the converse part in general, but only for some special cases, described in the forthcoming subsection. In the following, we provide an outer bound to the capacity region, assuming unreliable causal cribbing. Let $\mathcal{I}_{\text{mac}}^{O}$ be the convex hull of all rate quadruples (R_1, R_1', R_2, R_2'') satisfying

$$R_1 \le I(V;Y|X_2), \tag{25a}$$
$$R_2 \le I(X_2;Y|X_1), \tag{25b}$$
$$R_1 + R_2 \le I(V,X_2;Y), \tag{25c}$$
$$R_1' \le H(X_1|V), \tag{25d}$$
$$R_2'' \le I(X_2'';Y''|V,X_1), \tag{25e}$$
$$R_1 + R_1' + R_2'' \le I(X_1,X_2'';Y''), \tag{25f}$$

for some $P_{V,X_1,X_2,X_2'',Y,Y''} \in \mathcal{P}$. The following result is true also for the non-causal cribbing case, namely,

$$\boldsymbol{x}_2''(m_2'',\boldsymbol{x}_1) = (x_{2,1}''(m_2'',x_1^n),\ldots,x_{2,n}''(m_2'',x_1^n)), \tag{26}$$

or, equivalently, replace (11c) with:

$$f_{2,i}'' : \mathcal{N}_2'' \times \mathcal{X}_1^n \to \mathcal{X}_{2,i}, \quad i = 1,2,\ldots,n. \tag{27}$$

The following is proved in Subsection V-D.

Theorem 4 (Outer bound - causal (non-causal) case): For any MAC with unreliable causal (non-causal) cribbing

$$\mathcal{I}_{\text{mac}}^{O} \supseteq \mathcal{C}_{\text{mac}}.$$

Next, we consider a case in which we were able to derive the capacity region.

A. User #1 is always fully decoded

Consider the case where $R_1' = 0$, which means that there is no extra rate sent by User 1 to be decoded when cribbing is present. In this case, the first user is fully decoded no matter whether cribbing is present or not. Then, according to Theorem 3, it is easy to verify that an achievable region is given by:

$$R_1 \le I(V;Y|X_2), \tag{28a}$$
$$R_2 \le I(X_2;Y|V), \tag{28b}$$
$$R_1 + R_2 \le I(V,X_2;Y), \tag{28c}$$
$$R_2'' \le I(X_2'';Y''|X_1), \tag{28d}$$
$$R_1 + R_2'' \le I(X_1,X_2'';Y''), \tag{28e}$$

for some $P_{V,X_1,X_2,X_2'',Y,Y''} \in \mathcal{P}$ of the form

$$P_{V,X_1} P_{X_2} P_{Y|X_1,X_2} P_{X_2''|X_1} P_{Y''|X_1,X_2''}. \tag{29}$$

Let $\tilde{\mathcal{I}}_{\text{mac}}^{A}$ be the collection of all quadruples (R_1, R_2, R_2'') satisfying (28) and (29). In this stage, one may realize that for $R_1' = 0$, the auxiliary RV V should be superfluous, and we can actually substitute X_1 instead. This is indeed reasonable due to the fact that V is used to convey the message M_1, and the extra messages from the first user, that is M_1', is encoded by X_1. Accordingly, let $\mathcal{I}_{\text{mac}}^{A}$ be the collection of all quadruples (R_1, R_2, R_2'') satisfying:

$$R_1 \le I(X_1;Y|X_2), \tag{30a}$$
$$R_2 \le I(X_2;Y|X_1), \tag{30b}$$
$$R_1 + R_2 \le I(X_1,X_2;Y), \tag{30c}$$
$$R_2'' \le I(X_2'';Y''|X_1), \tag{30d}$$
$$R_1 + R_2'' \le I(X_1,X_2'';Y''), \tag{30e}$$

for some $P_{X_1,X_2,X_2'',Y,Y''}$ of the form

$$P_{X_1} P_{X_2} P_{Y|X_1,X_2} P_{X_2''|X_1} P_{Y''|X_1,X_2''}. \tag{31}$$

The following lemma is proved in Appendix B.

Lemma 1: The following relation holds:

$$\tilde{\mathcal{I}}_{\text{mac}}^{A} = \mathcal{I}_{\text{mac}}^{A}. \tag{32}$$

Using Lemma 1, we obtain the following result.

Theorem 5: For any MAC with unreliable causal (non-causal) cribbing, if $R_1' = 0$, then $\mathcal{I}_{\text{mac}}^{A}$ is the capacity region.

Proof: The result follows directly by substituting $R_1' = 0$ in the outer bound (25), and noticing that it coincides with the achievable region in (30). □

According to (30), if the first user is fully decoded no matter whether cribbing is present or not, then there is no bound on the individual rate of the first user when cribbing is present (we have only bounds on the rate of the second user (30d) and on the sum rate (30e)). Instead, as can be seen from (30d)-(30e), it is assumed that X_1 is already known to the receiver when cribbing is present. The reason is that since cribbing can only help in recovering X_1, the bound on the individual rate of the first user when cribbing is absent dominates (or, more strict). To illustrate the result in Theorem 5, we consider the following example.

Example 1: Consider the example where the channel output, Y, is given by:

$$Y = X_1 \oplus X_2 \oplus Z_1 \oplus Z_2 \tag{33}$$

where X_1, X_2, N_1, and N_2, are binary RVs, where Z_1 is Bernoulli with $\Pr\{Z_1 = 0\} = p_1$, $Z_2 = 0$ if $X_1 = 0$, and it is Bernoulli with $\Pr\{Z_2 = 0\} = p_2$, otherwise (i.e., if $X_1 = 1$). Here, X_1, X_2, Z_1 and Z_2, are independent. When cribbing is present, the channel output, Y'', is given by:

$$Y'' = X_1 \oplus X_2'' \oplus Z_1 \oplus Z_2 \tag{34}$$

where now X_1 may depend on X_2''. Let $\Pr\{X_i = 0\} \triangleq P_{X_i}$, for $i = 1, 2$, $\Pr\{X_2'' = 0|X_1 = 0\} = \mu_1$, and $\Pr\{X_2'' = 0|X_1 = 1\} = \mu_2$. Also, for two real numbers $0 \leq a, b \leq 1$, define $a * b \triangleq a \cdot \bar{b} + \bar{a} \cdot b$, and $a \star b \triangleq a \cdot b + \bar{a} \cdot \bar{b}$, where $\bar{a} \triangleq 1 - a$. Finally, let:

$$\alpha \triangleq (p_1 \star p_2) \cdot P_{X_2} + (p_1 * p_2) \cdot \bar{P}_{X_2}, \tag{35}$$

$$\beta \triangleq (p_1 \star p_2) \cdot \mu_2 + (p_1 * p_2) \cdot \bar{\mu}_2. \tag{36}$$

Then, using the above definition, it is a simple exercise to check that

$$\begin{aligned}
R_1 &\leq h_2(P_{X_1} p_1 + \bar{P}_{X_1}(p_1 \star p_2)) - P_{X_1} h_2(p_1) \\
&\quad - \bar{P}_{X_1} h_2(p_1 * p_2), \\
R_2 &\leq P_{X_1} h_2(p_1 \star P_{X_2}) + \bar{P}_{X_1} h_2(\alpha) - P_{X_1} h_2(p_1) \\
&\quad - \bar{P}_{X_1} h_2(p_1 * p_2), \\
R_1 + R_2 &\leq h_2(P_{X_1}(p_1 \star P_{X_2}) + \bar{P}_{X_1}\bar{\alpha}) - P_{X_1} h_2(p_1) \\
&\quad - \bar{P}_{X_1} h_2(p_1 * p_2), \\
R_2'' &\leq P_{X_1} h_2(p_1 \star \mu_1) + \bar{P}_{X_1} h_2(\beta) - P_{X_1} h_2(p_1) \\
&\quad - \bar{P}_{X_1} h_2(p_1 * p_2), \\
R_1 + R_2'' &\leq h_2(P_{X_1}(p_1 \star \mu_1) + \bar{P}_{X_1}\bar{\beta}) - P_{X_1} h_2(p_1) \\
&\quad - \bar{P}_{X_1} h_2(p_1 * p_2).
\end{aligned} \tag{37}$$

where $h_2(\cdot)$ is the binary entropy. Fig. 3 depicts the capacity region for the case where $p_1 = 0.01$ and $p_2 = 0.1$. The capacity region was numerically evaluated using (37). In the figure, we present two curves corresponding to the constraints related to the rates (R_1, R_2) which refer to the case where cribbing is absent (the blue curve), and the constraints related to the rates (R_1, R_2'') which refer to the case where cribbing is present (the red curve). It is evident that higher rates can be achieved for the second user due to the cribbing.

V. Proofs

A. Proof of Theorem 1

In this subsection, we prove Theorem 1. The direct part uses random selection and strong typicality arguments.

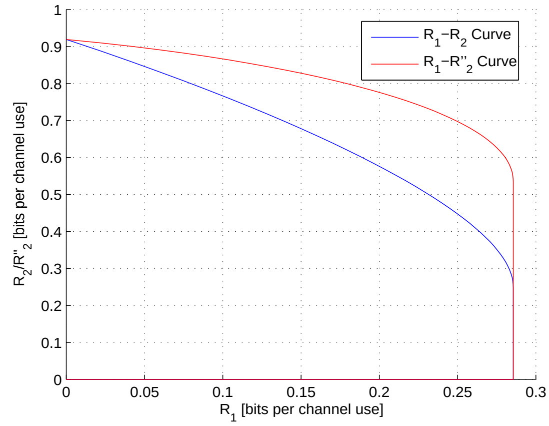

Fig. 3. The capacity region for $p_1 = 0.01$ and $p_2 = 0.1$.

Direct Part. We use the binning approach suggested in [4]. We will start with the code construction.

Codebook construction: Fix a joint distribution $P_{U,V,X}$.

1) Generate e^{nR_2} codewords $\boldsymbol{u}(j)$, $j = 1, 2, \ldots e^{nR_2}$, i.i.d., according to P_U.
2) For every $\boldsymbol{u}(j)$, generate $e^{nR_2'}$ codewords $\boldsymbol{v}(k|j)$, $k = 1, 2, \ldots e^{nR_2'}$, independently according to $\prod_{i=1}^n P_{V|U}(v_i|u_i(j))$.
3) For every j, distribute the $e^{nR_2'}$ codewords $\boldsymbol{v}(k|j)$, $k = 1, 2, \ldots, e^{nR_2'}$, into $e^{nC_{1,2}}$ bins, evenly and independently of each other. Thus, in every bin there are $e^{n(R_2' - C_{1,2})}$ codewords $\boldsymbol{v}(k|j)$ with a fixed index j. Denote by $b(k|j)$ the bin number to which $\boldsymbol{v}(k|j)$ belongs. Note that
$$1 \leq b(k|j) \leq e^{nC_{1,2}}. \tag{38}$$
4) For every pair $(\boldsymbol{u}(j), \boldsymbol{v}(k|j))$, $j = 1, 2, \ldots, e^{nR_2}$, $k = 1, 2, \ldots, e^{nR_2'}$, generate e^{nR_1} vectors $\boldsymbol{x}(l|j,k)$, $l = 1, 2, \ldots, e^{nR_1}$, independently of each other, according to $\prod_{i=1}^n P_{X|U,V}(x_i|u_i(j), v_i(k|j))$.

These codewords form the codebook, which is revealed to the encoder and the decoders.

Encoding: Given a triple (j, k, l), where $j = 1, 2, \ldots, e^{nR_2}$, $k = 1, 2, \ldots, e^{nR_2'}$, $l = 1, 2, \ldots, e^{nR_1}$, the encoder sends via the channel the codeword $\boldsymbol{x}(l|j,k)$.

Decoding: We assume first that the conference link is absent. Decoder 2 has $\boldsymbol{y}_2$ at hand. He looks for the unique index $\hat{j}$ in $\{1, 2, \ldots, \exp(nR_2)\}$ such that

$$(\boldsymbol{u}(\hat{j}), \boldsymbol{y}_2) \in T_\epsilon^{(n)}(UY_2).$$

If such $\hat{j}$ does not exist, or there is more than one such index, an error is declared. By classical results, if

$$R_2 < I(U; Y_2), \tag{39}$$

the index j is decoded correctly with high probability.

Decoder 1 has $\boldsymbol{y}_1$ at hand. He looks for the unique index $\hat{\hat{j}}$ in $\{1, 2, \ldots, \exp(nR_2)\}$ such that

$$(\boldsymbol{u}(\hat{\hat{j}}), \boldsymbol{y}_1) \in T_\epsilon^{(n)}(UY_1).$$

If such $\hat{\hat{j}}$ does not exist, or there is more than one such index, an error is declared. By classical results, if

$$R_2 < I(U;Y_1), \tag{40}$$

Decoder 1 succeeds to decode correctly the index j with high probability. Since the channel is degraded, if (39) holds, it implies (40). Next, Decoder 1 looks for the unique index $\hat{\hat{k}}$ in $\{1,2,\ldots,\exp(nR_2')\}$ such that

$$(\boldsymbol{u}(\hat{\hat{j}}), \boldsymbol{v}(\hat{\hat{k}}|\hat{\hat{j}}), \boldsymbol{y}_1) \in T_\epsilon^{(n)}(VY_1|U). \tag{41}$$

If such $\hat{\hat{k}}$ does not exist, or there is more than one such, an error is declared. By classical results, the index $k \in \{1,2,\ldots,\exp(nR_2')\}$ is decoded correctly with high probability if

$$R_2' < I(V;Y_1|U). \tag{42}$$

Having the pair $(\hat{\hat{j}}, \hat{\hat{k}})$ at hand, Decoder 1 looks for the unique index $\hat{\hat{l}} \in \{1,2,\ldots,\exp(nR_1)\}$ satisfying

$$(\boldsymbol{u}(\hat{\hat{j}}), \boldsymbol{v}(\hat{\hat{k}}|\hat{\hat{j}}), \boldsymbol{x}(\hat{\hat{l}}|\hat{\hat{k}},\hat{\hat{j}}), \boldsymbol{y}_1) \in T_\epsilon^{(n)}(XY_1|UV). \tag{43}$$

By classical results, this step succeeds if the rate R_1 satisfies

$$R_1 < I(X;Y_1|U,V). \tag{44}$$

This concludes the decoding process when the conference link is absent. By (39), (42) and (44), the conditions for correct decoding when there is no conferencing are

$$R_2 \le I(U;Y_2), \tag{45a}$$
$$R_2' \le I(V;Y_1|U), \tag{45b}$$
$$R_1 \le I(X;Y_1|U,V). \tag{45c}$$

Observe that, although the rate R_2' is decoded by Decoder 1 (if (45b) is satisfied), it does not arrive to User 2, since the conferencing link is absent. The bound (45b) is still needed in order to guarantee that Decoder 1 can proceed and decode the index l (the message intended to him).

We turn now to the case where the conference link is present. Decoder 1 operates exactly as in the case of no conference, and decodes the indices $\hat{\hat{j}}$, $\hat{\hat{k}}$, and $\hat{\hat{l}}$. If (45) hold, these steps succeed with high probability. He then sends $b(\hat{\hat{k}}|\hat{\hat{j}})$, the index of the bin to which $\boldsymbol{v}(\hat{\hat{k}}|\hat{\hat{j}})$ belongs, via the conference link. Due to (38), the link capacity suffices, and Decoder 2 receives $b(\hat{\hat{k}}|\hat{\hat{j}})$ without an error.

Decoder 2 decodes the index $\hat{j}$ as in the case of no conference. After receiving from Decoder 1 the bin index $b(\hat{\hat{k}}|\hat{\hat{j}})$, he looks in this bin for the unique index $\hat{k}$ such that

$$(\boldsymbol{v}(\hat{k}|\hat{j}), \boldsymbol{u}(\hat{j}), \boldsymbol{y}_2) \in T_\epsilon^{(n)}(VY_2|U). \tag{46}$$

If such an index does not exist, or there is more than one such, an error is declared. From the code construction, every bin contains approximately $e^{n(R_2'-C_{1,2})}$ codewords $\boldsymbol{v}$. Assuming that the previous decoding steps were successful (i.e., $\hat{\hat{j}}$, $\hat{\hat{k}}$, $\hat{j}$ are the correct indices for j, k, and j, respectively), by classical results $\hat{k}$ is correct with high probability if

$$R_2' - C_{1,2} \le I(V;Y_2|U). \tag{47}$$

The region defined by (45) and (47) coincides with $\mathcal{R}(C_{1,2})$. This concludes the proof of the achievability part.

Converse Part. We start with a sequence of codes $(n, e^{nR_1}, e^{nR_2}, e^{nR_2'}, e^{nC_{1,2}}, \epsilon_n)$ with increasing blocklength n, satisfying $\lim_{n\to\infty}\epsilon_n = 0$. We denote by M_k the random message from $\mathcal{N}_k$, $k=1,2$, and by M_2' the message from $\mathcal{N}_2'$. The conference message is denoted by $M_{1,2}$. By Fano's inequality we can bound the rate R_2 as

$$\begin{aligned} nR_2 - n\delta_n \;&\le\; I(M_2;Y_2^n) &(48)\\ &\overset{(a)}{=} \sum_{i=1}^n I(M_2;Y_{2,i}|Y_2^{i-1}) \\ &\le \sum_{i=1}^n I(M_2,Y_2^{i-1};Y_{2,i}), &(49)\end{aligned}$$

where $\lim_{n\to\infty}\delta_n = 0$, due to $\lim_{n\to\infty}\epsilon_n = 0$, and (a) follows from the chain rule. We now bound the rate R_2' as follows. If the conference link is present, then the messages M_2' can be decoded by Decoder 2 based on Y_2^n and the message transmitted via the conference link, $M_{1,2}$. Therefore

$$\begin{aligned} nR_2' - n\delta_n &\le I(M_2';Y_2^n,M_{1,2}|M_2) &(50)\\ &= I(M_2';Y_2^n|M_2) + I(M_2';M_{1,2}|M_2,Y_2^n) \\ &\le I(M_2';Y_2^n|M_2) + H(M_{1,2}) \\ &= \sum_{i=1}^n I(M_2';Y_{2,i}|M_2,Y_2^{i-1}) + H(M_{1,2}) \\ &\le \sum_{i=1}^n I(M_2',Y_1^{i-1};Y_{2,i}|M_2,Y_2^{i-1}) + H(M_{1,2}). \end{aligned}$$

Moreover, the message M_2' can be decoded by Decoder 1, regardless of the conference link. Hence:

$$\begin{aligned} nR_2' - n\delta_n \;&\le\; I(M_2';Y_1^n|M_2) \\ &= \sum_{i=1}^n I(M_2';Y_{1,i}|M_2,Y_1^{i-1}) \\ &\overset{(a)}{=} \sum_{i=1}^n I(M_2';Y_{1,i}|M_2,Y_1^{i-1},Y_2^{i-1}) \\ &\le \sum_{i=1}^n I(M_2',Y_1^{i-1};Y_{1,i}|M_2,Y_2^{i-1}), &(51)\end{aligned}$$

where (a) is true because the channel is physically degraded. The rate R_1 can be bounded by

$$\begin{aligned} nR_1 - n\delta_n \;&\le\; I(M_1;Y_1^n|M_2,M_2') \\ &= \sum_{i=1}^n I(M_1;Y_{1,i}|M_2,M_2',Y_1^{i-1}) \\ &\overset{(a)}{=} \sum_{i=1}^n I(M_1;Y_{1,i}|M_2,M_2',Y_1^{i-1},Y_2^{i-1}) \end{aligned}$$

$$\overset{(b)}{=} \sum_{i=1}^{n} I(X_i; Y_{1,i}|M_2, M_2', Y_1^{i-1}, Y_2^{i-1}), \quad (52)$$

where (a) is true since the channel is physically degraded. Equality (b) holds since X_i is a deterministic function of the messages M_1, M_2, and M_2', and since $Y_{1,i}$ is independent of $(M_2, M_2', Y_2^{i-1}, Y_1^{i-1}, M_1)$ when conditioned on X_i. Defining $U_i = (M_2, Y_2^{i-1})$, $V_i = (M_2', Y_1^{i-1})$ and using the fact that

$$\frac{1}{n}H(M_{1,2}) \leq C_{1,2}, \quad (53)$$

we obtain from (49), (50), (51), and (52) the bounds

$$n(R_2 - \delta_n) \leq \sum_{i=1}^{n} I(U_i; Y_{2,i}), \quad (54a)$$

$$n(R_2' - \delta_n) \leq \sum_{i=1}^{n} I(V_i; Y_{2,i}|U_i) + nC_{1,2}, \quad (54b)$$

$$n(R_2' - \delta_n) \leq \sum_{i=1}^{n} I(U_i; Y_{1,i}|V_i), \quad (54c)$$

$$n(R_1 - \delta_n) \leq \sum_{i=1}^{n} I(X_i; Y_{1,i}|U_i, V_i). \quad (54d)$$

Using the standard time-sharing argument as in [13, Ch. 14.3], one can rewrite (54) by introducing an appropriate time-sharing random variable. Therefore, if $\epsilon_n \to 0$ as $n \to \infty$, the convex hull of this region can be shown to be equivalent to the convex hull of the region in (6).

Finally, the bounds on the cardinalities of U and V follow from Fenchel-Eggleston-Carathéodry Theorem, similarly as used for the 3-receiver degraded BC [12, Appendix C]. □

B. Proof of Theorem 2

The proof of Theorem 2 is based on the combination of superposition coding and block-Markov coding. The transmission is always performed in B sub-blocks, of length n each. In each sub-block, the messages of User 1 are encoded in two layers. First the message M_1 is encoded with V, and then the message M_1' is encoded with X_1, using superposition coding around the cloud centers V. If the cribbing link is absent, Encoder 2 encodes his messages independently of Encoder 1. The decoder can then decode only the messages of V and X_2. If the cribbing link is present, block Markov coding is employed, similarly to the scheme used in [9] for one sided strictly causal cribbing.

It is important to emphasize that User 1 must employ a universal encoding scheme, in the sense of being *independent* of the cribbing. User 2 and the decoder, however, can employ different encoding and decoding schemes, in accordance to existence or absence of the cribbing. Accordingly, in the sequel, we describe the encoding scheme for the first user separately.

We use a random coding argument to demonstrate the achievability part. The messages $M_{1,b} \in \{1, 2, \dots, \exp(nR_1)\}$ and $M_{1,b}' \in \{1, 2, \dots, \exp(nR_1')\}$, for $b = 1, 2, \dots, B-1$, which are uniformly distributed and independent of each other, will be sent over the MAC in B blocks, each of n transmissions. Note that if $B \to \infty$, the overall rates are $R_1(B-1)/B \to R_1$ and $R_1'(B-1)/B \to R_1'$. In each of the B blocks the same codebook is used, and is constructed, for the first user, as follows.

Codebook construction for User 1: Fix a joint distribution $P_U P_V P_{X_1|U,V}$, and a sufficiently small $\epsilon > 0$.

1) Generate $e^{n(R_1+R_1)}$ codewords $\boldsymbol{v}$, i.i.d., according to P_V. Label them $\boldsymbol{v}(m_0, m_1)$, for $m_0, m_1 \in \{1, 2, \dots, \exp(nR_1)\}$.
2) Generate $e^{n(R_1+R_1')}$ codewords $\boldsymbol{u}$, independently according to P_U. Label them $\boldsymbol{u}(m_0, m_0')$, for $m_0 \in \{1, 2, \dots, \exp(nR_1)\}$ and $m_0' \in \{1, 2, \dots, \exp(nR_1')\}$.
3) For every $\boldsymbol{v}(m_0, m_1)$ and $\boldsymbol{u}(m_0, m_0')$, generate $e^{nR_1'}$ codewords $\boldsymbol{x}_1$, independently according to $\prod_{i=1}^{n} P_{X_1|U,V}(x_{1,i}|u_i(m_0, m_0'), v_i(m_0, m_1))$. Label them $\boldsymbol{x}_1(m_1', \boldsymbol{u}(m_0, m_0'), \boldsymbol{v}(m_0, m_1))$, for $m_1' \in \{1, 2, \dots, \exp(nR_1')\}$.

We now present the achievability scheme for the case where cribbing is absent.

1) Cribbing is absent: The message $M_{2,b} \in \{1, 2, \dots, \exp(nR_2)\}$, for $b = 1, 2, \dots, B-1$, is uniformly distributed, independent of the messages of the first user, and will be sent over the MAC in B blocks, each of n transmissions. If $B \to \infty$, the overall rate is $R_2(B-1)/B \to R_2$. In each of the B blocks the same codebook is used, and is constructed, for the second user, as follows.

Codebook construction for User 2: Fix a distribution P_{X_2}, and a sufficiently small $\epsilon > 0$. Generate e^{nR_2} codewords $\boldsymbol{x}_2$, i.i.d., according to P_{X_2}. Label them $\boldsymbol{x}_2(m_2)$, for $m_2 \in \{1, 2, \dots, \exp(nR_2)\}$.

The codewords of Users 1 and 2 form the codebook, which is revealed to the encoders and the decoder. The messages $m_{1,b} \in \{1, \dots, \exp(nR_1)\}$, $m_{1,b}' \in \{1, \dots, \exp(nR_1')\}$, and $m_{2,b} \in \{1, \dots, \exp(nR_2)\}$, $b = 1, \dots, B-1$, are encoded in the following way.

Encoding: In block 1, the encoders send:

$$\boldsymbol{x}_{1,1} = \boldsymbol{x}_1(m_{1,1}', \boldsymbol{u}(1,1), \boldsymbol{v}(1, m_{1,1})) \quad (55a)$$

$$\boldsymbol{x}_{2,1} = \boldsymbol{x}_2(m_{2,1}). \quad (55b)$$

Then, in block b, $b = 2, 3, \dots, B$, the encoders send (57), shown at the top of the next page.

Decoding: We employ simultaneous joint typicality decoding. At the end of the first block, the decoder looks for $(\hat{m}_{1,1}, \hat{m}_{2,1})$ such that:

$$(\boldsymbol{v}(1, \hat{m}_{1,1}), \boldsymbol{x}_2(\hat{m}_{2,1}), \boldsymbol{y}) \in T_\epsilon^{(n)}(VX_2Y). \quad (58)$$

Next, assume that the decoder has correctly found $\hat{m}_{1,1}$. Then, to find the transmitted information at the end of the second block, the decoder looks for $(\hat{m}_{1,2}, \hat{m}_{2,2})$ such that:

$$(\boldsymbol{v}(\hat{m}_{1,1}, \hat{m}_{1,2}), \boldsymbol{x}_2(\hat{m}_{2,2}), \boldsymbol{y}) \in T_\epsilon^{(n)}(VX_2Y). \quad (59)$$

With the knowledge of $\hat{m}_{1,2}$ the information at the end of the third block can be decoded in a similar manner. In general,

$$\boldsymbol{x}_{1,b} = \boldsymbol{x}_1(m'_{1,b}, \boldsymbol{u}(m_{1,b-1}, m'_{1,b-1}), \boldsymbol{v}(m_{1,b-1}, m_{1,b})), \quad b = 2, 3, \ldots, B-1 \tag{57a}$$
$$\boldsymbol{x}_{2,b} = \boldsymbol{x}_2(m_{2,b}), \quad b = 2, 3, \ldots, B-1 \tag{57b}$$
$$\boldsymbol{x}_{1B} = \boldsymbol{x}_1(1, \boldsymbol{u}(m_{1,B-1}, m'_{1,B-1}), \boldsymbol{v}(m_{1,B-1}, 1)), \tag{57c}$$
$$\boldsymbol{x}_{2B} = \boldsymbol{x}_2(m_{2,B}). \tag{57d}$$

---

at the end of block b the decoder looks for $(\hat{m}_{1,b}, \hat{m}_{2,b})$ such that:

$$(\boldsymbol{v}(\hat{m}_{1,b-1}, \hat{m}_{1,b}), \boldsymbol{x}_2(\hat{m}_{2,b}), \boldsymbol{y}) \in T_\epsilon^{(n)}(VX_2Y) \tag{60}$$

where $\hat{m}_{1,b-1}$ was decoded in the previous block.

Error Analysis: By classical results (e.g., standard MAC), there exists a sequence of codes with a probability of error that goes to zero as the block length goes to infinity, if:

$$R_1 \le I(V;Y|X_2), \tag{61a}$$
$$R_2 \le I(X_2;Y|V), \tag{61b}$$
$$R_1 + R_2 \le I(V, X_2; Y). \tag{61c}$$

This concludes the decoding process when the conference link is absent.

2) Cribbing is present: We turn now to the case where the cribbing link is present. The message $M''_{2,b} \in \{1, 2, \ldots, \exp(nR''_2)\}$, for $b = 1, 2, \ldots, B-1$, is uniformly distributed, independent of the messages of the first user, and will be sent over the MAC in B blocks, each of n transmissions. In each of the B blocks the same codebook is used, and is constructed, for the second user, as follows.

Codebook construction for User 2: Fix a distribution $P_{X''_2|U}$, and a sufficiently small $\epsilon > 0$. For every $\boldsymbol{u}(m_0, m'_0)$, generate $e^{nR''_2}$ codewords $\boldsymbol{x}''_2$, independently according to $\prod_{i=1}^n P_{X''_2|U}(x_{2,i}|u_i(m_0, m'_0))$. Label them $\boldsymbol{x}''_2(m''_2, \boldsymbol{u}(m_0, m'_0))$, for $m''_2 \in \{1, 2, \ldots, \exp(nR''_2)\}$. The codewords of Users 1 and 2 form the codebook, which is revealed to the encoders and the decoder.

Encoding: The messages $m_{1,b} \in \{1, \ldots, \exp(nR_1)\}$, $m'_{1,b} \in \{1, \ldots, \exp(nR'_1)\}$, and $m''_{2,b} \in \{1, \ldots, \exp(nR''_2)\}$, $b = 1, \ldots, B-1$, are encoded in the following way: In block 1, the encoders send[1]:

$$\boldsymbol{x}_{1,1} = \boldsymbol{x}_1(m'_{1,1}, \boldsymbol{u}(1,1), \boldsymbol{v}(1, m_{1,1})) \tag{62a}$$
$$\boldsymbol{x}''_{2,1} = \boldsymbol{x}''_2(m''_{2,1}, \boldsymbol{u}(1,1)). \tag{62b}$$

Assume that as a result of cribbing from encoder 1, after block b, $b = 1, 2, \ldots, B-1$, encoder 2 has estimates $\hat{m}_{1,b}$ and $\hat{m}'_{1,b}$, for $m_{1,b}$ and $m'_{1,b}$, respectively. To this end, encoder 2 first chooses $\hat{m}_{1,b}$ such that:

$$(\boldsymbol{v}(\hat{m}_{1,b-1}, \hat{m}_{1,b}), \boldsymbol{x}_{1,b}) \in T_\epsilon^{(n)}(VX_1) \tag{63}$$

where $\hat{m}_{1,b-1}$ was determined at the end of block $b-1$ (recall that $m_{1,0} = 1$). Then, given $\hat{m}_{1,b}$, he chooses $\hat{m}'_{1,b}$ according to (64), shown at the top of the next page, where $\hat{m}'_{1,b-1}$ was determined at the end of block $b-1$. Finally, in block b, $b = 2, 3, \ldots, B$, the encoders send (65), shown at the top of the next page.

Decoding: Here, the principle of backward decoding [9] is used to find the transmitted information. In the last block, block B, the decoder looks for $(\hat{m}_{1,B-1}, \hat{m}'_{1,B-1})$ such that

$$\begin{aligned}\big\{&\boldsymbol{u}(\hat{m}_{1,B-1}, \hat{m}'_{1,B-1}), \boldsymbol{v}(\hat{m}_{1,B-1}, 1),\\ &\boldsymbol{x}_1(1, \boldsymbol{u}(\hat{m}_{1,B-1}, \hat{m}'_{1,B-1}), \boldsymbol{v}(\hat{m}_{1,B-1}, 1)),\\ &\boldsymbol{x}''_2(1, \boldsymbol{u}(\hat{m}_{1,B-1}, \hat{m}'_{1,B-1})), \boldsymbol{y}''_B\big\}\\ &\qquad\qquad \in T_\epsilon^{(n)}(UVX_1X''_2Y''). \end{aligned} \tag{66}$$

Next, in block $B-1$, the decoder has at hand an estimate of the fresh information sent in block $B-1$, namely, $(\hat{m}_{1,B-1}, \hat{m}'_{1,B-1})$, and to find the transmitted information in block $B-1$ the decoder looks for[2] $(\hat{m}_{1,B-2}, \hat{m}'_{1,B-2}, \hat{m}''_{2,B-1})$ according to (67), shown at the top of the next page. Then, in block $B-2$, the decoder has at hand an estimate of the fresh information sent in block $B-2$, namely, $(\hat{m}_{1,B-2}, \hat{m}'_{1,B-2})$, and the information sent in block $B-2$ can be decoded next, etc. In general, in block b, the decoder has at hand an estimate of the fresh information sent in block b, namely, $(\hat{m}_{1,b}, \hat{m}'_{1,b})$, and to find the transmitted information in block b, the decoder looks for $(\hat{m}_{1,b-1}, \hat{m}'_{1,b-1}, \hat{m}''_{2,b})$ according to (68), shown at the top of the next page.

According to the above decoding rule, the decoding of User 1 and User 2 are staggered: at some block $b \in \{1, 2, \ldots, B-1\}$, the message of User 2 is decoded jointly with the *resolution information* of User 1, and the latter estimates are actually the fresh messages of block $b-1$.

If in a decoding step (second encoder or the decoder) there is no message index (or no index pair) to satisfy the decoding rule, or if there is more than one index (or index pair), then an index (or an index pair) is chosen at random.

Error Analysis: The following lemma will enable us to bound the probability of error of the super block nB by bounding the probability of error of each block.

Lemma 2: Let $\{\mathcal{A}_l\}_{l=1}^L$ be a set of events and let $\mathcal{A}_j^c$ be the complement of the event $\mathcal{A}_j$. Then,

$$\Pr\left\{\bigcup_{l=1}^L \mathcal{A}_l\right\} \le \sum_{l=1}^L \Pr\left\{\mathcal{A}_l | \mathcal{A}_1^c, \mathcal{A}_2^c, \ldots, \mathcal{A}_{l-1}^c\right\} \tag{69}$$

where $\mathcal{A}_0 = \emptyset$.

[1] Recall that User 1 must employ the same encoding scheme as in the case of absent cribbing.

[2] The messages $(m_{1,B-2}, m'_{1,B-2})$ are the resolution information of user 1 at block $B-1$, which are actually the fresh messages of $B-2$.

$$\left\{\boldsymbol{u}(\hat{m}_{1,b-1},\hat{m}'_{1,b-1}),\boldsymbol{v}(\hat{m}_{1,b-1},\hat{m}_{1,b}),\ \ \boldsymbol{x}_1(\hat{m}'_{1,b},\boldsymbol{u}(\hat{m}_{1,b-1},\hat{m}'_{1,b-1}),\boldsymbol{v}(\hat{m}_{1,b-1},\hat{m}_{1,b})),\boldsymbol{x}_{1,b}\right\}\in T_\epsilon^{(n)}(UVX_1X_1) \tag{64}$$

$$\boldsymbol{x}_{1,b}=\boldsymbol{x}_1(m'_{1,b},\boldsymbol{u}(m_{1,b-1},m'_{1,b-1}),\boldsymbol{v}(m_{1,b-1},m_{1,b})),\quad b=2,3,\ldots,B-1 \tag{65a}$$
$$\boldsymbol{x}''_{2,b}=\boldsymbol{x}''_2(m''_{2,b},\boldsymbol{u}(\hat{m}_{1,b-1},\hat{m}'_{1,b-1})),\quad b=2,3,\ldots,B-1 \tag{65b}$$
$$\boldsymbol{x}_{1B}=\boldsymbol{x}_1(1,\boldsymbol{u}(m_{1,B-1},m'_{1,B-1}),\boldsymbol{v}(m_{1,B-1},1)), \tag{65c}$$
$$\boldsymbol{x}''_{2B}=\boldsymbol{x}''_2(1,\boldsymbol{u}(\hat{m}_{1,B-1},\hat{m}'_{1,B-1})). \tag{65d}$$

$$\begin{aligned}\big\{&\boldsymbol{u}(\hat{m}_{1,B-2},\hat{m}'_{1,B-2}),\boldsymbol{v}(\hat{m}_{1,B-2},\hat{m}_{1,B-1}),\boldsymbol{x}_1(\hat{m}'_{1,B-1},\boldsymbol{u}(\hat{m}_{1,B-2},\hat{m}'_{1,B-2}),\boldsymbol{v}(\hat{m}_{1,B-2},\hat{m}_{1,B-1}))\\&\boldsymbol{x}''_2(\hat{m}''_{2,B-1},\boldsymbol{u}(\hat{m}_{1,B-2},\hat{m}'_{1,B-2})),\boldsymbol{y}''_{B-1}\big\}\in T_\epsilon^{(n)}(UVX_1X''_2Y'').\end{aligned} \tag{67}$$

$$\begin{aligned}\big\{&\boldsymbol{u}(\hat{m}_{1,b-1},\hat{m}'_{1,b-1}),\boldsymbol{v}(\hat{m}_{1,b-1},\hat{m}_{1,b}),\boldsymbol{x}_1(\hat{m}'_{1,b},\boldsymbol{u}(\hat{m}_{1,b-1},\hat{m}'_{1,b-1}),\boldsymbol{v}(\hat{m}_{1,b-1},\hat{m}_{1,b}))\\&\boldsymbol{x}''_2(\hat{m}''_{2,b},\boldsymbol{u}(\hat{m}_{1,b-1},\hat{m}'_{1,b-1})),\boldsymbol{y}''_b\big\}\in T_\epsilon^{(n)}(UVX_1X''_2Y'').\end{aligned} \tag{68}$$

Using Lemma 2, we bound the probability of error in the super block nB by the sum of the probability of having an error in each block b given that in previous blocks, the messages were decoded correctly.

First let us bound the probability that for some b, encoder 2 decodes the messages of encoder 1 incorrectly at the end of that block. Using Lemma 2, it suffices to show that the probability of decoding error in each block goes to zero, assuming that all previous messages in blocks $(1,2,\ldots,b-1)$ were decoded correctly.

Let $E_{\text{enc},b}=E^{(1)}_{\text{enc},b}\cup E^{(2)}_{\text{enc},b}$ be the event that encoder 2 has an error in decoding $m_{1,b}$ or $m'_{1,b}$. The event $E^{(1)}_{\text{enc},b}$ refers to an error in decoding $m_{1,b}$, while $E^{(2)}_{\text{enc},b}$ refers to an error in decoding $m'_{1,b}$. The term $\Pr\left\{E_{\text{enc},b}|E^c_{\text{enc},b-1}\right\}$ is the probability that encoder 2 incorrectly decoded $m_{1,b}$ or $m'_{1,b}$, given that $m_{1,b-1}$ and $m'_{1,b-1}$ were decoded correctly. We have,

$$\begin{aligned}\Pr\left\{E_{\text{enc},b}|E^c_{\text{enc},b-1}\right\}\le{}&\Pr\left\{E^{(1)}_{\text{enc},b}|E^c_{\text{enc},b-1}\right\}\\&+\Pr\left\{E^{(2)}_{\text{enc},b}|E^c_{\text{enc},b-1},(E^{(1)}_{\text{enc},b})^c\right\}.\end{aligned} \tag{70}$$

Define the sets

$$\mathcal{E}_{b,m_{1,b}}\triangleq(\boldsymbol{v}(m_{1,b-1},m_{1,b}),\boldsymbol{x}_{1,b})\in T_\epsilon^{(n)}(VX_1), \tag{71}$$

and the set $\mathcal{E}_{b,m'_{1,b}}$ in (72), shown at the top of the next page, given $m_{1,b-1}$ and $m'_{1,b-1}$. Assume without loss of generality that $m_{1,b-1}=m'_{1,b-1}=m_{1,b}=1$. Then, according to (63),

$$\Pr\left\{E^{(1)}_{\text{enc},b}|E^c_{\text{enc},b-1}\right\}\le\Pr\left\{\bigcup_{m_{1,b}\neq1}\mathcal{E}_{b,m_{1,b}}|E^c_{\text{enc},b-1}\right\}. \tag{73}$$

The probability at the right hand side of (73), is the probability of the event in (71), given that $m_{1,b-1}$, was decoded correctly. Then, to evaluate (73), we can equivalently evaluate the probability of the event

$$\mathcal{E}_{b,m_{1,b}}\triangleq(\boldsymbol{v}(1,m_{1,b}),\boldsymbol{x}_{1,b})\in T_\epsilon^{(n)}(VX_1), \tag{74}$$

for $m_{1,b}\neq1$. Whence, by classical results, we have,

$$\begin{aligned}\Pr\left\{E^{(1)}_{\text{enc},b}|E^c_{\text{enc},b-1}\right\}&\le\sum_{m_{1,b}\neq1}\Pr\left\{\mathcal{E}_{b,m_{1,b}}|E^c_{\text{enc},b-1}\right\} &&(75)\\&\le\sum_{m_{1,b}\neq1}e^{-n(I(V;X_1)-3\epsilon)} &&(76)\\&\le e^{n(R_1-I(V;X_1)+3\epsilon)}. &&(77)\end{aligned}$$

Next, recall that encoder 2 decodes $m'_{1,b}$ according to (64), given that he already decoded $\hat{m}_{1,b}$ in the first stage, and $\hat{m}_{1,b-1}$ and $\hat{m}'_{1,b-1}$ at the end of block $b-1$. Accordingly, we have,

$$\begin{aligned}&\Pr\left\{E^{(2)}_{\text{enc},b}|E^c_{\text{enc},b-1},(E^{(1)}_{\text{enc},b})^c\right\}\\&=\Pr\left\{\bigcup_{m'_{1,b}\neq1}\mathcal{E}_{b,m'_{1,b}}|E^c_{\text{enc},b-1},(E^{(1)}_{\text{enc},b})^c\right\}\end{aligned}$$

$$\mathcal{E}_{b,m'_{1,b}} \triangleq \left\{\boldsymbol{u}(m_{1,b-1}, m'_{1,b-1}), \boldsymbol{v}(m_{1,b-1}, m_{1,b}), \boldsymbol{x}_1(m'_{1,b}, \boldsymbol{u}(m_{1,b-1}, m'_{1,b-1}), \boldsymbol{v}(m_{1,b-1}, m_{1,b})), \boldsymbol{x}_{1,b}\right\} \in T_\epsilon^{(n)}(UVX_1X'_1). \tag{72}$$

$$\leq \sum_{m'_{1,b}\neq 1} \Pr\left\{\mathcal{E}_{b,m'_{1,b}} | E^c_{\text{enc},b-1}, (E^{(1)}_{\text{enc},b})^c\right\}. \tag{78}$$

Again, the probability at the right hand side of (78), is the probability of the event in (72), given that $m_{1,b-1}$, $m'_{1,b-1}$, and $m_{1,b}$, were decoded correctly. Then, to evaluate (78), we can equivalently evaluate the probability of the event in (79), shown at the top of the next page, for $m'_{1,b} \neq 1$. We get

$$\begin{aligned}
&\Pr\left\{\mathcal{E}_{b,m'_{1,b}} | E^c_{\text{enc},b-1}, (E^{(1)}_{\text{enc},b})^c\right\} = \\
&\sum_{T_\epsilon^{(n)}(UVX_1X_1)} P(\boldsymbol{u})P(\boldsymbol{v})P(\boldsymbol{x}_1|\boldsymbol{u},\boldsymbol{v})P(\boldsymbol{x}'_1|\boldsymbol{u},\boldsymbol{v}) \\
&\quad \leq \exp(n(H(U,V,X_1,X_1)+\epsilon)) \\
&\quad\quad \cdot \exp(-n(H(U,V,X_1)-4\epsilon)) \\
&\quad\quad \cdot \exp(-n(H(X_1|U,V)-\epsilon)) \\
&\quad = \exp(-n(H(X_1|U,V)-6\epsilon)).
\end{aligned} \tag{80}$$

Therefore,

$$\begin{aligned}
\Pr\left\{E^{(2)}_{\text{enc},b} | E^c_{\text{enc},b-1}, (E^{(1)}_{\text{enc},b})^c\right\} &\leq \sum_{m'_{1,b}\neq 1} e^{-n(H(X_1|U,V)-6\epsilon)} \\
&\leq e^{n(R'_1 - H(X_1|U,V)+6\epsilon)}.
\end{aligned} \tag{81}$$

Wrapping up, using (77) and (81), by Lemma 2, if $R_1 \leq I(V;X_1)$ and $R'_1 \leq H(X_1|U,V)$, then encoder 2 can decode all the messages (i.e., over all the B blocks) of encoder 1 correctly, with a probability of error that goes to zero as the block length goes to infinity.

Next, at the receiver side, recall first the decoding rule in (68), where in block b, the decoder looks for $(\hat{m}_{1,b-1}, \hat{m}'_{1,b-1}, \hat{m}''_{2,b})$ assuming that $(\hat{m}_{1,b}, \hat{m}'_{1,b})$ were already decoded in block $b+1$. In the following, we upper bound the overall error probability of the receiver. To this end, we use once again Lemma 2, as follows. The error probability of the receiver is upper bounded by the sum of the probabilities that in each block b the receiver incorrectly decodes the messages $m_{1,b-1}$, $m'_{1,b-1}$, and $m''_{2,b}$, given that: (1) at block $b+1$ the messages $m_{1,b}$ and $m'_{1,b}$ were decoded correctly, and (2) encoder 2 decoded correctly all the messages of encoder 1 (in all the B blocks).

Define the event in (82), shown at the top of the next page, and without loss of generality, assume that $m_{1,b} = m'_{1,b} = 1$. Assuming that $m_{1,b-1} = m'_{1,b-1} = m''_{2,b} = 1$, an error occurs if either the correct codewords are not jointly typical with the received sequences, i.e., $E^c_{1,1,1,b}$, or if there exists a different tuple $(m_1, m'_1, m''_2) \neq (1,1,1)$ such that $E_{m_1,m'_1,m''_2,b}$ occurs. Let $P^{(n)}_{e,b}$ be the decoding error probability at block b given that in blocks $(b+1,\ldots,B)$, there was no decoding error. From the union bound, we obtain that:

$$\begin{aligned}
P^{(n)}_{e,b} \leq \Pr\left\{E^c_{1,1,1,b}\right\} &+ \sum_{m_1>1} \Pr\left\{E_{m_1,1,1,b}\right\} \\
&+ \sum_{m'_1>1} \Pr\left\{E_{1,m'_1,1,b}\right\} + \sum_{m''_2>1} \Pr\left\{E_{1,1,m''_2,b}\right\} \\
&+ \sum_{m_1>1,m'_1>1} \Pr\left\{E_{m_1,m'_1,1,b}\right\} \\
&+ \sum_{m_1>1,m''_2>1} \Pr\left\{E_{m_1,1,m''_2,b}\right\} \\
&+ \sum_{m'_1>1,m''_2>1} \Pr\left\{E_{1,m'_1,m''_2,b}\right\} \\
&+ \sum_{m_1>1,m'_1>1,m''_2>1} \Pr\left\{E_{m'_1,m_1,m''_2,b}\right\}.
\end{aligned} \tag{83}$$

Let us upper bound each term in (83).

1) Upper-bounding $\Pr\left\{E^c_{1,1,1,b}\right\}$: Since we assume that encoder 2 encodes the right messages $m_{1,b-1}$ and $m'_{1,b-1}$ in block b, and that the receiver decoded the right messages $m_{1,b}$ and $m'_{1,b}$ at block $b+1$, by the LLN $\Pr\left\{E^c_{1,1,1,b}\right\} \to 0$ as $n\to\infty$.
2) Upper-bounding $\sum_{m''_2>1}\Pr\left\{E_{1,1,m''_2,b}\right\}$: Let $\mathcal{S}$ be the set of all sequences $(\boldsymbol{u},\boldsymbol{v},\boldsymbol{x}_1,\boldsymbol{x}''_2,\boldsymbol{y}'')$ that belong to $T_\epsilon^{(n)}(UVX_1X''_2Y'')$. We then have

$$\begin{aligned}
&\Pr\left\{E_{1,1,m''_2,b}\right\} \\
&= \sum_{\mathcal{S}} P(\boldsymbol{u})P(\boldsymbol{v})P(\boldsymbol{x}_1|\boldsymbol{u},\boldsymbol{v})P(\boldsymbol{x}''_2|\boldsymbol{u}) \\
&\qquad \times P(\boldsymbol{y}''|\boldsymbol{u},\boldsymbol{v},\boldsymbol{x}_1) \\
&\leq \exp(n(H(U,V,X_1,X''_2,Y'')+\epsilon)) \\
&\quad \cdot \exp(-n(H(U,V,X_1)-4\epsilon)) \\
&\quad \cdot \exp(-n(H(X''_2|U)-\epsilon)) \\
&\quad \cdot \exp(-n(H(Y''|U,V,X_1)-\epsilon)) \\
&= \exp(-n(I(X''_2;Y''|U,V,X_1)-7\epsilon)).
\end{aligned} \tag{84}$$

Hence, we obtain

$$\sum_{m''_2>1} \Pr\left\{E_{1,1,m''_2,b}\right\} \leq e^{n(R''_2 - I(X''_2;Y''|U,V,X_1)+7\epsilon)}. \tag{85}$$

3) Upper-bounding $\sum_{m_1>1}\Pr\left\{E_{m_1,1,1,b}\right\}$: We have

$$\begin{aligned}
&\Pr\left\{E_{m_1,1,1,b}\right\} \\
&= \sum_{\mathcal{S}} P(\boldsymbol{u})P(\boldsymbol{v})P(\boldsymbol{x}_1|\boldsymbol{u},\boldsymbol{v})P(\boldsymbol{x}''_2|\boldsymbol{u})P(\boldsymbol{y}'') \\
&\leq \exp(n(H(U,V,X_1,X''_2,Y'')+\epsilon)) \\
&\quad \cdot \exp(-n(H(U,V,X_1)-4\epsilon))
\end{aligned}$$

$$\tilde{\mathcal{E}}_{b,m'_{1,b}} \triangleq \left\{ \boldsymbol{u}(1,1), \boldsymbol{v}(1,1), \boldsymbol{x}_1(m'_{1,b}, \boldsymbol{u}(1,1), \boldsymbol{v}(1,1)), \boldsymbol{x}_{1,b} \right\} \in T_\epsilon^{(n)}(UVX_1X_1). \tag{79}$$

$$E_{m_1,m'_1,m''_2,b} \triangleq \left\{ \boldsymbol{u}(m_1,m'_1), \boldsymbol{v}(m_1,m_{1,b}), \boldsymbol{x}_1(m'_{1,b}, \boldsymbol{u}(m_1,m'_1), \boldsymbol{v}(m_1,m_{1,b})),\ \boldsymbol{x}''_2(m''_2, \boldsymbol{u}(m_1,m'_1)), \boldsymbol{y}''_b \right\} \\ \in T_\epsilon^{(n)}(UVX_1X''_2Y''). \tag{82}$$

$$\cdot \exp(-n(H(X''_2|U) - \epsilon)) \\ \cdot \exp(-n(H(Y'') - \epsilon)) \\ = \exp(-n(I(U,V,X_1,X''_2;Y'') - 7\epsilon)). \tag{86}$$

Hence, we obtain

$$\sum_{m_1>1} \Pr\left\{E_{m_1,1,1,b}\right\} \le e^{n(R_1 - I(U,V,X_1,X''_2;Y'')+7\epsilon)}. \tag{87}$$

4) Upper-bounding $\sum_{m'_1>1} \Pr\left\{E_{1,m'_1,1,b}\right\}$: We have

$$\Pr\left\{E_{1,m'_1,1,b}\right\} \\ = \sum_{\mathcal{S}} P(\boldsymbol{u})P(\boldsymbol{v})P(\boldsymbol{x}_1|\boldsymbol{u},\boldsymbol{v})P(\boldsymbol{x}''_2|\boldsymbol{u})P(\boldsymbol{y}''|\boldsymbol{v}) \\ \le \exp(n(H(U,V,X_1,X''_2,Y'') + \epsilon)) \\ \cdot \exp(-n(H(U,V,X_1) - 4\epsilon)) \\ \cdot \exp(-n(H(X''_2|U) - \epsilon)) \\ \cdot \exp(-n(H(Y''|V) - \epsilon)) \\ = \exp(-n(I(U,X_1,X''_2;Y''|V) - 7\epsilon)). \tag{88}$$

Hence, we obtain

$$\sum_{m'_1>1} \Pr\left\{E_{1,m'_1,1,b}\right\} \le e^{n(R'_1 - I(U,X_1,X''_2;Y''|V)+7\epsilon)}. \tag{89}$$

5) Upper-bounding $\sum_{m_1>1,m'_1>1} \Pr\left\{E_{m_1,m'_1,1,b}\right\}$: We have

$$\Pr\left\{E_{m_1,m'_1,1,b}\right\} \\ = \sum_{\mathcal{S}} P(\boldsymbol{u})P(\boldsymbol{v})P(\boldsymbol{x}_1|\boldsymbol{u},\boldsymbol{v})P(\boldsymbol{x}''_2|\boldsymbol{u})P(\boldsymbol{y}'') \\ \le \exp(n(H(U,V,X_1,X''_2,Y'') + \epsilon)) \\ \cdot \exp(-n(H(U,V,X_1) - 4\epsilon)) \\ \cdot \exp(-n(H(X''_2|U) - \epsilon)) \\ \cdot \exp(-n(H(Y'') - \epsilon)) \\ = \exp(-n(I(U,V,X_1,X''_2;Y'') - 7\epsilon)). \tag{90}$$

Hence, we obtain

$$\sum_{m_1>1,m'_1>1} \Pr\left\{E_{m_1,m'_1,1,b}\right\} \\ \le e^{n(R_1+R'_1 - I(U,V,X_1,X''_2;Y'')+7\epsilon)}. \tag{91}$$

6) Upper-bounding $\sum_{m_1>1,m''_2>1} \Pr\left\{E_{m_1,1,m''_2,b}\right\}$: We have

$$\Pr\left\{E_{m_1,1,m''_2,b}\right\} \\ = \sum_{\mathcal{S}} P(\boldsymbol{u})P(\boldsymbol{v})P(\boldsymbol{x}_1|\boldsymbol{u},\boldsymbol{v})P(\boldsymbol{x}''_2|\boldsymbol{u})P(\boldsymbol{y}'') \\ \le \exp(n(H(U,V,X_1,X''_2,Y'') + \epsilon)) \\ \cdot \exp(-n(H(U,V,X_1) - 4\epsilon)) \\ \cdot \exp(-n(H(X''_2|U) - \epsilon)) \\ \cdot \exp(-n(H(Y'') - \epsilon)) \\ = \exp(-n(I(U,V,X_1,X''_2;Y'') - 7\epsilon)). \tag{92}$$

Hence, we obtain

$$\sum_{m_1>1,m''_2>1} \Pr\left\{E_{m_1,1,m''_2,b}\right\} \\ \le e^{n(R_1+R''_2 - I(U,V,X_1,X''_2;Y'')+7\epsilon)}. \tag{93}$$

7) Upper-bounding $\sum_{m'_1>1,m''_2>1} \Pr\left\{E_{1,m'_1,m''_2,b}\right\}$: We have

$$\Pr\left\{E_{1,m'_1,m''_2,b}\right\} \\ = \sum_{\mathcal{S}} P(\boldsymbol{u})P(\boldsymbol{v})P(\boldsymbol{x}_1|\boldsymbol{u},\boldsymbol{v})P(\boldsymbol{x}''_2|\boldsymbol{u})P(\boldsymbol{y}''|\boldsymbol{v}) \\ \le \exp(n(H(U,V,X_1,X''_2,Y'') + \epsilon)) \\ \cdot \exp(-n(H(U,V,X_1) - 4\epsilon)) \\ \cdot \exp(-n(H(X''_2|U) - \epsilon)) \\ \cdot \exp(-n(H(Y''|V) - \epsilon)) \\ = \exp(-n(I(U,X_1,X''_2;Y''|V) - 7\epsilon)). \tag{94}$$

Hence, we obtain

$$\sum_{m'_1>1,m''_2>1} \Pr\left\{E_{1,m'_1,m''_2,b}\right\} \\ \le e^{n(R'_1+R''_2 - I(U,X_1,X''_2;Y''|V)+7\epsilon)}. \tag{95}$$

8) Upper-bounding $\sum_{m_1>1,m'_1>1,m''_2>1} \Pr\left\{E_{m_1,m'_1,m''_2,b}\right\}$: We have

$$\Pr\left\{E_{m_1,m'_1,m''_2,b}\right\} \\ = \sum_{\mathcal{S}} P(\boldsymbol{u})P(\boldsymbol{v})P(\boldsymbol{x}_1|\boldsymbol{u},\boldsymbol{v})P(\boldsymbol{x}''_2|\boldsymbol{u})P(\boldsymbol{y}'') \\ \le \exp(n(H(U,V,X_1,X''_2,Y'') + \epsilon)) \\ \cdot \exp(-n(H(U,V,X_1) - 4\epsilon))$$

$$\cdot \exp(-n(H(X_2''|U)-\epsilon))$$
$$\cdot \exp(-n(H(Y'')-\epsilon))$$
$$= \exp(-n(I(U,V,X_1,X_2'';Y'')-7\epsilon)). \tag{96}$$

Hence, we obtain

$$\sum_{m_1>1,m_1'>1,m_2''>1} \Pr\left\{E_{m_1,m_1',m_2'',b}\right\} \leq e^{n(R_1+R_1'+R_2''-I(U,V,X_1,X_2'';Y'')+7\epsilon)}. \tag{97}$$

Thus, if (R_1, R_1', R_2'') satisfy:

$$R_1 \leq I(V;X_1), \tag{98a}$$
$$R_1' \leq H(X_1|U,V), \tag{98b}$$
$$R_1 \leq I(U,V,X_1,X_2'';Y''), \tag{98c}$$
$$R_1' \leq I(U,X_1,X_2'';Y''|V), \tag{98d}$$
$$R_1+R_1' \leq I(U,V,X_1,X_2'';Y''), \tag{98e}$$
$$R_2'' \leq I(X_2'';Y''|U,V,X_1), \tag{98f}$$
$$R_1+R_2'' \leq I(U,V,X_1,X_2'';Y''), \tag{98g}$$
$$R_1'+R_2'' \leq I(U,X_1,X_2'';Y''|V), \tag{98h}$$
$$R_1+R_1'+R_2'' \leq I(U,V,X_1,X_2'';Y''), \tag{98i}$$

then there exists a sequence of codes with a probability of error that goes to zero as the block length goes to infinity. We note to the following simplifications. First, we can remove (98c), (98e), and (98g), due to (98i), and (98d) can be removed due to (98h). Second, (98h) and (98i) can be replaced with $R_1' + R_2'' \leq I(X_1, X_2''; Y''|V)$ and $R_1+R_1'+R_2'' \leq I(X_1,X_2'';Y'')$, respectively, due to the Markov chain $(U,V) \multimap (X_1,X_2'') \multimap Y''$. Finally, the constraint in (98a), is superfluous due to (61a). Indeed,

$$I(V;Y|X_2) = H(V|X_2) - H(V|X_2,Y) \tag{99}$$
$$\stackrel{(a)}{\leq} H(V) - H(V|X_1,X_2,Y) \tag{100}$$
$$\stackrel{(b)}{=} H(V) - H(V|X_1) \tag{101}$$
$$= I(V;X_1) \tag{102}$$

where (a) follows from the fact that conditioning reduces entropy, and (b) follows from the Markov chain $(X_2,Y) \multimap X_1 \multimap V$. Thus, to summarize, using the above simplifications, the achievable region for the MAC with unreliable strictly causal cribbing is given (recall, (61))

$$R_1 \leq I(V;Y|X_2), \tag{103a}$$
$$R_2 \leq I(X_2;Y|V), \tag{103b}$$
$$R_1+R_2 \leq I(V,X_2;Y), \tag{103c}$$
$$R_1' \leq H(X_1|U,V), \tag{103d}$$
$$R_2'' \leq I(X_2'';Y''|U,V,X_1), \tag{103e}$$
$$R_1'+R_2'' \leq I(X_1,X_2'';Y''|V), \tag{103f}$$
$$R_1+R_1'+R_2'' \leq I(X_1,X_2'';Y''), \tag{103g}$$

for some $P_{U,V,X_1,X_2,X_2'',Y,Y''}$ of the form

$$P_U P_V P_{X_1|U,V} P_{X_2} P_{X_2''|U} P_{Y|X_1,X_2} P_{Y''|X_1,X_2''}, \tag{104}$$

as stated in Theorem 2.

C. *Proof of Theorem 3*

In order to show that all the rate pairs in (22) are achievable, we employ Shannon strategies [9]. Consider all different strategies (functions), with members $t \in \mathscr{T} \triangleq \mathcal{X}_2^{|\mathcal{X}_1|}$ that map inputs $x_1 \in \mathcal{X}_1$ into inputs $x_2'' \in \mathcal{X}_2$. Denote by $t(\cdot)$ the strategy with member t as an operator.

Definition 3: For a DMMAC $(\mathcal{X}_1 \times \mathcal{X}_2, P(y''|x_1,x_2''), \mathcal{Y})$ the DM *derived* MAC is denoted by $(\mathcal{X}_1 \times \mathscr{T}, P^{\triangle}(y''|x_1,t), \mathcal{Y})$ where $P^{\triangle}(y''|x_1,t) \triangleq P(y''|x_1,x_2'' = t(x_1))$ for all $x_1 \in \mathcal{X}_1$, $t \in \mathscr{T}$, and $y'' \in \mathcal{Y}$.

Let $\mathcal{R}_S$ be the set of rates (R_1, R_1', R_2, R_2'') satisfying

$$R_1 \leq I(V;Y|X_2), \tag{105a}$$
$$R_2 \leq I(X_2;Y|V), \tag{105b}$$
$$R_1+R_2 \leq I(V,X_2;Y), \tag{105c}$$
$$R_1' \leq H(X_1|U,V), \tag{105d}$$
$$R_2'' \leq I(T;Y''|U,V,X_1), \tag{105e}$$
$$R_1'+R_2'' \leq I(X_1,T;Y''|V), \tag{105f}$$
$$R_1+R_1'+R_2'' \leq I(X_1,T;Y''), \tag{105g}$$

for some joint distribution $P(u,v,x_1,x_2,t,y,y'')$ of the form

$$P(u,v,x_1,x_2,t,y,y'') = P(u)P(v)P(x_1|u,v)P(x_2)P(t|u)P(y|x_1,x_2)P^{\triangle}(y''|x_1,t). \tag{106}$$

By the achievability scheme for the strictly causal case (Theorem 2), all rate pairs inside $\mathcal{R}_S$ are achievable for the above derived MAC. Therefore for the MAC with causal cribbing all rate pairs inside $\mathcal{R}_S$ must be achievable. If we now restrict the distributions in (106) to satisfy

$$P(u,v,x_1,x_2,t,y,y'') = P(u)P(v)P(x_1|v)P(x_2)P(t) \times P(y|x_1,x_2)P^{\triangle}(y''|x_1,t), \tag{107}$$

then

$$H(X_1|U,V) = H(X_1|V), \tag{108a}$$
$$I(T;Y''|U,V,X_1) = I(X_2'';Y''|V,X_1), \tag{108b}$$
$$I(X_1,T;Y''|V) = I(X_1,X_2'';Y''|V), \tag{108c}$$
$$I(X_1,T;Y'') = I(X_1,X_2'';Y''), \tag{108d}$$

and[3]

$$P(v,x_1,x_2'',y'') = P(v,x_1) \sum_{t:\, t(x_1)=x_2''} P(t)P(y''|x_1,x_2''). \tag{109}$$

[3]Recall that for a discrete random variable X with probability mass function $P_X(\cdot)$, the probability mass function $P_Y(\cdot)$ of the discrete random variable $Y = g(X)$ is given by

$$P_Y(y) = \sum_{x:\, y=g(x)} P_X(x).$$

Now, given an arbitrary distribution $P^0(v, x_1, x_2'') = P^0(v, x_1)P^0(x_2''|x_1)$, we note that there always exists a product distribution $P(v, x_1, t) = P(v, x_1)P(t)$ such that

$$P(v, x_1) \sum_{t:\ t(x_1)=x_2''} P(t) = P^0(v, x_1, x_2''). \tag{110}$$

Indeed, this holds for the following choice:

$$P(v, x_1) = \sum_{x_2''} P^0(v, x_1, x_2''), \tag{111a}$$

$$P(t) = \sum_{x_1'} \frac{P^0(x_1', x_2'' = t(x_1'))}{P(x_1')}. \tag{111b}$$

From the above (recall also (61)), we conclude that all rate pairs

$$R_1 \le I(V; Y|X_2), \tag{112a}$$
$$R_2 \le I(X_2; Y|V), \tag{112b}$$
$$R_1 + R_2 \le I(V, X_2; Y), \tag{112c}$$
$$R_1' \le H(X_1|V), \tag{112d}$$
$$R_2'' \le I(X_2''; Y''|V, X_1), \tag{112e}$$
$$R_1' + R_2'' \le I(X_1, X_2''; Y''|V), \tag{112f}$$
$$R_1 + R_1' + R_2'' \le I(X_1, X_2''; Y''), \tag{112g}$$

for some $P_{V,X_1,X_2,X_2'',Y,Y''}$ of the form

$$P_{V,X_1} P_{X_2} P_{X_2''|X_1} P_{Y|X_1,X_2} P_{Y''|X_1,X_2''}, \tag{113}$$

are achievable for the MAC with causal cribbing. Finally, note that due to the Markov chains $V \multimap X_1 \multimap X_2''$ and $V \multimap (X_1, X_2'') \multimap Y''$ we can replace (112e) with $R_2'' \le I(X_2''; Y''|X_1)$, and we obtain Theorem 3. Finally, we show that each constraint in (112) can be upper bounded by each constraint in (22), and that that the upper bounds can be achieved by taking $V = \emptyset$, which of course satisfy (113). Indeed, for example, for (112a), we have

$$\begin{aligned} I(V; Y|X_2, U) &= H(Y|X_2, U) - H(Y|X_2, U, V) \\ &\le H(Y|X_2) - H(Y|X_2, V) \\ &= I(V; Y|X_2) \end{aligned} \tag{114}$$

where the inequality follows from the fact that conditioning reduces entropy, and the Markov chain $U \multimap (V, X_2) \multimap Y$ which follows from the Markov chains $U \multimap (X_1, X_2) \multimap Y$ and $U \multimap V \multimap X_1$. In the same way, we get that $I(X_2; Y|V, U) \le I(X_2; Y|V)$ and $I(V, X_2; Y|U) \le I(V, X_2; Y)$. Finally, for (112e), we have

$$\begin{aligned} I(X_2''; Y''|U, V, X_1) &= H(Y''|U, V, X_1) \\ &\quad - H(Y''|U, V, X_1, X_2'') \\ &\le H(Y''|V, X_1) - H(Y''|V, X_1, X_2'') \\ &= I(X_2''; Y''|V, X_1) \end{aligned} \tag{115}$$

where the inequality follows from the fact that conditioning reduces entropy, and the Markov chains $(U, V) \multimap (X_1, X_2'') \multimap Y''$. However, we readily see that the above upper bounds can be achieved by simply taking $V = \emptyset$, which coincides with Theorem 3. □

D. Proof of Theorem 4

We next show that $\mathcal{I}^O_{\text{mac}}$, defined in (25), is an outer bound to the capacity region. We start with a sequence of codes $(n, e^{nR_1}, e^{nR_1'}, e^{nR_2}, e^{nR_2''}, \epsilon_n)$ with increasing blocklength n, satisfying $\lim_{n\to\infty} \epsilon_n = 0$. We denote by M_k the random message from $\mathcal{N}_k$, for $k = 1, 2$, and by M_1' and M_2'' the messages from $\mathcal{N}_1'$ and $\mathcal{N}_2''$, respectively. If the cribbing is absent, by Fano's inequality we can bound the rate R_1 as follows

$$nR_1 - n\delta_n \le I(M_1; Y^n|M_2) \tag{116}$$
$$= \sum_{i=1}^n I\left(M_1; Y_i|Y^{i-1}, M_2\right) \tag{117}$$
$$\overset{(a)}{\le} \sum_{i=1}^n I\left(M_1, Y^{i-1}; Y_i|M_2\right) \tag{118}$$
$$\overset{(b)}{=} \sum_{i=1}^n I(M_1; Y_i|M_2) + I(Y^{i-1}; Y_i|M_1, M_2) \tag{119}$$
$$\overset{(c)}{=} \sum_{i=1}^n I(M_1; Y_i|M_2, X_{2,i}) + I(Y^{i-1}; Y_i|M_1, M_2) \tag{120}$$
$$\overset{(a)}{\le} \sum_{i=1}^n I(M_1, M_2; Y_i|X_{2,i}) + I(Y^{i-1}; Y_i|M_1, M_2) \tag{121}$$
$$\begin{aligned} &\overset{(b)}{=} \sum_{i=1}^n I(M_1; Y_i|X_{2,i}) + I(M_2; Y_i|M_1, X_{2,i}) \\ &\qquad + I(Y^{i-1}; Y_i|M_1, M_2) \end{aligned} \tag{122}$$
$$\overset{(d)}{=} \sum_{i=1}^n I(M_1; Y_i|X_{2,i}) + I(Y^{i-1}; Y_i|M_1, M_2) \tag{123}$$

where $\lim_{n\to\infty} \delta_n = 0$, due to $\lim_{n\to\infty} \epsilon_n = 0$, (a) follows from the chain rule for mutual information and the non-negativity of the mutual information, (b) follows from the chain rule for mutual information, (c) is due to the fact that $X_{2,i}$ is a deterministic function of M_2, and (d) follows from the Markov chain $M_2 \multimap (M_1, X_{2,i}) \multimap Y_i$, proved in Appendix A (see, Lemma 3). Thus, $I(M_2; Y_i|M_1, X_{2,i}) = 0$. Continuing, note that $I(Y^{i-1}; Y_i|M_1, M_2)$, appearing in (123), can be upper bounded as follows

$$I(Y^{i-1}; Y_i|M_1, M_2) \overset{(a)}{=} I(Y^{i-1}; Y_i|M_1, M_2, X_2^i) \tag{124}$$
$$\overset{(b)}{\le} I(X_1^{i-1}; Y_i|M_1, M_2, X_2^i) \tag{125}$$
$$\le I(X_2^{i-1}, M_2, X_1^{i-1}; Y_i|M_1, X_{2i}) \tag{126}$$
$$\begin{aligned} &\overset{(c)}{=} I(X_1^{i-1}; Y_i|M_1, X_{2,i}) \\ &\quad + I(X_2^{i-1}, M_2; Y_i|M_1, X_{2,i}, X_1^{i-1}) \end{aligned} \tag{127}$$
$$\overset{(d)}{=} I(X_1^{i-1}; Y_i|M_1, X_{2,i}) \tag{128}$$

where (a) is due to the fact that X_2^i is a deterministic function of M_2, (b) follows from the fact that $Y^{i-1} \multimap (X_1^{i-1}, X_2^i, M_1, M_2) \multimap Y_i$ (see, Lemma 3), and the fact that for any Markov chain $Z \multimap (X,W) \multimap Y$ we have[4]

$$I(Z;Y|W) \leq I(X;Y|W). \tag{129}$$

Passage (c) follows from the chain rule of mutual information, and finally (d) is due to the Markov chain $(M_2, X_2^{i-1}) \multimap \left(M_1, X_1^{i-1}, X_{2,i}\right) \multimap Y_i$ (see, Lemma 3). Wrapping up, we obtained

$$nR_1 - n\delta_n \leq \sum_{i=1}^{n} I\left(M_1; Y_i|X_{2,i}\right) + I(X_1^{i-1}; Y_i|M_1, X_{2,i}) \tag{130}$$

$$= \sum_{i=1}^{n} I\left(M_1, X_1^{i-1}; Y_i|X_{2,i}\right). \tag{131}$$

Next, for R_2 we have:

$$nR_2 - n\delta_n \leq I(M_2; Y^n|M_1, M_1') \tag{132}$$

$$= \sum_{i=1}^{n} I(M_2; Y_i|M_1, M_1', Y^{i-1}) \tag{133}$$

$$\overset{(a)}{=} \sum_{i=1}^{n} I(X_{2,i}, M_2; Y_i|M_1, M_1', X_{1,i}, Y^{i-1}) \tag{134}$$

$$\leq \sum_{i=1}^{n} I(X_{2,i}, M_1, M_1', M_2, Y^{i-1}; Y_i|X_{1,i}) \tag{135}$$

$$\overset{(b)}{=} \sum_{i=1}^{n} I(X_{2,i}; Y_i|X_{1,i}) + I(M_1, M_1', M_2, Y^{i-1}; Y_i|X_{1,i}, X_{2,i}) \tag{136}$$

$$\overset{(c)}{=} \sum_{i=1}^{n} I(X_{2,i}; Y_i|X_{1,i}) \tag{137}$$

where (a) follows from the fact that $X_{2,i}$ and $X_{1,i}$ are deterministic functions of M_2 and (M_1, M_1'), respectively, (b) is due to the chain rule for mutual information, and (c) follows from the Markov chain $(M_1, M_1', M_2, Y^{i-1}) \multimap (X_{1,i}, X_{2,i}) \multimap Y_i$. Finally, for the sum rate we have

$$n(R_1 + R_2) - n\delta_n \leq \sum_{i=1}^{n} I\left(M_1, M_2; Y_i|Y^{i-1}\right) \tag{138}$$

$$\leq \sum_{i=1}^{n} I\left(M_1, M_2, Y^{i-1}; Y_i\right) \tag{139}$$

$$= \sum_{i=1}^{n} I\left(M_1, M_2; Y_i\right) + I(Y^{i-1}; Y_i|M_1, M_2) \tag{140}$$

where the last equality follows from the chain rule. However, we already saw that (recall (128)):

$$I(Y^{i-1}; Y_i|M_1, M_2) \leq I(X_1^{i-1}; Y_i|M_1, X_{2,i}), \tag{141}$$

[4]Take $Z = Y^{i-1}$, $X = X_1^{i-1}$, $Y = Y_i$, and $W = (M_1, M_2, X_2^i)$.

and thus

$$n(R_1 + R_2) - n\delta_n \leq \sum_{i=1}^{n} I\left(M_1, M_2; Y_i\right) + I(X_1^{i-1}; Y_i|M_1, X_{2,i}) \tag{142}$$

$$\overset{(a)}{=} \sum_{i=1}^{n} I\left(M_1, M_2, X_{2,i}; Y_i\right) + I(X_1^{i-1}; Y_i|M_1, X_{2,i}) \tag{143}$$

$$\overset{(b)}{=} \sum_{i=1}^{n} I\left(M_1, X_{2,i}; Y_i\right) + I(X_1^{i-1}; Y_i|M_1, X_{2,i}) \tag{144}$$

$$= \sum_{i=1}^{n} I\left(M_1, X_1^{i-1}, X_{2,i}; Y_i\right) \tag{145}$$

where in (a) we use the fact that $X_{2,i}$ is a deterministic function of M_2, and (b) is due to the fact that $I\left(M_1, M_2, X_{2,i}; Y_i\right) = I\left(M_1, X_{2,i}; Y_i\right) + I\left(M_2; Y_i|M_1, X_{2,i}\right)$ and that $M_2 \multimap (M_1, X_{2,i}) \multimap Y_i$.

Now, when cribbing is present, by Fano's inequality we bound the rate R_1' as follows:

$$nR_1' - n\delta_n \leq I(M_1'; Y^{n''}|M_1) \tag{146}$$

$$= I(M_1, M_1'; Y^{n''}|M_1) \tag{147}$$

$$\overset{(a)}{=} I(M_1, M_1', X_1^n; Y^{n''}|M_1) \tag{148}$$

$$\overset{(b)}{=} I(X_1^n; Y^{n''}|M_1) + I(M_1, M_1'; Y^{n''}|M_1, X_1^n) \tag{149}$$

$$\overset{(c)}{=} I(X_1^n; Y^{n''}|M_1) \tag{150}$$

$$\leq H(X_1^n|M_1) \tag{151}$$

$$\overset{(d)}{=} \sum_{i=1}^{n} H(X_{1,i}|M_1, X_1^{i-1}) \tag{152}$$

where (a) follows the fact that X_1^n is a deterministic function of (M_1, M_1'), (b) is due to the chain rule for mutual information, (c) follows from the Markov chain $(M_1, M_1') \multimap X_1^n \multimap Y^{n''}$ (see, Lemma 3), and (d) is due to the entropy chain rule. Next, for R_2'' we have:

$$nR_2'' - n\delta_n \leq I(M_2''; Y^{n''}|M_1, M_1') \tag{153}$$

$$= \sum_{i=1}^{n} I(M_2''; Y_i''|Y^{i-1''}, M_1, M_1') \tag{154}$$

$$\overset{(a)}{=} \sum_{i=1}^{n} I(M_2''; Y_i''|Y^{i-1''}, M_1, M_1', X_1^{i-1}, X_{1,i}) \tag{155}$$

$$\leq \sum_{i=1}^{n} I(M_1, X_1^{i-1}, Y^{i-1''}, M_1', M_2''; Y_i''|X_{1,i}) \tag{156}$$

$$\overset{(b)}{=} \sum_{i=1}^{n} I(M_1, X_1^{i-1}, Y^{i-1''}, M_1', M_2'', X_{2,i}''; Y_i''|X_{1,i}) \tag{157}$$

$$\overset{(c)}{=} \sum_{i=1}^{n} I(X_{2,i}''; Y_i''|X_{1,i}) \tag{158}$$

where (a) is due to the fact that X_1^i is a deterministic function of M_1 and M_1', (b) follows the fact that $X_{2,i}''$ is a deterministic function of (M_2'', X_1^i), and (c) follows from the Markov chain $(M_1, X_1^{i-1}, Y^{i-1''}, M_1', M_2'') \multimap (X_{1,i}, X_{2,i}'') \multimap Y_i''$. Finally, for the sum rate $R_1 + R_1' + R_2''$, we have:

$$n(R_1 + R_1' + R_2'') - n\delta_n \leq I(M_1, M_1', M_2''; Y^{n''}) \quad (159)$$

$$\leq \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}''; Y_i''). \quad (160)$$

So, hitherto we have that:

$$n(R_1 - \delta_n) \leq \sum_{i=1}^{n} I(M_1, X_1^{i-1}; Y_i | X_{2,i}) \quad (161a)$$

$$n(R_2 - \delta_n) \leq \sum_{i=1}^{n} I(X_{2,i}; Y_i | X_{1,i}) \quad (161b)$$

$$n(R_1 + R_2 - \delta_n) \leq \sum_{i=1}^{n} I(M_1, X_1^{i-1}, X_{2,i}; Y_i) \quad (161c)$$

$$n(R_1' - \delta_n) \leq \sum_{i=1}^{n} H(X_{1i} | M_1, X_1^{i-1}) \quad (161d)$$

$$n(R_2'' - \delta_n) \leq \sum_{i=1}^{n} I(X_{2,i}''; Y_i'' | M_1, X_1^{i-1}, X_{1,i}) \quad (161e)$$

$$n(R_1' + R_2'' - \delta_n) \leq \sum_{i=1}^{n} H(X_{1i} | M_1, X_1^{i-1}) + I(X_{2,i}''; Y_i'' | M_1, X_1^{i-1}, X_{1,i}) \quad (161f)$$

$$n(R_1 + R_1' + R_2'' - \delta_n) \leq \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}''; Y_i''). \quad (161g)$$

We are now in a position to define our auxiliary RV. From (161a)-(161g), letting $V_i \triangleq \left(M_1, X_1^{i-1}\right)$, and thus preserving the Markov chain induced by $\mathcal{P}$, we have that

$$n(R_1 - \delta_n) \leq \sum_{i=1}^{n} I(V_i; Y_i | X_{2,i}) \quad (162a)$$

$$n(R_2 - \delta_n) \leq \sum_{i=1}^{n} I(X_{2,i}; Y_i | X_{1,i}) \quad (162b)$$

$$n(R_1 + R_2 - \delta_n) \leq \sum_{i=1}^{n} I(V_i, X_{2,i}; Y_i) \quad (162c)$$

$$n(R_1' - \delta_n) \leq \sum_{i=1}^{n} H(X_{1i} | V_i) \quad (162d)$$

$$n(R_2'' - \delta_n) \leq \sum_{i=1}^{n} I(X_{2,i}''; Y_i'' | V_i, X_{1,i}) \quad (162e)$$

$$n(R_1' + R_2'' - \delta_n) \leq \sum_{i=1}^{n} H(X_{1i} | V_i) + I(X_{2,i}''; Y_i'' | V_i, X_{1,i}) \quad (162f)$$

$$n(R_1 + R_1' + R_2'' - \delta_n) \leq \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}''; Y_i''). \quad (162g)$$

Using the standard time-sharing argument as in [13, Ch. 14.3], one can rewrite (162) by introducing an appropriate time-sharing random variable. Therefore, if $\epsilon_n \to 0$ as $n \to \infty$, the convex hull of this region can be shown to be equivalent to the convex hull of the region in (25).

Remark 1: As was mentioned in the paragraph preceding Theorem 4, one can obtain the same outer bound also for the case of non-causal cribbing (see, (26)). Indeed, it is evident that the only places where the casual assumption play a role are in the bounds on R_2'' and $R_1 + R_1' + R_2''$. It is easy to see that the bound on $R_1 + R_1' + R_2''$ will not change, and regarding R_2'', we have (see, (158)):

$$nR_2'' - n\delta_n \leq I(M_2''; Y^{n''} | M_1, M_1') \quad (163)$$

$$= \sum_{i=1}^{n} I(M_2''; Y_i'' | Y^{i-1''}, M_1, M_1') \quad (164)$$

$$\overset{(a)}{=} \sum_{i=1}^{n} I(M_2''; Y_i'' | Y^{i-1''}, M_1, M_1', X_1^n) \quad (165)$$

$$\leq \sum_{i=1}^{n} I(M_1, X_1^{n/i}, Y^{i-1''}, M_1', M_2''; Y_i'' | X_{1,i}) \quad (166)$$

$$\overset{(b)}{=} \sum_{i=1}^{n} I(M_1, X_1^{n/i}, Y^{i-1''}, M_1', M_2'', X_{2,i}''; Y_i'' | X_{1,i}) \quad (167)$$

$$\overset{(c)}{=} \sum_{i=1}^{n} I(X_{2,i}''; Y_i'' | X_{1,i}) \quad (168)$$

where (a) is due to the fact that X_1^n is a deterministic function of M_1 and M_1', (b) follows the fact that $X_{2,i}''$ is a deterministic function of (M_2'', X_1^n), and (c) follows from the Markov chain $(M_1, X_1^{n/i}, Y^{i-1''}, M_1', M_2'') \multimap (X_{1,i}, X_{2,i}'') \multimap Y_i''$, where $X^{n/i} = (X^{i-1}, X_{i+1}^n)$.

$\square$

Appendix A
Auxiliary Markov Chains Relations

Lemma 3: The following relations hold:

1) $M_2 \multimap (M_1, X_{2,i}) \multimap Y_i$
2) $(M_2, X_2^{i-1}) \multimap \left(M_1, X_1^{i-1}, X_{2,i}\right) \multimap Y_i$
3) $Y^{i-1} \multimap \left(X_1^{i-1}, X_2^{i-1}\right) \multimap Y_i$
4) $Y^{i-1} \multimap \left(X_1^{i-1}, X_2^{i-1}, M_1, M_2\right) \multimap Y_i$
5) $Y^{i-1} \multimap \left(X_1^{i-1}, X_2^{i}, M_1, M_2\right) \multimap Y_i$
6) $(M_1, M_1') \multimap X_1^n \multimap Y^{n''}$

Proof of Lemma 3: First, recall that:

$$\left(M_1, M_2, Y^{i-1}, X_1^{i-1}, X_2^{i-1}\right) \multimap (X_{1,i}, X_{2,i}) \multimap Y_i. \quad (A.1)$$

Thus, the first item of Lemma 3 follows from:

$$P_{Y_i|M_1,X_{2,i},M_2} = \sum_{x_{1,i}} P_{Y_i|M_1,X_{2,i},M_2,X_{1,i}} \times P_{X_{1i}|M_1,X_{2,i},M_2} \quad (A.2)$$

$$= \sum_{x_{1,i}} P_{Y_i|M_1,X_{2,i},X_{1,i}} P_{X_{1i}|M_1,X_{2,i}} \quad (A.3)$$

$$= P_{Y_i|M_1,X_{2,i}}, \quad (A.4)$$

where in the second equality we have used (A.1), and the fact that X_1 is independent of M_2. The second item of Lemma 3 follows exactly in the same way as above. Indeed,

$$\begin{aligned} P_{Y_i|M_1,X_2^i,M_2,X_1^{i-1}} &= \sum_{x_{1,i}} P_{Y_i|M_1,X_2^i,M_2,X_1^i} \\ &\quad \times P_{X_{1i}|M_1,X_2^i,M_2,X_1^{i-1}} && \text{(A.5)} \\ &= \sum_{x_{1,i}} P_{Y_i|M_1,X_1^i,X_{2,i}} \\ &\quad \times P_{X_{1i}|M_1,X_1^{i-1},X_{2,i}} && \text{(A.6)} \\ &= P_{Y_i|M_1,X_1^{i-1},X_{2,i}}. && \text{(A.7)} \end{aligned}$$

Next, the third item is true because:

$$\begin{aligned} P_{Y_i|X_1^{i-1},X_2^{i-1},Y^{i-1}} &= \sum_{x_{1,i},x_{2,i}} P_{Y_i|X_1^{i-1},X_2^{i-1},Y^{i-1},X_{1,i},X_{2,i}} \\ &\quad \times P_{X_{1i},X_{2i}|X_1^{i-1},X_2^{i-1},Y^{i-1}} && \text{(A.8)} \\ &= \sum_{x_{1,i},x_{2,i}} P_{Y_i|X_1^i,X_2^i} P_{X_{1i},X_{2i}|X_1^{i-1},X_2^{i-1}} && \text{(A.9)} \\ &= \sum_{x_{1,i},x_{2,i}} P_{X_{1,i},X_{2,i},Y_i|X_1^{i-1},X_2^{i-1}} && \text{(A.10)} \\ &= P_{Y_i|X_1^{i-1},X_2^{i-1}} && \text{(A.11)} \end{aligned}$$

where the second equality follows from the fact that the channel is memoryless and the fact that there is no feedback. The forth item follows in exactly the same way. The fifth item follows from:

$$\begin{aligned} &P_{Y_i|X_1^{i-1},X_2^i,Y^{i-1},M_1,M_2} \\ &= \sum_{x_{1,i}} P_{Y_i|X_1^{i-1},X_2^i,Y^{i-1},X_{1,i},M_1,M_2} \\ &\qquad \times P_{X_{1i}|X_1^{i-1},X_2^i,Y^{i-1},M_1,M_2} && \text{(A.12)} \\ &= \sum_{x_{1,i}} P_{Y_i|X_1^i,X_2^i,M_1,M_2} P_{X_{1i}|X_1^{i-1},X_2^i,M_1,M_2} && \text{(A.13)} \\ &= \sum_{x_{1,i}} P_{X_{1,i},Y_i|X_1^{i-1},X_2^i,M_1,M_2} && \text{(A.14)} \\ &= P_{Y_i|X_1^{i-1},X_2^i,M_1,M_2} && \text{(A.15)} \end{aligned}$$

where again the second equality follows from the fact that the channel is memoryless and the fact that there is no feedback. Finally, we obtain the sixth item due to the same reasons:

$$\begin{aligned} &P_{Y^{n\prime\prime}|X_1^n,M_1,M_1'} \\ &= \sum_{x_2^{n\prime\prime}} P_{Y^{n\prime\prime}|X_1^n,X_2^{n\prime\prime},M_1,M_1'} P_{X_2^{n\prime\prime}|X_1^n,M_1,M_1'} && \text{(A.16)} \\ &= \sum_{x_2^{n\prime\prime}} P_{Y^{n\prime\prime}|X_1^n,X_2^{n\prime\prime}} P_{X_2^{n\prime\prime}|X_1^n} && \text{(A.17)} \\ &= \sum_{x_{1,i}} P_{Y^{n\prime\prime},X_2^{n\prime\prime}|X_1^n} && \text{(A.18)} \\ &= P_{Y^{n\prime\prime}|X_1^n}. && \text{(A.19)} \end{aligned}$$

□

APPENDIX B
PROOF OF LEMMA 1

Proof: In the following, we upper bound each constraint in (28), and show that that the upper bounds can be achieved by taking $V = X_1$. We have:

$$\begin{aligned} R_1 &\le I(V;Y|X_2) && \text{(B.1)} \\ &\le I(V,X_1;Y|X_2) && \text{(B.2)} \\ &= I(X_1;Y|X_2), && \text{(B.3)} \end{aligned}$$

where we have used the fact that $V \multimap (X_1,X_2) \multimap Y$. Next,

$$\begin{aligned} R_2 &\le I(X_2;Y|V) && \text{(B.4)} \\ &= H(X_2|V) - H(X_2|V,Y) && \text{(B.5)} \\ &\le H(X_2|X_1) - H(X_2|X_1,Y) && \text{(B.6)} \\ &= I(X_2;Y|X_1) && \text{(B.7)} \end{aligned}$$

where the inequality follows from the fact that X_2 is independent of (V,X_1), and the fact that:

$$\begin{aligned} H(X_2|X_1,Y) &= H(X_2|X_1,V,Y) && \text{(B.8)} \\ &\le H(X_2|V,Y) && \text{(B.9)} \end{aligned}$$

where the inequality is due to the fact that conditioning reduces entropy, and the equality follows from the relation $V \multimap (X_1,Y) \multimap X_2$. Indeed, first note that:

$$\begin{aligned} P_{X_2,V|X_1,Y} &= \frac{P_{X_1X_2Y}P_{V|X_1,X_2,Y}}{P_{X_1,Y}} && \text{(B.10)} \\ &= P_{X_2|X_1,Y}P_{V|X_1,X_2,Y} && \text{(B.11)} \\ &= P_{X_2|X_1,Y}P_{V|X_1,X_2} && \text{(B.12)} \\ &= P_{X_2|X_1,Y}P_{V|X_1} && \text{(B.13)} \\ &= P_{X_2|X_1,Y}P_{V|X_1,Y} && \text{(B.14)} \end{aligned}$$

where the third and last equalities follow from the relations $V \multimap (X_1,X_2) \multimap Y$ and $V \multimap X_1 \multimap Y$, respectively, which are true due to (21). For the sum rate, we have:

$$\begin{aligned} R_1 + R_2 &\le I(V,X_2;Y) && \text{(B.15)} \\ &\le I(V,X_1,X_2;Y) && \text{(B.16)} \\ &= I(X_1,X_2;Y) && \text{(B.17)} \end{aligned}$$

in which the last equality follow from $V \multimap (X_1,X_2) \multimap Y$. Similarly, for R_2'', we obtain:

$$\begin{aligned} R_2'' &\le I(X_2'';Y''|X_1,V) && \text{(B.18)} \\ &= H(Y''|X_1,V) - H(Y''|X_1,V,X_2'') && \text{(B.19)} \\ &\le H(Y''|X_1) - H(Y''|X_1,X_2'') && \text{(B.20)} \\ &= I(X_2'';Y''|X_1) && \text{(B.21)} \end{aligned}$$

where the inequality follows from the fact that conditioning reduces entropy, and the relation $V \multimap (X_1,X_2'') \multimap Y''$. Finally, the result follows by noticing that the obtained upper bounds in (B.3), (B.7), (B.17), and (B.21) are independent of V, and can be achieved by taking $V = X$. □

REFERENCES

[1] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. London, UK.: Academic Press, 1981.

[2] R. Dabora and S. Servetto, "Broadcast channels with cooperating receivers: a downlink for sensor reachback problem," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, June 27-July 2, 2004, p. 176.

[3] R. Dabora and S. Servetto, "Broadcast channels with cooperating decoders," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5438-5454, December 2006.

[4] L. Dikstein, H. Permuter, and Y. Steinberg, "The state-dependent broadcast channel with cooperation," in *Proc. of the 51st Annual Allerton Conference on Communication, Control, and Computing*, Allerton House, Monticello, Illinois, October 2–4, 2013.

[5] R. Karasik, O. Simeone, and S. Shamai, "Robust uplink communications over fading channels with variable backhaul connectivity," *IEEE Trans.Wireless Commun.*, vol. 12, no. 11, pp. 5788-5799, Nov. 2013.

[6] Y. Liang and V. V. Veeravalli, "The impact of relaying on the capacity of broadcast channels," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, June 27-July 2, 2004, p. 403.

[7] Y. Liang and V. V. Veeravalli, "Cooperative relay broadcast channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 900-1028, March 2007.

[8] O. Simeone, N. Levy, A. Sanderovich, O. Somekh, B. Zaidel, H. Poor, and S. Shamai, "Cooperative wireless cellular systems: an information theoretic view," *Foundations Trends Commun. Inf. Theory*, vol. 8, no. 1-2, pp.1-177, 2012.

[9] F.M.J. Willems and E. C. Van Der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 313-327, May 1985.

[10] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572584, Sep. 1979.

[11] A. E. Gamal, N. Hassanpour, and J. P. Mammen, "Relay networks with delays," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 341334317, Oct. 2007.

[12] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[13] T. M. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.