

Channels with Cooperation Links that May Be Absent

Wasim Huleihel* and Yossef Steinberg†

*Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA
wasimh@mit.edu

†Andrew & Erna Viterbi Faculty of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel
ysteinbe@ee.technion.ac.il

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. Unfortunately, in many modern communication networks the availability of cooperation links cannot be guaranteed a priori, due to the dynamic nature of the network. 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, as they depend on parameters that the system designer does not have any control on. For example, the cooperation can depend on the battery status of intermediate users (relays), on weather, or just on the willingness of peers in the network to help. A typical situation, therefore, is that the users are aware of the *possibility* that cooperation will take place, but it cannot be assured before transmission begins. Moreover, in many instances it is not possible to inform the transmitter whether or not a potential relay/helper decides to help. Thus, the traditional approach leaves the designer with two design options. Option 1 is the pessimistic one: assume that none of

the unreliable relays exists, and design a system without cooperation. Option 2 takes the optimistic view: assume that the potential relays exist, and design a system with cooperation. The pros and cons of each of the designs are clear. Option 1 is “safe,” but results in relatively low rates. Option 2 aims to transmit at the maximal rates, but runs the risk that some of the relays/helpers are absent, in which case the coding scheme collapses.

In this work we suggest a third option: design a robust system, which takes advantage of the links if they are present, but can operate also if they are absent, although possibly at reduced rates. This design problem becomes simple if all the users in the system can be informed about the situation of the helpers before transmission begins. We study models in which, at least for part of the users, this information is not available before transmission. In general, this set of problems can be viewed as channel coding counterparts of well known problems in source coding, like multiple descriptions [4], [5], successive refinement [6], [7], and rate-distortion when side information may be absent [9]. We focus on two channel models - 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 [11], [12], 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 [11] is replaced with a constant rate bit pipe.

The MAC with cribbing encoders was introduced by

This paper was presented in part at the 2014 and 2016 IEEE International Symposium on Information Theory. This work was supported by the ISRAEL SCIENCE FOUNDATION (ISF) (grant no. 684/11).

Willems and Van Der Meulen in [14]. 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 [14] all consistent scenarios of cribbing (strictly causal, causal, non-causal, and symmetric or asymmetric), and characterized the capacity region of these models. Another relevant recent work is [17], where the MAC with partial cribbing encoders was considered, motivated by the additive noise Gaussian MAC model, where perfect cribbing means full cooperation between the encoders and requires an infinite entropy link. Finally, we mention [18], which considers the MAC channel with state and cribbing encoders. Accordingly, the state can be specialized to capture the availability of the cribbing links, which would lead to a setup similar to the one considered in our paper. Nonetheless, different from our paper, this state information is assumed to be causally known at the cribbed encoder and not to the cribbing encoder.

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 case where a constraint on the rates of the users is added. For the physically degraded BC, the results are conclusive. The results derived here were partially presented in [15], [16].

It should be noted that multi-user communication systems with uncertainty in part of the network links have been studied in the literature - see, e.g., [13] and [10], 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 II, we establish our notation. The physically degraded BC with unreliably cooperating decoders is presented and studied in Section III. In Section IV, we consider the MAC with cribbing encoders where the cribbing link may be absent. The proofs are provided in Section V.

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 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., \mathbf{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 $X \circlearrowleft Y_1 \circlearrowleft Y_2$ holds, i.e.,

$$P_{X, Y_1, Y_2} = P_X P_{Y_1, Y_2|X} = P_X P_{Y_1|X} P_{Y_2|Y_1}. \quad (1)$$

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

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

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, $\mathbf{x}(M_1, M_2, M'_2)$. 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.

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

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

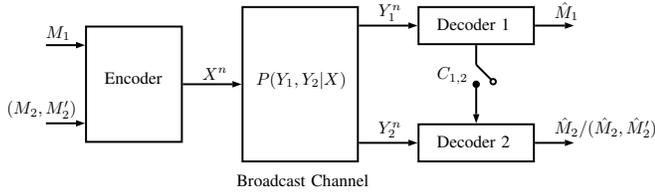


Fig. 1. Broadcast channel with unreliable cooperating decoders.

a conference mapping

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

and three decoding maps:

$$g_1 : \mathcal{Y}_1^n \rightarrow \mathcal{N}_1, \quad (3a)$$

$$g_2 : \mathcal{Y}_2^n \rightarrow \mathcal{N}'_2, \quad (3b)$$

$$g'_2 : \mathcal{Y}_2^n \times \mathcal{N}_{1,2} \rightarrow \mathcal{N}'_2, \quad (3c)$$

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

$$P_e = \frac{1}{\nu_1 \nu_2 \nu'_2} \sum_{m_1, m_2, m'_2} P_{Y_1, Y_2|X}(S_e | f(m_1, m_2, m'_2)) \quad (4a)$$

$$P'_e = \frac{1}{\nu_1 \nu_2 \nu'_2} \sum_{m_1, m_2, m'_2} P_{Y_1, Y_2|X}(S'_e | f(m_1, m_2, m'_2)) \quad (4b)$$

where the sets S_e and S'_e are defined as

$$\begin{aligned} S_e(m_1, m_2) &= \{(\mathbf{y}_1, \mathbf{y}_2) : g_1(\mathbf{y}_1) \neq m_1 \text{ or } g_2(\mathbf{y}_2) \neq m_2\}, \\ S'_e(m_1, m_2, m'_2) &= S_e(m_1, m_2) \\ &\cup \{(\mathbf{y}_1, \mathbf{y}_2) : g'_2(\mathbf{y}_2, h(\mathbf{y}_1)) \neq m'_2\}, \end{aligned} \quad (5)$$

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:

$$C_{1,2} = \frac{\log \nu_{1,2}}{n}, \quad R_k = \frac{\log \nu_k}{n}, \quad k = 1, 2, \quad R'_2 = \frac{\log \nu'_2}{n}.$$

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:

$$R_1 \leq I(X; Y_1|U, V), \quad (6a)$$

$$R_2 \leq I(U; Y_2), \quad (6b)$$

$$R'_2 \leq \min \{I(V; Y_2|U) + C_{1,2}, I(V; Y_1|U)\}, \quad (6c)$$

for some joint distribution of the form

$$P_{U,V,X,Y_1,Y_2} = P_{U,V} P_{X|U,V} P_{Y_1,Y_2|X}, \quad (6d)$$

where $|U| \leq |\mathcal{X}| + 3$, 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 make the following observations:

- The direct part in the proof of Theorem 1 is based on a combination of superposition coding and binning. The intuitive explanation/interpretation of the various auxiliary random variables in (6) is as follows. First, the information of User 2 is encoded with U , which depends on the message M_2 . This information is always decoded by both decoders, whether the conference link is present or not. The extra message of User 2, which is M'_2 is encoded with V , which is superimposed on top of U . Finally, the message of User 1, namely M_1 is encoded with X , which is again superimposed on top of U and V . The information encoded with U , V , and X , are always decoded by the first decoder, whether the conference link is present or not. The extra information encoded with V , however, is decoded (with the help of the conference link) by the second decoder only if the conference link is present. This is done by using the binning approach.

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

$$R_1 \leq I(X; Y_1|U, V). \quad (7a)$$

$$R_2 \leq I(U; Y_2), \quad (7b)$$

$$R'_2 \leq I(V; Y_2|U), \quad (7c)$$

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:

$$R_1 \leq I(X; Y_1|U, V), \quad (8a)$$

$$R'_2 + R_2 \leq I(U, V; Y_2), \quad (8b)$$

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

$$R_1 \leq I(X; Y_1|V), \quad (9a)$$

$$R'_2 \leq \min\{I(V; Y_2) + C_{1,2}, I(V; Y_1)\}, \quad (9b)$$

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 $\mathbb{E}[\sum_{i=1}^n \Gamma(X_i)] \leq nP$. In this case the achievable region is given by Theorem 1 where the additional constraint $\mathbb{E}[\Gamma(X_1)] \leq P$ is needed. Note that the achievability and the converse proofs of Theorem 1 with the input constraint remain the same, where in the achievability part, we make use of the fact that by the law of large numbers the constraints are satisfied with high probability.

• It is interesting to check what happens in case that the rate to User 2 is smaller than the capacity of the cooperation link, namely, $R'_2 \leq C_{1,2}$. When the cooperation link is reliable (i.e., always available), which is the model considered in [3, Theorem 1], it can be shown that the capacity region is the convex hull of all rate pairs (R_1, R'_2) such that

$$R_1 + R'_2 \leq I(X; Y_1), \quad (10a)$$

$$R'_2 \leq C_{1,2}. \quad (10b)$$

for some joint distribution $P_{X,Y_1} = P_X P_{Y_1|X}$. This result is indeed reasonable due to the fact that in this case User 1 can transmit all the information through the cooperation link. To show (10), first note that the intersection between (9) and $R'_2 \leq C_{1,2}$ gives

$$R_1 \leq I(X; Y_1|V), \quad (11a)$$

$$R'_2 \leq \min\{C_{1,2}, I(V; Y_1)\}, \quad (11b)$$

for some joint distribution $P_{V,X,Y_1} = P_V P_{X|V} P_{Y_1|X}$. Let \mathcal{A} and \mathcal{B} denote the regions in (10) and (11), respectively. Then, it is evident that $\mathcal{B} \subseteq \mathcal{A}$ due to the Markov chain $V \leftrightarrow X \leftrightarrow Y_1$. We now proceed to show the reverse inclusion, i.e., $\mathcal{A} \subseteq \mathcal{B}$. To this end, let $(R_1, R'_2) \in \mathcal{A}$, achieved by some X . We consider two cases: if $R_1 = I(X; Y_1)$, then by using (10a), we get $R'_2 = 0$. However, from (11a) we see that $R_1 = I(X; Y_1)$ if and only if $V = \emptyset$, from which we also get that $R'_2 = 0$. Thus, $(R_1, R'_2) \in \mathcal{B}$. If, however, $R_1 < I(X; Y_1)$, then let $R_1 = I(X; Y_1) - \alpha$, for $\alpha > 0$. We define

$$V \triangleq \begin{cases} \emptyset, & \text{w.p. } \beta \\ X, & \text{w.p. } 1 - \beta \end{cases}, \quad (12)$$

for some $\beta \in [0, 1)$. Obviously, we have the Markov chain $V \leftrightarrow X \leftrightarrow Y_1$, and it is easy to see that

$$I(X; Y_1|V) = \beta \cdot I(X; Y_1). \quad (13)$$

Now, since β is arbitrary, we can choose as

$$\beta = \frac{I(X; Y_1) - \alpha}{I(X; Y_1)}, \quad (14)$$

and we readily get that

$$R_1 = I(X; Y_1) - \alpha = I(X; Y_1|V), \quad (15)$$

and

$$R_1 + R'_2 \leq I(X; Y_1) = I(V; Y_1). \quad (16)$$

Combining (10b), (15), and (16), we have that (R_1, R_2) satisfy

$$R_1 \leq I(X; Y_1|V), \quad (17a)$$

$$R'_2 \leq \min\{C_{1,2}, I(V; Y_1)\}, \quad (17b)$$

for $V \leftrightarrow X \leftrightarrow Y_1$, which implies that $(R_1, R'_2) \in \mathcal{B}$.

When the cooperation link is unreliable, however, using the same arguments as above, it can be shown that the capacity region when $R'_2 \leq C_{1,2}$ is the convex hull of all rate triples (R_1, R_2, R'_2) that satisfy

$$R_1 + R'_2 \leq I(X; Y_1|U), \quad (18a)$$

$$R_2 \leq I(U; Y_2), \quad (18b)$$

$$R'_2 \leq C_{1,2}, \quad (18c)$$

for some distribution $P_{U,X,Y_1,Y_2} = P_U P_{X|U} P_{Y_1,Y_2|X}$. This result makes sense because of the fact that when the cooperation link is absent, we still would like to transmit some information to the User 2, which is captured by U .

To illustrate the general result in Theorem 1, we consider the following simple example.

Example 1: Consider the example where the channel output Y_1 is clean, namely, $Y_1 = X \in \{0, 1\}$, and Y_2 is the output of a binary symmetric channel, i.e., $Y_2 = X \oplus Z$, where Z is Bernoulli with $\Pr\{Z = 0\} = p$ and statistically independent of X . In this case, we obtain from Theorem 1 that the capacity region is:

$$R_1 \leq H(X|U, V), \quad (19a)$$

$$R_2 \leq I(U; Y_2), \quad (19b)$$

$$R'_2 \leq \min\{I(V; Y_2|U) + C_{1,2}, I(V; X|U)\}. \quad (19c)$$

Fig. 2 depicts the capacity region in (19), assuming that $C_{1,2} = 0.5$, for several values of R_2 . We present four curves corresponding to the capacity region of the standard BC without cooperation (black curve), the rates (R_1, R_2) which refer to the case where (unreliable) conferencing/cooperation is absent (blue dashed curve), the rates $(R_1, R_2 + R'_2)$ which refer to the case where (unreliable) conferencing/cooperation is present (red dotted curve), and the capacity region in case of reliable/perfect cooperation [3] (green dashed-dotted curve), i.e., regular cooperation with reliable link. It can be seen that (total) higher rates can be achieved in case of unreliable cooperation compared to the case where there is no cooperation at all, as expected. Also, comparing the $(R_1, R_2 + R'_2)$ curve and the reliable cooperation curve, it can be noticed that there is some degradation due to the fact that the cooperation link is unreliable. Finally, from the (R_1, R_2) and the standard BC curves it can be seen that there is some price in terms of the rate R_2 (namely, when there is no cooperation) due

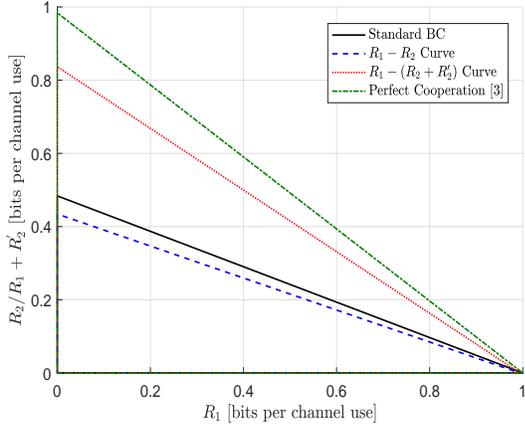


Fig. 2. The capacity region of the BC in Example 1 with $C_{1,2} = 0.5$, compared to the capacity region of the BC with reliable cooperation [3].

to the universality of the coding scheme in case of unreliable cribbing.

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 $\mathcal{N}'_1 = \{1, 2, \dots, \nu'_1\}$ and $\mathcal{N}''_2 = \{1, 2, \dots, \nu''_2\}$ 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 $\mathbf{x}_1 = \mathbf{x}_1(M_1, M'_1)$. If the cribbing link is absent, Encoder 2 maps the message M_2 to an input sequence $\mathbf{x}_2 = \mathbf{x}_2(M_2)$. If the cribbing link is present, Encoder 2 knows \mathbf{x}_1 strictly causally, thus maps the pair (M''_2, \mathbf{x}_1) to an input sequence \mathbf{x}_2'' , in a strictly causal manner:

$$\mathbf{x}_2''(m''_2, \mathbf{x}_1) = (x_{2,1}''(m''_2), x_{2,2}''(m''_2, x_{1,1}), \dots, x_{2,n}''(m''_2, x_{1,1}^{n-1})). \quad (20)$$

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

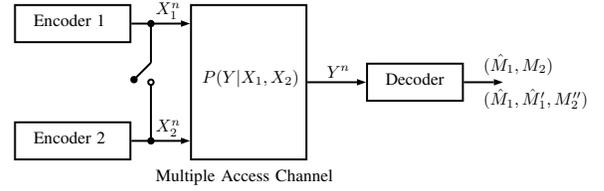


Fig. 3. MAC with unreliable cribbing encoders.

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. 3.

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 of $n + 2$ encoding maps

$$f_1 : \mathcal{N}_1 \times \mathcal{N}'_1 \rightarrow \mathcal{X}_1^n, \quad (21a)$$

$$f_2 : \mathcal{N}_2 \rightarrow \mathcal{X}_2^n, \quad (21b)$$

$$f''_{2,i} : \mathcal{N}''_2 \times \mathcal{X}_1^{i-1} \rightarrow \mathcal{X}_2^i, \quad i = 1, 2, \dots, n, \quad (21c)$$

and a pair of decoding maps

$$g : \mathcal{Y}^n \rightarrow \mathcal{N}_1 \times \mathcal{N}_2, \quad (22a)$$

$$g' : \mathcal{Y}^n \rightarrow \mathcal{N}_1 \times \mathcal{N}'_1 \times \mathcal{N}''_2, \quad (22b)$$

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

$$P_e = \frac{1}{\nu_1 \nu'_1 \nu_2} \sum_{m_1, m'_1, m_2} P_{Y|X_1, X_2}(\mathcal{Q}_e | f_1(m_1, m'_1), f_2(m_2)) \quad (23a)$$

$$P'_e = \frac{1}{\nu_1 \nu'_1 \nu''_2} \sum_{m_1, m'_1, m''_2} P_{Y|X_1, X_2}(\mathcal{Q}'_e | f_1(m_1, m'_1), \mathbf{f}''_2(m''_2, f_1(m_1, m'_1))) \quad (23b)$$

where $\mathbf{f}''_2(m''_2, f_1(m_1, m'_1))$ is the sequence of maps $f''_{2,i}$ in (21c), the sets \mathcal{Q}_e and \mathcal{Q}'_e are defined as

$$\mathcal{Q}_e(m_1, m_2) = \{\mathbf{y} : g(\mathbf{y}) \neq (m_1, m_2)\}, \quad (24a)$$

$$\mathcal{Q}'_e(m_1, m'_1, m''_2) = \{\mathbf{y} : g'(\mathbf{y}) \neq (m_1, m'_1, m''_2)\}, \quad (24b)$$

and the dependence of the sets $\mathcal{Q}_e, \mathcal{Q}'_e$ on the messages is dropped in (23), 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

$$P_U P_V P_{X_1|U, V} P_{X_2} P_{Y|X_1, X_2} P_{X'_2|U} P_{Y'|X_1, X'_2} \quad (25)$$

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 convex hull of all rate quadruples (R_1, R_1', R_2, R_2') satisfying

$$R_1 \leq I(V; Y|X_2), \quad (26a)$$

$$R_2 \leq I(X_2; Y|V), \quad (26b)$$

$$R_1 + R_2 \leq I(V, X_2; Y), \quad (26c)$$

$$R_1' \leq H(X_1|U, V), \quad (26d)$$

$$R_2'' \leq I(X_2''; Y''|U, V, X_1), \quad (26e)$$

$$R_1' + R_2'' \leq I(X_1, X_2''; Y''|V), \quad (26f)$$

$$R_1 + R_1' + R_2'' \leq I(X_1, X_2''; Y''), \quad (26g)$$

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

$$|\mathcal{U}| \leq \min \{|\mathcal{X}_1| \cdot |\mathcal{X}_2| + 1, |\mathcal{Y}| + 2\} \quad (27)$$

$$|\mathcal{V}| \leq \min \{|\mathcal{X}_1| \cdot |\mathcal{X}_2| + 4, |\mathcal{Y}| + 5\}. \quad (28)$$

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,

$$\mathbf{x}_2''(m_2'', \mathbf{x}_1) = (x_{2,1}''(m_2'', x_{1,1}), \dots, x_{2,n}''(m_2'', x_{1,n})), \quad (29)$$

or, equivalently, replace (21c) with:

$$f_{2,i}'' : \mathcal{N}_2'' \times \mathcal{X}_1^i \rightarrow \mathcal{X}_{2,i}, \quad i = 1, 2, \dots, n. \quad (30)$$

The capacity \mathcal{C}_{mac} of the MAC with unreliable causal cribbing is defined similarly to the strictly causal case, but with (29) and (30), replacing (20) and (21c), respectively.

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

$$P_{V, X_1} P_{X_2} P_{Y|X_1, X_2} P_{X_2''|X_1} P_{Y''|X_1, X_2''}. \quad (31)$$

The interpretation of the coding random variables and their 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}_{mac} be the convex hull of all rate quadruples (R_1, R_1', R_2, R_2') satisfying

$$R_1 \leq I(V; Y|X_2), \quad (32a)$$

$$R_2 \leq I(X_2; Y|V), \quad (32b)$$

$$R_1 + R_2 \leq I(V, X_2; Y), \quad (32c)$$

$$R_1' \leq H(X_1|V), \quad (32d)$$

$$R_2'' \leq I(X_2''; Y''|V, X_1), \quad (32e)$$

$$R_1' + R_2'' \leq I(X_1, X_2''; Y''|V), \quad (32f)$$

$$R_1 + R_1' + R_2'' \leq I(X_1, X_2''; Y''), \quad (32g)$$

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

$$|\mathcal{V}| \leq \min \{|\mathcal{X}_1| \cdot |\mathcal{X}_2| + 4, |\mathcal{Y}| + 5\}. \quad (33)$$

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 the above results.

- The bounds on the cardinalities of U , and V , are derived in a similar manner as in [14, Appendix B], and is based on Fenchel-Eggleston-Carathéodry 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 is 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 [14] for one sided causal cribbing. In this case, the decoder decodes the messages of V, U, X_1 , and X_2'' . Finally, to prove Theorem 3 we employ Shannon strategies.

- 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}_{mac} yields the capacity region of the MAC without cribbing, as expected.

- The r.h.s. of (26e) is smaller than that of (32e). Indeed,

$$\begin{aligned} I(X_2''; Y''|U, V, X_1) &= H(Y''|U, V, X_1) \\ &\quad - H(Y''|U, V, X_1, X_2'') \\ &\leq H(Y''|V, X_1) - H(Y''|V, X_1, X_2'') \\ &= I(X_2''; Y''|V, X_1) \end{aligned} \quad (34)$$

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

Unfortunately, we were not able to show the converse part in general, but only for some special case, described in the sequel. In the following, we provide first 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

$$R_1 \leq I(V; Y|X_2), \quad (35a)$$

$$R_2 \leq I(X_2; Y|X_1), \quad (35b)$$

$$R_1 + R_2 \leq I(V, X_2; Y), \quad (35c)$$

$$R'_1 \leq H(X_1|V), \quad (35d)$$

$$R''_2 \leq I(X''_2; Y''|V, X_1), \quad (35e)$$

$$R_1 + R'_1 + R''_2 \leq I(X_1, X''_2; Y''), \quad (35f)$$

for some $P_{V, X_1, X_2, X''_2, Y, Y''} \in \mathcal{P}^O$ of the form

$$P_{V, X_1} P_{X_2} P_{Y|X_1, X_2} P_{X''_2|V, X_1} P_{Y''|X_1, X''_2}. \quad (36)$$

The forthcoming result is true also for the non-causal cribbing case, namely,

$$\mathbf{x}''_2(m''_2, \mathbf{x}_1) = (x''_{2,1}(m''_2, x_1^n), \dots, x''_{2,n}(m''_2, x_1^n)), \quad (37)$$

or, equivalently, replace (21c) with:

$$f''_{2,i} : \mathcal{N}''_2 \times \mathcal{X}_1^n \rightarrow \mathcal{X}_{2,i}, \quad i = 1, 2, \dots, n. \quad (38)$$

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.

Assume that $R'_1 = 0$, 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:

$$R_1 \leq I(V; Y|X_2), \quad (39a)$$

$$R_2 \leq I(X_2; Y|V), \quad (39b)$$

$$R_1 + R_2 \leq I(V, X_2; Y), \quad (39c)$$

$$R''_2 \leq I(X''_2; Y''|V, X_1), \quad (39d)$$

$$R_1 + R''_2 \leq I(X_1, X''_2; Y''), \quad (39e)$$

for some $P_{V, X_1, X_2, X''_2, Y, Y''} \in \mathcal{P}$ of the form

$$P_{V, X_1} P_{X_2} P_{Y|X_1, X_2} P_{X''_2|X_1} P_{Y''|X_1, X''_2}. \quad (40)$$

Let $\hat{\mathcal{I}}_{\text{mac}}^I$ be the convex hull of all rate triples (R_1, R_2, R''_2) satisfying (39) and (40). Next, let $\hat{\mathcal{I}}_{\text{mac}}^O$ be the convex hull of all rate triples (R_1, R_2, R''_2) satisfying:

$$R_1 \leq I(V; Y|X_2), \quad (41a)$$

$$R_2 \leq I(X_2; Y|X_1), \quad (41b)$$

$$R_1 + R_2 \leq I(V, X_2; Y), \quad (41c)$$

$$R''_2 \leq I(X''_2; Y''|V, X_1), \quad (41d)$$

$$R_1 + R''_2 \leq I(X_1, X''_2; Y''), \quad (41e)$$

for some $P_{V, X_1, X_2, X''_2, Y, Y''} \in \mathcal{P}^O$ in (36). It is easy to see that $\hat{\mathcal{I}}_{\text{mac}}^O$ is obtained upon substitution of $R'_1 = 0$ in (35), and

thus according to Theorem 4, it is an outer bound on (39). 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 $\hat{\mathcal{I}}_{\text{mac}}$ be the convex hull of all rate triples (R_1, R_2, R''_2) satisfying:

$$R_1 \leq I(X_1; Y|X_2), \quad (42a)$$

$$R_2 \leq I(X_2; Y|X_1), \quad (42b)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y), \quad (42c)$$

$$R''_2 \leq I(X''_2; Y''|X_1), \quad (42d)$$

$$R_1 + R''_2 \leq I(X_1, X''_2; Y''), \quad (42e)$$

for some $P_{X_1, X_2, X''_2, Y, Y''}$ of the form

$$P_{X_1} P_{X_2} P_{Y|X_1, X_2} P_{X''_2|X_1} P_{Y''|X_1, X''_2}. \quad (43)$$

The subsequent lemma is proved in Appendix B.

Lemma 1: Let $\hat{\mathcal{I}}_{\text{mac}}^I$, $\hat{\mathcal{I}}_{\text{mac}}^O$, and $\hat{\mathcal{I}}_{\text{mac}}$, be defined in (39), (41), and (42), respectively. Then,

$$\hat{\mathcal{I}}_{\text{mac}}^I = \hat{\mathcal{I}}_{\text{mac}}^O = \hat{\mathcal{I}}_{\text{mac}}. \quad (44)$$

Hence, using Lemma 1, we obtain the following result.

Theorem 5: For any MAC with unreliable causal (non-causal) cribbing, if $R'_1 = 0$, then $\hat{\mathcal{I}}_{\text{mac}}$ is the capacity region.

According to (42), 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 (42d) and on the sum rate (42e)). Instead, as can be seen from (42d)-(42e), 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).

An interesting conclusion arises from the region in (42)-(43). Note that (42a)-(42c), when evaluated over all product distributions $P_{X_1} P_{X_2}$ (as in (43)), coincides with the capacity region of the standard MAC, without cribbing. Therefore, for the case of $R'_1 = 0$, there is no loss of performance when using a robust coding scheme, relative to the case of no cribbing at all. To illustrate the results in Theorems 3 and 5, and the above conclusion, we consider the following examples.

Example 2: Consider the channel model depicted in Fig. 4, where the channel output is $Y = (Y_1, Y_2)$, $\rho_0 \triangleq \Pr\{Y_2 = 1|X_2 = 0\} = \Pr\{Y_2 = 0|X_2 = 1\}$ and $\rho_1 \triangleq \Pr\{Y_2 = 1|X_2 = 2\} = \Pr\{Y_2 = 0|X_2 = 3\}$. The crossover probabilities ρ_0 and ρ_1 depend on X_1 in the following way: if $X_1 = 0$, then $(\rho_0, \rho_1) = (0, 1/2)$, and if $X_1 = 1$, then $(\rho_0, \rho_1) = (1/2, 0)$. Accordingly, if cribbing is present, then User 2 can transmit his information via a noiseless channel. In this case, the total rate that User 2 can transmit is 1, which is the maximal possible rate for him since the output Y_2 is binary. When cribbing is absent, User 2 cannot know which of the channels is clean, and thus cannot transmit at the maximal

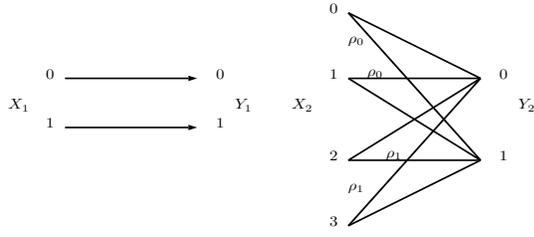


Fig. 4. The channel model in Example 1.

rate 1. Let $P_{X_1} \triangleq \Pr\{X_1 = 0\}$, and $p_i \triangleq \Pr\{X_2 = i\}$, for $i = 0, 1, 2, 3$. Using (42), it is a simple exercise to check that

$$R_1 \leq \mathcal{H}_2(P_{X_1}), \quad (45a)$$

$$R_2 \leq P_{X_1} \mathcal{H}_2(p_0 + \frac{p_2 + p_3}{2}) + \bar{P}_{X_1} \mathcal{H}_2(p_2 + \frac{p_0 + p_1}{2}) - P_{X_1} \cdot (p_2 + p_3) - \bar{P}_{X_1} \cdot (p_0 + p_1), \quad (45b)$$

$$R_2'' \leq 1, \quad (45c)$$

where $\mathcal{H}_2(\cdot)$ is the binary entropy, and note that in this example the sum-rate constraints in (42c) and (42e), are redundant because they are given by the sum of the individual rate constraints. Also, note that the optimal distribution $P_{X_2''|X_1}$ in this case, is given by

$$P_{X_2''|X_1=0} = \begin{cases} 0.5, & \text{if } X_2'' = 0 \\ 0.5, & \text{if } X_2'' = 1 \\ 0, & \text{if } X_2'' = 2 \\ 0, & \text{if } X_2'' = 3 \end{cases}, \quad (46)$$

and

$$P_{X_2''|X_1=1} = \begin{cases} 0, & \text{if } X_2'' = 0 \\ 0, & \text{if } X_2'' = 1 \\ 0.5, & \text{if } X_2'' = 2 \\ 0.5, & \text{if } X_2'' = 3 \end{cases}. \quad (47)$$

Fig. 5 presents three curves corresponding to the capacity region of the standard MAC without cribbing (green “+” curve), the rates (R_1, R_2) which refer to the case where cribbing is absent (blue dotted curve), and the rates (R_1, R_2'') which refer to the case where cribbing is present (red curve). Each value of R_1 is associated with two rates R_2 and R_2'' . For example, if $R_1 = 0.7$ then $R_2 \approx 0.8$, $R_2'' = 1$. It is evident that higher rates can be achieved for the second user due to the cribbing, as expected. Also, it can be seen that the (R_1, R_2) curve coincide with the capacity region of the standard MAC without cribbing, as expected from (42). This means that the coding scheme for the case of unreliable cribbing is robust for the case of $R_1' = 0$. That is, when $R_1' = 0$, the uncertainty about the cribbing link does not have negative influence on the performance compared to the case of no cribbing at all. Finally, we mention that in this example, the capacity region in case of reliable cribbing [14] coincides with the (R_1, R_2'') curve.

Example 3: Consider the model in the previous example, but now $Y_1 = X_1 \oplus Z$, where Z is a Bernoulli RV, independent of

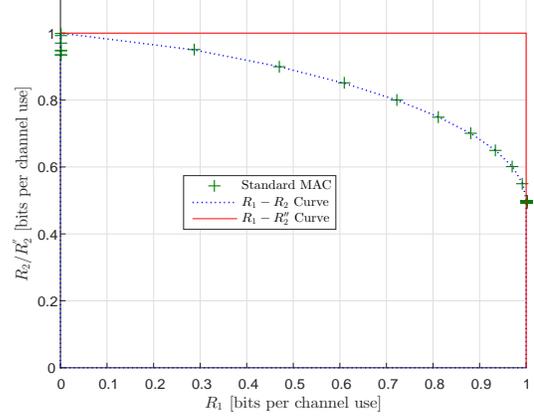


Fig. 5. The capacity region of Example 2 ((45a)-(45c)), compared to the capacity region of the standard MAC without cribbing. Each value of R_1 is associated with two rates R_2 and R_2'' . The capacity region in case of reliable cribbing coincides with the (R_1, R_2'') curve.

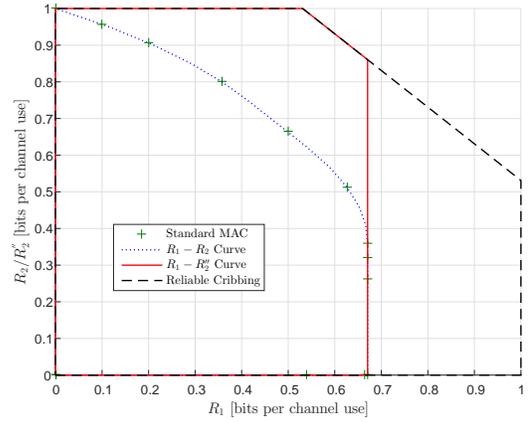


Fig. 6. The capacity region in Example 3, compared to the capacity region of the standard MAC without cribbing, and the capacity region in case of reliable cribbing, for $q = 0.1$. Each value of R_1 is associated with two rates R_2 and R_2'' .

X_1 , and $\Pr\{Z = 0\} = q$. The capacity region in this case is given in Appendix C (see (C.13)). Fig. 6 presents four curves corresponding to the capacity region of the standard MAC without cribbing (green “+” curve), the rates (R_1, R_2) (blue dotted curve), the rates (R_1, R_2'') (red curve), and the capacity region in case of reliable cribbing [14] (black dashed curve). In the simulations we choose $q = 0.1$. As before, higher rates can be achieved for the second user due to the cribbing, as expected, and it can be seen that the (R_1, R_2) curve coincide with the capacity region of the standard MAC. Finally, contrary to the previous example, here, there is some degradation compared to the reliable cribbing case.

Example 4: Consider the example where the channel output, Y , is given by:

$$Y = X_1 \oplus X_2 \oplus Z_1 \oplus Z_2 \quad (48)$$

where X_1, X_2, Z_1 , and Z_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:

$$Y'' = X_1 \oplus X_2'' \oplus Z_1 \oplus Z_2 \quad (49)$$

where now X_2'' may depend on X_1 . 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:

$$\alpha \triangleq (p_1 \star p_2) \cdot P_{X_2} + (p_1 * p_2) \cdot \bar{P}_{X_2}, \quad (50)$$

$$\beta \triangleq (p_1 \star p_2) \cdot \mu_2 + (p_1 * p_2) \cdot \bar{\mu}_2. \quad (51)$$

We wish to evaluate the region in (32) where R_1' may be positive. To this end, we choose the auxiliary RV V to be:

$$V \triangleq \begin{cases} \emptyset, & \text{w.p. } \gamma \\ X_1, & \text{w.p. } 1 - \gamma \end{cases}, \quad (52)$$

which may be sub-optimal. We define the following quantities:

$$\begin{aligned} I_1 &\triangleq \mathcal{H}_2(P_{X_1} p_1 + \bar{P}_{X_1} (p_1 \star p_2)) - P_{X_1} \mathcal{H}_2(p_1) \\ &\quad - \bar{P}_{X_1} \mathcal{H}_2(p_1 * p_2), \\ I_2 &\triangleq \mathcal{H}_2(P_{X_1} (p_1 \star P_{X_2}) + \bar{P}_{X_1} \bar{\alpha}) - \\ &\quad - \mathcal{H}_2(P_{X_1} p_1 + \bar{P}_{X_1} (p_1 \star p_2)), \\ I_3 &\triangleq P_{X_1} \mathcal{H}_2(p_1 \star P_{X_2}) + \bar{P}_{X_1} \mathcal{H}_2(\alpha) - P_{X_1} \mathcal{H}_2(p_1) \\ &\quad - \bar{P}_{X_1} \mathcal{H}_2(p_1 * p_2), \\ I_4 &\triangleq \mathcal{H}_2(P_{X_1} (p_1 \star P_{X_2}) + \bar{P}_{X_1} \bar{\alpha}) - P_{X_1} \mathcal{H}_2(p_1) \\ &\quad - \bar{P}_{X_1} \mathcal{H}_2(p_1 * p_2), \\ I_5 &\triangleq P_{X_1} \mathcal{H}_2(p_1 \star \mu_1) + \bar{P}_{X_1} \mathcal{H}_2(\beta) - P_{X_1} \mathcal{H}_2(p_1) \\ &\quad - \bar{P}_{X_1} \mathcal{H}_2(p_1 * p_2), \\ I_6 &\triangleq \mathcal{H}_2(P_{X_1} (p_1 \star \mu_1) + \bar{P}_{X_1} \bar{\beta}) - P_{X_1} \mathcal{H}_2(p_1) \\ &\quad - \bar{P}_{X_1} \mathcal{H}_2(p_1 * p_2). \end{aligned} \quad (53)$$

Then, using the above definitions, it is a simple exercise to check that (32) boils down to:

$$R_1 \leq (1 - \gamma) \cdot I_1, \quad (54a)$$

$$R_2 \leq \gamma \cdot I_2 + (1 - \gamma) \cdot I_3, \quad (54b)$$

$$R_1 + R_2 \leq \gamma \cdot I_2 + (1 - \gamma) \cdot I_4, \quad (54c)$$

$$R_1' \leq \gamma \cdot \mathcal{H}_2(P_{X_1}), \quad (54d)$$

$$R_2'' \leq I_5, \quad (54e)$$

$$R_1' + R_2'' \leq \gamma \cdot I_6 + (1 - \gamma) \cdot I_5, \quad (54f)$$

$$R_1 + R_1' + R_2'' \leq I_6. \quad (54g)$$

Fig. 7 depicts the achievable region in (54) for the case where $p_1 = 0.01$ and $p_2 = 0.4$. Since (54) is parametrized by four rates, it is convenient to fix the rate R_1' on some value, which was chosen in our calculations to be $R_1' = 0.3$. We present five curves corresponding to the capacity region of the standard MAC without cribbing (blue “+” curve), the rates (R_1, R_2)

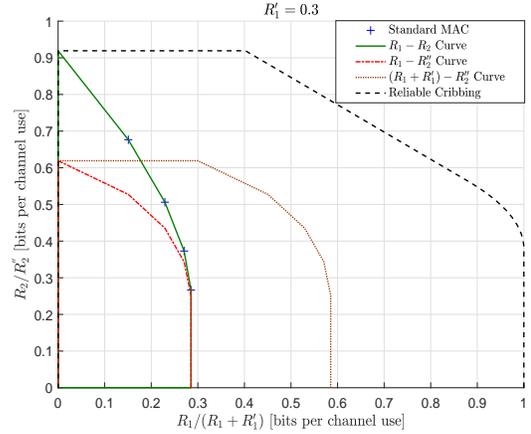


Fig. 7. The achievable region in Example 4 (see (54)), for $p_1 = 0.01$, $p_2 = 0.4$, and $R_1' = 0.3$, compared to the capacity region of the standard MAC without cribbing, and reliable cribbing. Each value of R_1 is associated with two rates R_2 and R_2'' on the (R_1, R_2) and (R_1, R_2'') curves.

(green curve), the rates (R_1, R_2'') (red dashed-dotted curve), the rates $(R_1 + R_1', R_2'')$ which refer to the total rate of User 1 versus the rate of User 2 when cribbing is present (brown dotted curve), and the capacity region in case of reliable cribbing [14] (black dashed curve). Each value of R_1 is associated with two rates R_2 and R_2'' . For example, if $R_1 = 0.1$ then $R_2 \approx 0.75$, $R_2'' \approx 0.56$, and $R_1' = 0.3$. This means that when cribbing is present User 2 reduces his rate R_2'' in favor of increasing the rate of User 1. This conclusion is noticeable from the fact that the (R_1, R_2) curve is on top of the (R_1, R_2'') curve for any R_1 . Finally, the best results are obtained in the case of reliable cribbing, as expected, and accordingly there is some degradation due to the fact that the cribbing is unreliable.

V. PROOFS

A. Proof of Theorem 1

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

Direct Part. We start with the code construction.

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

- 1) Generate e^{nR_2} codewords $\mathbf{u}(j)$, $j = 1, 2, \dots, e^{nR_2}$, i.i.d., according to P_U .
- 2) For every $\mathbf{u}(j)$, generate $e^{nR_2'}$ codewords $\mathbf{v}(k|j)$, $k = 1, 2, \dots, e^{nR_2'}$, 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 $\mathbf{v}(k|j)$, $k = 1, 2, \dots, e^{nR_2'}$, 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 $\mathbf{v}(k|j)$ with a fixed index j . Denote by $b(k|j)$ the bin number to which $\mathbf{v}(k|j)$ belongs. Note that
$$1 \leq b(k|j) \leq e^{nC_{1,2}}. \quad (55)$$
- 4) For every pair $(\mathbf{u}(j), \mathbf{v}(k|j))$, $j = 1, 2, \dots, e^{nR_2}$, $k = 1, 2, \dots, e^{nR_2'}$, generate e^{nR_1} vectors $\mathbf{x}(l|j, k)$, $l = 1, 2, \dots, e^{nR_1}$, 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 $j = 1, 2, \dots, e^{nR_2}$, $k = 1, 2, \dots, e^{nR'_2}$, $l = 1, 2, \dots, e^{nR_1}$, the encoder sends via the channel the codeword $\mathbf{x}(l|j, k)$.

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

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

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

$$R_2 < I(U; Y_2), \quad (56)$$

the index j is decoded correctly with high probability.

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

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

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

$$R_2 < I(U; Y_1), \quad (57)$$

Decoder 1 succeeds to decode correctly the index j with high probability. Since the channel is degraded, if (56) holds, it implies (57). Next, Decoder 1 looks for the unique index \hat{k} in $\{1, 2, \dots, \exp(nR'_2)\}$ such that

$$(\mathbf{u}(\hat{j}), \mathbf{v}(\hat{k}|\hat{j}), \mathbf{y}_1) \in T_\epsilon^{(n)}(UVY_1). \quad (58)$$

If such \hat{k} does not exist, or there is more than one such, an error is declared. By classical results, the index $k \in \{1, 2, \dots, \exp(nR'_2)\}$ is decoded correctly with high probability if

$$R'_2 < I(V; Y_1|U). \quad (59)$$

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

$$(\mathbf{u}(\hat{j}), \mathbf{v}(\hat{k}|\hat{j}), \mathbf{x}(\hat{l}|\hat{k}, \hat{j}), \mathbf{y}_1) \in T_\epsilon^{(n)}(UVXY_1). \quad (60)$$

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

$$R_1 < I(X; Y_1|U, V). \quad (61)$$

This concludes the decoding process when the conference link is absent. By (56), (59) and (61), the conditions for correct decoding when there is no conferencing are

$$R_2 \leq I(U; Y_2), \quad (62a)$$

$$R'_2 \leq I(V; Y_1|U), \quad (62b)$$

$$R_1 \leq I(X; Y_1|U, V). \quad (62c)$$

Observe that, although the rate R'_2 is decoded by Decoder 1 (if (62b) is satisfied), it does not arrive to User 2, since the conferencing link is absent. The bound (62b) 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{j} , \hat{k} , and \hat{l} . If (62) hold, these steps succeed with high probability. He then sends $b(\hat{k}|\hat{j})$, the index of the bin to which $\mathbf{v}(\hat{k}|\hat{j})$ belongs, via the conference link. Due to (55), the link capacity suffices, and Decoder 2 receives $b(\hat{k}|\hat{j})$ 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 $b(\hat{k}|\hat{j})$, he looks in this bin for the unique index \hat{k} such that

$$(\mathbf{u}(\hat{j}), \mathbf{v}(\hat{k}|\hat{j}), \mathbf{y}_2) \in T_\epsilon^{(n)}(UVY_2). \quad (63)$$

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 \mathbf{v} . Assuming that the previous decoding steps were successful (i.e., \hat{j} , \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

$$R'_2 - C_{1,2} \leq I(V; Y_2|U). \quad (64)$$

The region defined by (62) and (64) 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 $\lim_{n \rightarrow \infty} \epsilon_n = 0$. 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

$$nR_2 - n\delta_n \leq I(M_2; Y_2^n) \quad (65)$$

$$\begin{aligned} &\stackrel{(a)}{=} \sum_{i=1}^n I(M_2; Y_{2,i} | Y_2^{i-1}) \\ &\leq \sum_{i=1}^n I(M_2; Y_2^{i-1}; Y_{2,i}), \end{aligned} \quad (66)$$

where $\lim_{n \rightarrow \infty} \delta_n = 0$, due to $\lim_{n \rightarrow \infty} \epsilon_n = 0$, 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

$$nR'_2 - n\delta_n \leq I(M'_2; Y_2^n, M_{1,2} | M_2) \quad (67)$$

$$\begin{aligned} &= I(M'_2; Y_2^n | M_2) + I(M'_2; M_{1,2} | M_2, Y_2^n) \\ &\leq I(M'_2; Y_2^n | M_2) + H(M_{1,2}) \\ &= \sum_{i=1}^n I(M'_2; Y_{2,i} | M_2, Y_2^{i-1}) + H(M_{1,2}) \\ &\leq \sum_{i=1}^n I(M'_2; Y_1^{i-1}; Y_{2,i} | M_2, Y_2^{i-1}) + H(M_{1,2}). \end{aligned}$$

Moreover, the message M'_2 can be decoded by Decoder 1, regardless of the conference link. Hence:

$$\begin{aligned}
nR'_2 - n\delta_n &\leq I(M'_2; Y_1^n | M_2) \\
&= \sum_{i=1}^n I(M'_2; Y_{1,i} | M_2, Y_1^{i-1}) \\
&\stackrel{(a)}{=} \sum_{i=1}^n I(M'_2; Y_{1,i} | M_2, Y_1^{i-1}, Y_2^{i-1}) \\
&\leq \sum_{i=1}^n I(M'_2, Y_1^{i-1}; Y_{1,i} | M_2, Y_2^{i-1}), \quad (68)
\end{aligned}$$

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

$$\begin{aligned}
nR_1 - n\delta_n &\leq I(M_1; Y_1^n | M_2, M'_2) \\
&= \sum_{i=1}^n I(M_1; Y_{1,i} | M_2, M'_2, Y_1^{i-1}) \\
&\stackrel{(a)}{=} \sum_{i=1}^n I(M_1; Y_{1,i} | M_2, M'_2, Y_1^{i-1}, Y_2^{i-1}) \\
&\stackrel{(b)}{=} \sum_{i=1}^n I(X_i; Y_{1,i} | M_2, M'_2, Y_1^{i-1}, Y_2^{i-1}), \quad (69)
\end{aligned}$$

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 $U_i \triangleq (M_2, Y_2^{i-1})$ and $V_i \triangleq (M'_2, Y_1^{i-1})$, which due to (2) satisfy the Markov chain $(U_i, V_i) \circlearrowleft X_i \circlearrowleft (Y_{1,i}, Y_{1,i})$, and using the fact that

$$\frac{1}{n} H(M_{1,2}) \leq C_{1,2}, \quad (70)$$

we obtain from (66), (67), (68), and (69) the bounds

$$n(R_2 - \delta_n) \leq \sum_{i=1}^n I(U_i; Y_{2,i}), \quad (71a)$$

$$n(R'_2 - \delta_n) \leq \sum_{i=1}^n I(V_i; Y_{2,i} | U_i) + nC_{1,2}, \quad (71b)$$

$$n(R'_2 - \delta_n) \leq \sum_{i=1}^n I(U_i; Y_{1,i} | V_i), \quad (71c)$$

$$n(R_1 - \delta_n) \leq \sum_{i=1}^n I(X_i; Y_{1,i} | U_i, V_i). \quad (71d)$$

Using the standard time-sharing argument as in [22, Ch. 14.3], one can rewrite (71) by introducing an appropriate time-sharing random variable. Therefore, if $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$, 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 [21, Appendix C]. \square

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 ‘‘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 [14] for one sided 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 $B \rightarrow \infty$, the overall rates are $R_1(B - 1)/B \rightarrow R_1$ and $R'_1(B - 1)/B \rightarrow R'_1$. 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 \mathbf{v} , i.i.d., according to P_V . Label them $\mathbf{v}(m_0, m_1)$, for $m_0, m_1 \in \{1, 2, \dots, \exp(nR_1)\}$.
- 2) Generate $e^{n(R_1 + R'_1)}$ codewords \mathbf{u} , independently according to P_U . Label them $\mathbf{u}(m_0, m'_0)$, for $m_0 \in \{1, 2, \dots, \exp(nR_1)\}$ and $m'_0 \in \{1, 2, \dots, \exp(nR'_1)\}$.
- 3) For every $\mathbf{v}(m_0, m_1)$ and $\mathbf{u}(m_0, m'_0)$, generate $e^{nR'_1}$ codewords \mathbf{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 $\mathbf{x}_1(m'_1, \mathbf{u}(m_0, m'_0), \mathbf{v}(m_0, m_1))$, 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 $B \rightarrow \infty$, the overall rate is $R_2(B - 1)/B \rightarrow R_2$. 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 \mathbf{x}_2 , i.i.d., according to P_{X_2} . Label them $\mathbf{x}_2(m_2)$, 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:

$$\mathbf{x}_{1,1} = \mathbf{x}_1(m'_{1,1}, \mathbf{u}(1, 1), \mathbf{v}(1, m_{1,1})) \quad (72a)$$

$$\mathbf{x}_{2,1} = \mathbf{x}_2(m_{2,1}). \quad (72b)$$

Then, in block b , $b = 2, 3, \dots, B$, the encoders send (73), shown at the top of the 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:

$$(\mathbf{v}(1, \hat{m}_{1,1}), \mathbf{x}_2(\hat{m}_{2,1}), \mathbf{y}) \in T_\epsilon^{(n)}(VX_2Y). \quad (74)$$

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:

$$(\mathbf{v}(\hat{m}_{1,1}, \hat{m}_{1,2}), \mathbf{x}_2(\hat{m}_{2,2}), \mathbf{y}) \in T_\epsilon^{(n)}(VX_2Y). \quad (75)$$

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

$$(\mathbf{v}(\hat{m}_{1,b-1}, \hat{m}_{1,b}), \mathbf{x}_2(\hat{m}_{2,b}), \mathbf{y}) \in T_\epsilon^{(n)}(VX_2Y) \quad (76)$$

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:

$$R_1 \leq I(V; Y|X_2), \quad (77a)$$

$$R_2 \leq I(X_2; Y|V), \quad (77b)$$

$$R_1 + R_2 \leq I(V, X_2; Y). \quad (77c)$$

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 $M''_{2,b} \in \{1, 2, \dots, \exp(nR''_2)\}$, 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. 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 $\mathbf{u}(m_0, m'_0)$, generate $e^{nR''_2}$ codewords \mathbf{x}''_2 , independently according to $\prod_{i=1}^n P_{X''_2|U}(x_{2,i}|u_i(m_0, m'_0))$. Label them $\mathbf{x}''_2(m''_2, \mathbf{u}(m_0, m'_0))$, 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.

Encoding: 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: In block 1, the encoders send¹:

$$\mathbf{x}_{1,1} = \mathbf{x}_1(m'_{1,1}, \mathbf{u}(1, 1), \mathbf{v}(1, m_{1,1})) \quad (78a)$$

$$\mathbf{x}''_{2,1} = \mathbf{x}''_2(m''_{2,1}, \mathbf{u}(1, 1)). \quad (78b)$$

Assume that as a result of cribbing from encoder 1, after block b , $b = 1, 2, \dots, B-1$, 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:

$$(\mathbf{v}(\hat{m}_{1,b-1}, \hat{m}_{1,b}), \mathbf{x}_{1,b}) \in T_\epsilon^{(n)}(VX_1) \quad (79)$$

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 (80), shown at the top of the page, where $\hat{m}'_{1,b-1}$ was determined at the end of block $b-1$. Finally, in block b , $b = 2, 3, \dots, B$, the encoders send (81), shown at the top of the next page.

Decoding: Here, the principle of backward decoding [14] 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

$$\begin{aligned} & \{\mathbf{u}(\hat{m}_{1,B-1}, \hat{m}'_{1,B-1}), \mathbf{v}(\hat{m}_{1,B-1}, 1), \\ & \mathbf{x}_1(1, \mathbf{u}(\hat{m}_{1,B-1}, \hat{m}'_{1,B-1}), \mathbf{v}(\hat{m}_{1,B-1}, 1)), \\ & \mathbf{x}''_2(1, \mathbf{u}(\hat{m}_{1,B-1}, \hat{m}'_{1,B-1}), \mathbf{y}''_B)\} \\ & \in T_\epsilon^{(n)}(UVX_1X''_2Y''). \end{aligned} \quad (82)$$

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 for² $(\hat{m}_{1,B-2}, \hat{m}'_{1,B-2}, \hat{m}''_{2,B-1})$ according to (83), shown at the top of the 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 (84), 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 $b \in \{1, 2, \dots, B-1\}$, 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.

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

²The 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 $B-2$.

$$\mathbf{x}_{1,b} = \mathbf{x}_1(m'_{1,b}, \mathbf{u}(m_{1,b-1}, m'_{1,b-1}), \mathbf{v}(m_{1,b-1}, m_{1,b})), \quad b = 2, 3, \dots, B-1 \quad (73a)$$

$$\mathbf{x}_{2,b} = \mathbf{x}_2(m_{2,b}), \quad b = 2, 3, \dots, B-1 \quad (73b)$$

$$\mathbf{x}_{1B} = \mathbf{x}_1(1, \mathbf{u}(m_{1,B-1}, m'_{1,B-1}), \mathbf{v}(m_{1,B-1}, 1)), \quad (73c)$$

$$\mathbf{x}_{2B} = \mathbf{x}_2(m_{2,B}). \quad (73d)$$

$$\{\mathbf{u}(\hat{m}_{1,b-1}, \hat{m}'_{1,b-1}), \mathbf{v}(\hat{m}_{1,b-1}, \hat{m}_{1,b}), \mathbf{x}_1(\hat{m}'_{1,b}, \mathbf{u}(\hat{m}_{1,b-1}, \hat{m}'_{1,b-1}), \mathbf{v}(\hat{m}_{1,b-1}, \hat{m}_{1,b})), \mathbf{x}_{1,b}\} \in T_\epsilon^{(n)}(UVX_1X_1) \quad (80)$$

$$\mathbf{x}_{1,b} = \mathbf{x}_1(m'_{1,b}, \mathbf{u}(m_{1,b-1}, m'_{1,b-1}), \mathbf{v}(m_{1,b-1}, m_{1,b})), \quad b = 2, 3, \dots, B-1 \quad (81a)$$

$$\mathbf{x}''_{2,b} = \mathbf{x}''_2(m''_{2,b}, \mathbf{u}(\hat{m}_{1,b-1}, \hat{m}'_{1,b-1})), \quad b = 2, 3, \dots, B-1 \quad (81b)$$

$$\mathbf{x}_{1B} = \mathbf{x}_1(1, \mathbf{u}(m_{1,B-1}, m'_{1,B-1}), \mathbf{v}(m_{1,B-1}, 1)), \quad (81c)$$

$$\mathbf{x}''_{2B} = \mathbf{x}''_2(1, \mathbf{u}(\hat{m}_{1,B-1}, \hat{m}'_{1,B-1})). \quad (81d)$$

$$\{\mathbf{u}(\hat{m}_{1,B-2}, \hat{m}'_{1,B-2}), \mathbf{v}(\hat{m}_{1,B-2}, \hat{m}_{1,B-1}), \mathbf{x}_1(\hat{m}'_{1,B-1}, \mathbf{u}(\hat{m}_{1,B-2}, \hat{m}'_{1,B-2}), \mathbf{v}(\hat{m}_{1,B-2}, \hat{m}_{1,B-1})) \\ \mathbf{x}''_2(\hat{m}''_{2,B-1}, \mathbf{u}(\hat{m}_{1,B-2}, \hat{m}'_{1,B-2})), \mathbf{y}''_{B-1}\} \in T_\epsilon^{(n)}(UVX_1X_2''Y''). \quad (83)$$

$$\{\mathbf{u}(\hat{m}_{1,b-1}, \hat{m}'_{1,b-1}), \mathbf{v}(\hat{m}_{1,b-1}, \hat{m}_{1,b}), \mathbf{x}_1(\hat{m}'_{1,b}, \mathbf{u}(\hat{m}_{1,b-1}, \hat{m}'_{1,b-1}), \mathbf{v}(\hat{m}_{1,b-1}, \hat{m}_{1,b})) \\ \mathbf{x}''_2(\hat{m}''_{2,b}, \mathbf{u}(\hat{m}_{1,b-1}, \hat{m}'_{1,b-1})), \mathbf{y}''_b\} \in T_\epsilon^{(n)}(UVX_1X_2''Y''). \quad (84)$$

Error Analysis: The following lemma (see, e.g., [17, Lemma 4]) 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,

$$\Pr \left\{ \bigcup_{l=1}^L \mathcal{A}_l \right\} \leq \sum_{l=1}^L \Pr \{ \mathcal{A}_l | \mathcal{A}_1^c, \mathcal{A}_2^c, \dots, \mathcal{A}_{l-1}^c \} \quad (85)$$

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

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 $(1, 2, \dots, b-1)$ were decoded correctly.

Let $E_{\text{enc},b} = E_{\text{enc},b}^{(1)} \cup E_{\text{enc},b}^{(2)}$ be the event that encoder 2 has an error in decoding $m_{1,b}$ or $m'_{1,b}$. The event $E_{\text{enc},b}^{(1)}$ refers to an error in decoding $m_{1,b}$, while $E_{\text{enc},b}^{(2)}$ refers to an error in decoding $m'_{1,b}$. The term $\Pr \{ E_{\text{enc},b} | E_{\text{enc},b-1}^c \}$ 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,

$$\Pr \{ E_{\text{enc},b} | E_{\text{enc},b-1}^c \} \leq \Pr \{ E_{\text{enc},b}^{(1)} | E_{\text{enc},b-1}^c \} \\ + \Pr \{ E_{\text{enc},b}^{(2)} | E_{\text{enc},b-1}^c, (E_{\text{enc},b}^{(1)})^c \}. \quad (86)$$

Define the sets

$$\mathcal{E}_{b,m_{1,b}} \triangleq (\mathbf{v}(m_{1,b-1}, m_{1,b}), \mathbf{x}_{1,b}) \in T_\epsilon^{(n)}(VX_1), \quad (87)$$

and the set $\mathcal{E}_{b,m'_{1,b}}$ in (88), 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 (79),

$$\mathcal{E}_{b,m'_{1,b}} \triangleq \{\mathbf{u}(m_{1,b-1}, m'_{1,b-1}), \mathbf{v}(m_{1,b-1}, m_{1,b}), \mathbf{x}_1(m'_{1,b}, \mathbf{u}(m_{1,b-1}, m'_{1,b-1}), \mathbf{v}(m_{1,b-1}, m_{1,b})), \mathbf{x}_{1,b}\} \in T_\epsilon^{(n)}(UVX_1X_1). \quad (88)$$

$$\Pr \left\{ E_{\text{enc},b}^{(1)} | E_{\text{enc},b-1}^c \right\} \leq \Pr \left\{ \bigcup_{m_{1,b} \neq 1} \mathcal{E}_{b,m_{1,b}} | E_{\text{enc},b-1}^c \right\}. \quad (89)$$

The probability at the right hand side of (89), is the probability of the event in (87), given that $m_{1,b-1}$, was decoded correctly. Then, to evaluate (89), we can equivalently evaluate the probability of the event

$$\mathcal{E}_{b,m_{1,b}} \triangleq (\mathbf{v}(1, m_{1,b}), \mathbf{x}_{1,b}) \in T_\epsilon^{(n)}(VX_1), \quad (90)$$

for $m_{1,b} \neq 1$. Hence, by classical results, we have,

$$\Pr \left\{ E_{\text{enc},b}^{(1)} | E_{\text{enc},b-1}^c \right\} \leq \sum_{m_{1,b} \neq 1} \Pr \left\{ \mathcal{E}_{b,m_{1,b}} | E_{\text{enc},b-1}^c \right\} \quad (91)$$

$$\leq \sum_{m_{1,b} \neq 1} e^{-n(I(V;X_1)-3\epsilon)} \quad (92)$$

$$\leq e^{n(R_1-I(V;X_1)+3\epsilon)}. \quad (93)$$

Next, recall that encoder 2 decodes $m'_{1,b}$ according to (80), 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,

$$\begin{aligned} & \Pr \left\{ E_{\text{enc},b}^{(2)} | E_{\text{enc},b-1}^c, (E_{\text{enc},b}^{(1)})^c \right\} \\ &= \Pr \left\{ \bigcup_{m'_{1,b} \neq 1} \mathcal{E}_{b,m'_{1,b}} | E_{\text{enc},b-1}^c, (E_{\text{enc},b}^{(1)})^c \right\} \\ &\leq \sum_{m'_{1,b} \neq 1} \Pr \left\{ \mathcal{E}_{b,m'_{1,b}} | E_{\text{enc},b-1}^c, (E_{\text{enc},b}^{(1)})^c \right\}. \quad (94) \end{aligned}$$

Again, the probability at the right hand side of (94), is the probability of the event in (88), given that $m_{1,b-1}$, $m'_{1,b-1}$, and $m_{1,b}$, were decoded correctly. Then, to evaluate (94), we can equivalently evaluate the probability of the event in (95), shown at the top of the page, for $m'_{1,b} \neq 1$. We get

$$\begin{aligned} & \Pr \left\{ \mathcal{E}_{b,m'_{1,b}} | E_{\text{enc},b-1}^c, (E_{\text{enc},b}^{(1)})^c \right\} \\ &= \sum_{T_\epsilon^{(n)}(UVX_1X_1)} P(\mathbf{u})P(\mathbf{v})P(\mathbf{x}_1|\mathbf{u}, \mathbf{v})P(\mathbf{x}'_1|\mathbf{u}, \mathbf{v}) \\ &\leq \exp(n(H(U, V, X_1, X_1) + \epsilon)) \\ &\quad \cdot \exp(-n(H(U, V, X_1) - 4\epsilon)) \\ &\quad \cdot \exp(-n(H(X_1|U, V) - \epsilon)) \\ &= \exp(-n(H(X_1|U, V) - 6\epsilon)). \quad (96) \end{aligned}$$

Therefore,

$$\begin{aligned} \Pr \left\{ E_{\text{enc},b}^{(2)} | E_{\text{enc},b-1}^c, (E_{\text{enc},b}^{(1)})^c \right\} &\leq \sum_{m'_{1,b} \neq 1} e^{-n(H(X_1|U, V) - 6\epsilon)} \\ &\leq e^{n(R'_1 - H(X_1|U, V) + 6\epsilon)}. \quad (97) \end{aligned}$$

Wrapping up, using (93) and (97), 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 (84), 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 (98), shown at the top of the 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., $E_{1,1,1,b}^c$, 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 $P_{e,b}^{(n)}$ be the decoding error probability at block b given that in blocks $(b+1, \dots, B)$, there was no decoding error. From the union bound, we obtain that:

$$\begin{aligned} P_{e,b}^{(n)} &\leq \Pr \left\{ E_{1,1,1,b}^c \right\} + \sum_{m_1 > 1} \Pr \left\{ E_{m_1, 1, 1, b} \right\} \\ &\quad + \sum_{m'_1 > 1} \Pr \left\{ E_{1, m'_1, 1, b} \right\} + \sum_{m''_2 > 1} \Pr \left\{ E_{1, 1, m''_2, b} \right\} \\ &\quad + \sum_{m_1 > 1, m'_1 > 1} \Pr \left\{ E_{m_1, m'_1, 1, b} \right\} \\ &\quad + \sum_{m_1 > 1, m''_2 > 1} \Pr \left\{ E_{m_1, 1, m''_2, b} \right\} \\ &\quad + \sum_{m'_1 > 1, m''_2 > 1} \Pr \left\{ E_{1, m'_1, m''_2, b} \right\} \\ &\quad + \sum_{m_1 > 1, m'_1 > 1, m''_2 > 1} \Pr \left\{ E_{m_1, m'_1, m''_2, b} \right\}. \quad (99) \end{aligned}$$

Let us upper bound each term in (99).

$$\tilde{\mathcal{E}}_{b,m'_1,b} \triangleq \{\mathbf{u}(1,1), \mathbf{v}(1,1), \mathbf{x}_1(m'_{1,b}, \mathbf{u}(1,1), \mathbf{v}(1,1)), \mathbf{x}_{1,b}\} \in T_\epsilon^{(n)}(UVX_1X_1). \quad (95)$$

$$E_{m_1,m'_1,m'_2,b} \triangleq \{\mathbf{u}(m_1, m'_1), \mathbf{v}(m_1, m_{1,b}), \mathbf{x}_1(m'_{1,b}, \mathbf{u}(m_1, m'_1), \mathbf{v}(m_1, m_{1,b})), \mathbf{x}_2''(m'_2, \mathbf{u}(m_1, m'_1)), \mathbf{y}_b''\} \\ \in T_\epsilon^{(n)}(UVX_1X_2''Y''). \quad (98)$$

- 1) Upper-bounding $\Pr\{E_{1,1,1,b}^c\}$: 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 $\Pr\{E_{1,1,1,b}^c\} \rightarrow 0$ as $n \rightarrow \infty$.
- 2) Upper-bounding $\sum_{m'_2 > 1} \Pr\{E_{1,1,m'_2,b}\}$: Let \mathcal{S} be the set of all sequences $(\mathbf{u}, \mathbf{v}, \mathbf{x}_1, \mathbf{x}_2'', \mathbf{y}'')$ that belong to $T_\epsilon^{(n)}(UVX_1X_2''Y'')$. We then have

$$\begin{aligned} & \Pr\{E_{1,1,m'_2,b}\} \\ &= \sum_{\mathcal{S}} P(\mathbf{u})P(\mathbf{v})P(\mathbf{x}_1|\mathbf{u}, \mathbf{v})P(\mathbf{x}_2''|\mathbf{u}) \\ & \quad \times P(\mathbf{y}''|\mathbf{u}, \mathbf{v}, \mathbf{x}_1) \\ &\leq \exp(n(H(U, V, X_1, X_2'', Y'') + \epsilon)) \\ & \quad \cdot \exp(-n(H(U, V, X_1) - 4\epsilon)) \\ & \quad \cdot \exp(-n(H(X_2''|U) - \epsilon)) \\ & \quad \cdot \exp(-n(H(Y''|U, V, X_1) - \epsilon)) \\ &= \exp(-n(I(X_2''; Y''|U, V, X_1) - 7\epsilon)). \quad (100) \end{aligned}$$

where we have used the fact that $(V, X_1) \ominus U \ominus X_2''$. Hence, we obtain

$$\sum_{m'_2 > 1} \Pr\{E_{1,1,m'_2,b}\} \leq e^{n(R_2'' - I(X_2''; Y''|U, V, X_1) + 7\epsilon)}. \quad (101)$$

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

$$\begin{aligned} & \Pr\{E_{m_1,1,1,b}\} \\ &= \sum_{\mathcal{S}} P(\mathbf{u})P(\mathbf{v})P(\mathbf{x}_1|\mathbf{u}, \mathbf{v})P(\mathbf{x}_2''|\mathbf{u})P(\mathbf{y}'') \\ &\leq \exp(n(H(U, V, X_1, X_2'', Y'') + \epsilon)) \\ & \quad \cdot \exp(-n(H(U, V, X_1) - 4\epsilon)) \\ & \quad \cdot \exp(-n(H(X_2''|U) - \epsilon)) \\ & \quad \cdot \exp(-n(H(Y'') - \epsilon)) \\ &= \exp(-n(I(U, V, X_1, X_2''; Y'') - 7\epsilon)). \quad (102) \end{aligned}$$

where again we use $(V, X_1) \ominus U \ominus X_2''$. Hence, we obtain

$$\sum_{m_1 > 1} \Pr\{E_{m_1,1,1,b}\} \leq e^{n(R_1 - I(U, V, X_1, X_2''; Y'') + 7\epsilon)}. \quad (103)$$

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

$$\begin{aligned} & \Pr\{E_{1,m'_1,1,b}\} \\ &= \sum_{\mathcal{S}} P(\mathbf{u})P(\mathbf{v})P(\mathbf{x}_1|\mathbf{u}, \mathbf{v})P(\mathbf{x}_2''|\mathbf{u})P(\mathbf{y}''|\mathbf{v}) \\ &\leq \exp(n(H(U, V, X_1, X_2'', Y'') + \epsilon)) \\ & \quad \cdot \exp(-n(H(U, V, X_1) - 4\epsilon)) \\ & \quad \cdot \exp(-n(H(X_2''|U) - \epsilon)) \\ & \quad \cdot \exp(-n(H(Y''|V) - \epsilon)) \\ &= \exp(-n(I(U, X_1, X_2''; Y''|V) - 7\epsilon)). \quad (104) \end{aligned}$$

where again we use $(V, X_1) \ominus U \ominus X_2''$. Hence, we get

$$\sum_{m'_1 > 1} \Pr\{E_{1,m'_1,1,b}\} \leq e^{n(R_1' - I(U, X_1, X_2''; Y''|V) + 7\epsilon)}. \quad (105)$$

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

$$\begin{aligned} & \Pr\{E_{m_1, m'_1, 1, b}\} \\ &= \sum_{\mathcal{S}} P(\mathbf{u})P(\mathbf{v})P(\mathbf{x}_1|\mathbf{u}, \mathbf{v})P(\mathbf{x}_2''|\mathbf{u})P(\mathbf{y}'') \\ &\leq \exp(n(H(U, V, X_1, X_2'', Y'') + \epsilon)) \\ & \quad \cdot \exp(-n(H(U, V, X_1) - 4\epsilon)) \\ & \quad \cdot \exp(-n(H(X_2''|U) - \epsilon)) \\ & \quad \cdot \exp(-n(H(Y'') - \epsilon)) \\ &= \exp(-n(I(U, V, X_1, X_2''; Y'') - 7\epsilon)). \quad (106) \end{aligned}$$

where we use $(V, X_1) \ominus U \ominus X_2''$. Therefore,

$$\sum_{m_1 > 1, m'_1 > 1} \Pr\{E_{m_1, m'_1, 1, b}\} \leq e^{n(R_1 + R_1' - I(U, V, X_1, X_2''; Y'') + 7\epsilon)}. \quad (107)$$

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

$$\begin{aligned} & \Pr\{E_{m_1, 1, m'_2, b}\} \\ &= \sum_{\mathcal{S}} P(\mathbf{u})P(\mathbf{v})P(\mathbf{x}_1|\mathbf{u}, \mathbf{v})P(\mathbf{x}_2''|\mathbf{u})P(\mathbf{y}'') \\ &\leq \exp(n(H(U, V, X_1, X_2'', Y'') + \epsilon)) \\ & \quad \cdot \exp(-n(H(U, V, X_1) - 4\epsilon)) \\ & \quad \cdot \exp(-n(H(X_2''|U) - \epsilon)) \end{aligned}$$

$$\begin{aligned} & \cdot \exp(-n(H(Y'') - \epsilon)) \\ & = \exp(-n(I(U, V, X_1, X_2''; Y'') - 7\epsilon)). \end{aligned} \quad (108)$$

using $(V, X_1) \ominus U \ominus X_2''$. Thus,

$$\begin{aligned} & \sum_{m_1 > 1, m_2'' > 1} \Pr \{E_{m_1, 1, m_2'', b}\} \\ & \leq e^{n(R_1 + R_2'' - I(U, V, X_1, X_2''; Y'') + 7\epsilon)}. \end{aligned} \quad (109)$$

7) Upper-bounding $\sum_{m_1' > 1, m_2'' > 1} \Pr \{E_{1, m_1', m_2'', b}\}$: We have

$$\begin{aligned} & \Pr \{E_{1, m_1', m_2'', b}\} \\ & = \sum_{\mathcal{S}} P(\mathbf{u})P(\mathbf{v})P(\mathbf{x}_1|\mathbf{u}, \mathbf{v})P(\mathbf{x}_2''|\mathbf{u})P(\mathbf{y}''|\mathbf{v}) \\ & \leq \exp(n(H(U, V, X_1, X_2'', Y'') + \epsilon)) \\ & \quad \cdot \exp(-n(H(U, V, X_1) - 4\epsilon)) \\ & \quad \cdot \exp(-n(H(X_2''|U) - \epsilon)) \\ & \quad \cdot \exp(-n(H(Y''|V) - \epsilon)) \\ & = \exp(-n(I(U, X_1, X_2''; Y''|V) - 7\epsilon)). \end{aligned} \quad (110)$$

where the last step follows from $(V, X_1) \ominus U \ominus X_2''$. Hence, we get

$$\begin{aligned} & \sum_{m_1' > 1, m_2'' > 1} \Pr \{E_{1, m_1', m_2'', b}\} \\ & \leq e^{n(R_1' + R_2'' - I(U, X_1, X_2''; Y''|V) + 7\epsilon)}. \end{aligned} \quad (111)$$

8) Upper-bounding $\sum_{m_1 > 1, m_1' > 1, m_2'' > 1} \Pr \{E_{m_1, m_1', m_2'', b}\}$: We have

$$\begin{aligned} & \Pr \{E_{m_1, m_1', m_2'', b}\} \\ & = \sum_{\mathcal{S}} P(\mathbf{u})P(\mathbf{v})P(\mathbf{x}_1|\mathbf{u}, \mathbf{v})P(\mathbf{x}_2''|\mathbf{u})P(\mathbf{y}'') \\ & \leq \exp(n(H(U, V, X_1, X_2'', Y'') + \epsilon)) \\ & \quad \cdot \exp(-n(H(U, V, X_1) - 4\epsilon)) \\ & \quad \cdot \exp(-n(H(X_2''|U) - \epsilon)) \\ & \quad \cdot \exp(-n(H(Y'') - \epsilon)) \\ & = \exp(-n(I(U, V, X_1, X_2''; Y'') - 7\epsilon)). \end{aligned} \quad (112)$$

where again we use $(V, X_1) \ominus U \ominus X_2''$. Hence, we obtain

$$\begin{aligned} & \sum_{m_1 > 1, m_1' > 1, m_2'' > 1} \Pr \{E_{m_1, m_1', m_2'', b}\} \\ & \leq e^{n(R_1 + R_1' + R_2'' - I(U, V, X_1, X_2''; Y'') + 7\epsilon)}. \end{aligned} \quad (113)$$

Thus, using (93), (97), (101), (103), (105), (107), (109), (111), and (113), if (R_1, R_1', R_2') satisfy:

$$R_1 \leq I(V; X_1), \quad (114a)$$

$$R_1' \leq H(X_1|U, V), \quad (114b)$$

$$R_1 \leq I(U, V, X_1, X_2''; Y''), \quad (114c)$$

$$R_1' \leq I(U, X_1, X_2''; Y''|V), \quad (114d)$$

$$R_1 + R_1' \leq I(U, V, X_1, X_2''; Y''), \quad (114e)$$

$$R_2'' \leq I(X_2''; Y''|U, V, X_1), \quad (114f)$$

$$R_1 + R_2'' \leq I(U, V, X_1, X_2''; Y''), \quad (114g)$$

$$R_1' + R_2'' \leq I(U, X_1, X_2''; Y''|V), \quad (114h)$$

$$R_1 + R_1' + R_2'' \leq I(U, V, X_1, X_2''; Y''), \quad (114i)$$

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 the following simplifications. First, we can remove (114c), (114e), and (114g), due to (114i), and (114d) can be removed due to (114h). Second, (114h) and (114i) 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 $(U, V) \ominus (X_1, X_2'') \ominus Y''$. Finally, the constraint in (114a), is superfluous due to (77a). Indeed,

$$I(V; Y|X_2) = H(V|X_2) - H(V|X_2, Y) \quad (115)$$

$$\stackrel{(a)}{\leq} H(V) - H(V|X_1, X_2, Y) \quad (116)$$

$$\stackrel{(b)}{=} H(V) - H(V|X_1) \quad (117)$$

$$= I(V; X_1) \quad (118)$$

where (a) follows from the fact that conditioning reduces entropy, and (b) follows from the Markov chain $(X_2, Y) \ominus X_1 \ominus V$. Thus, to summarize, using the above simplifications, the achievable region for the MAC with unreliable strictly causal cribbing is given (recall (77))

$$R_1 \leq I(V; Y|X_2), \quad (119a)$$

$$R_2 \leq I(X_2; Y|V), \quad (119b)$$

$$R_1 + R_2 \leq I(V, X_2; Y), \quad (119c)$$

$$R_1' \leq H(X_1|U, V), \quad (119d)$$

$$R_2'' \leq I(X_2''; Y''|U, V, X_1), \quad (119e)$$

$$R_1' + R_2'' \leq I(X_1, X_2''; Y''|V), \quad (119f)$$

$$R_1 + R_1' + R_2'' \leq I(X_1, X_2''; Y''), \quad (119g)$$

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

$$P_U P_V P_{X_1|U, V} P_{X_2} P_{X_2''|U} P_{Y|X_1, X_2} P_{Y''|X_1, X_2''}, \quad (120)$$

as stated in Theorem 2.

C. Proof of Theorem 3

In order to show that all the rate pairs in (32) are achievable, we employ Shannon strategies [14]. Consider all different strategies (functions), with members $t \in \mathcal{T} \triangleq \mathcal{X}_2^{|\mathcal{X}_1|}$ 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 $(\mathcal{X}_1 \times \mathcal{T}, P^\Delta(y''|x_1, t), \mathcal{Y})$ where $P^\Delta(y''|x_1, t) \triangleq P(y''|x_1, x_2'' = t(x_1))$ for all $x_1 \in \mathcal{X}_1$, $t \in \mathcal{T}$, and $y'' \in \mathcal{Y}$.

Let \mathcal{R}_S be the set of rates (R_1, R'_1, R_2, R''_2) satisfying

$$R_1 \leq I(V; Y|X_2), \quad (121a)$$

$$R_2 \leq I(X_2; Y|V), \quad (121b)$$

$$R_1 + R_2 \leq I(V, X_2; Y), \quad (121c)$$

$$R'_1 \leq H(X_1|U, V), \quad (121d)$$

$$R''_2 \leq I(T; Y''|U, V, X_1), \quad (121e)$$

$$R'_1 + R''_2 \leq I(X_1, T; Y''|V), \quad (121f)$$

$$R_1 + R'_1 + R''_2 \leq I(X_1, T; Y''), \quad (121g)$$

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

$$\begin{aligned} P(u, v, x_1, x_2, t, y, y'') = \\ P(u)P(v)P(x_1|u, v)P(x_2)P(t|u)P(y|x_1, x_2)P^\Delta(y''|x_1, t). \end{aligned} \quad (122)$$

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 (122) to satisfy

$$\begin{aligned} P(u, v, x_1, x_2, t, y, y'') = P(u)P(v)P(x_1|v)P(x_2)P(t) \\ \cdot P(y|x_1, x_2)P^\Delta(y''|x_1, t), \end{aligned} \quad (123)$$

then

$$H(X_1|U, V) = H(X_1|V), \quad (124a)$$

$$I(T; Y''|U, V, X_1) = I(X''_2; Y''|V, X_1), \quad (124b)$$

$$I(X_1, T; Y''|V) = I(X_1, X''_2; Y''|V), \quad (124c)$$

$$I(X_1, T; Y'') = I(X_1, X''_2; Y''), \quad (124d)$$

and³

$$P(v, x_1, x''_2, y'') = P(v, x_1) \sum_{t: t(x_1)=x''_2} P(t)P(y''|x_1, x''_2). \quad (125)$$

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

$$P(v, x_1) \sum_{t: t(x_1)=x''_2} P(t) = P^0(v, x_1, x''_2). \quad (126)$$

Indeed, this holds for the following choice:

$$P(v, x_1) = \sum_{x''_2} P^0(v, x_1, x''_2), \quad (127a)$$

$$P(t) = \sum_{x'_1} \frac{P^0(x'_1, x''_2 = t(x'_1))}{P(x'_1)}. \quad (127b)$$

³Recall 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

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

Thus, using (121) and (124), we conclude that all rate pairs

$$R_1 \leq I(V; Y|X_2), \quad (128a)$$

$$R_2 \leq I(X_2; Y|V), \quad (128b)$$

$$R_1 + R_2 \leq I(V, X_2; Y), \quad (128c)$$

$$R'_1 \leq H(X_1|V), \quad (128d)$$

$$R''_2 \leq I(X''_2; Y''|V, X_1), \quad (128e)$$

$$R'_1 + R''_2 \leq I(X_1, X''_2; Y''|V), \quad (128f)$$

$$R_1 + R'_1 + R''_2 \leq I(X_1, X''_2; Y''), \quad (128g)$$

for some $P_{V, X_1, X_2, X''_2, Y, Y''}$ of the form

$$P_{V, X_1} P_{X_2} P_{X''_2|X_1} P_{Y|X_1, X_2} P_{Y''|X_1, X''_2}, \quad (129)$$

are achievable for the MAC with causal cribbing. This completes the proof of Theorem 3. \square

D. Proof of Theorem 4

We next show that $\mathcal{I}_{\text{mac}}^O$, defined in (35), 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 $\lim_{n \rightarrow \infty} \epsilon_n = 0$. 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

$$nR_1 - n\delta_n \leq I(M_1; Y^n|M_2) \quad (130)$$

$$= \sum_{i=1}^n I(M_1; Y_i|Y^{i-1}, M_2) \quad (131)$$

$$\stackrel{(a)}{\leq} \sum_{i=1}^n I(M_1, Y^{i-1}; Y_i|M_2) \quad (132)$$

$$\stackrel{(b)}{=} \sum_{i=1}^n I(M_1; Y_i|M_2) + I(Y^{i-1}; Y_i|M_1, M_2) \quad (133)$$

$$\stackrel{(c)}{=} \sum_{i=1}^n I(M_1; Y_i|M_2, X_{2,i}) + I(Y^{i-1}; Y_i|M_1, M_2) \quad (134)$$

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

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

$$\stackrel{(d)}{=} \sum_{i=1}^n I(M_1; Y_i|X_{2,i}) + I(Y^{i-1}; Y_i|M_1, M_2) \quad (137)$$

where $\lim_{n \rightarrow \infty} \delta_n = 0$, due to $\lim_{n \rightarrow \infty} \epsilon_n = 0$, (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 $M_2 \ominus (M_1, X_{2,i}) \ominus Y_i$, 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 (137), can be upper bounded as follows

$$I(Y^{i-1}; Y_i | M_1, M_2) \stackrel{(a)}{=} I(Y^{i-1}; Y_i | M_1, M_2, X_2^i) \quad (138)$$

$$\leq I(Y^{i-1}, X_1^{i-1}; Y_i | M_1, M_2, X_2^i) \quad (139)$$

$$\stackrel{(b)}{=} I(X_1^{i-1}; Y_i | M_1, M_2, X_2^i) \quad (140)$$

$$\leq I(X_2^{i-1}, M_2, X_1^{i-1}; Y_i | M_1, X_{2,i}) \quad (141)$$

$$\stackrel{(c)}{=} I(X_1^{i-1}; Y_i | M_1, X_{2,i}) + I(X_2^{i-1}, M_2; Y_i | M_1, X_{2,i}, X_1^{i-1}) \quad (142)$$

$$\stackrel{(d)}{=} I(X_1^{i-1}; Y_i | M_1, X_{2,i}) \quad (143)$$

where (a) is due to the fact that X_2^i is a deterministic function of M_2 , (b) follows from the fact that $Y^{i-1} \ominus (X_1^{i-1}, X_2^i, M_1, M_2) \ominus Y_i$ (see, Lemma 3), (c) follows from the chain rule of mutual information, and finally (d) is due to the Markov chain $(M_2, X_2^{i-1}) \ominus (M_1, X_1^{i-1}, X_{2,i}) \ominus Y_i$ (see, Lemma 3). Wrapping up, we obtained

$$\begin{aligned} nR_1 - n\delta_n &\leq \sum_{i=1}^n I(M_1; Y_i | X_{2,i}) + I(X_1^{i-1}; Y_i | M_1, X_{2,i}) \\ &= \sum_{i=1}^n I(M_1, X_1^{i-1}; Y_i | X_{2,i}). \end{aligned} \quad (144)$$

Next, for R_2 we have:

$$nR_2 - n\delta_n \leq I(M_2; Y^n | M_1, M_1') \quad (145)$$

$$= \sum_{i=1}^n I(M_2; Y_i | M_1, M_1', Y^{i-1}) \quad (146)$$

$$\stackrel{(a)}{=} \sum_{i=1}^n I(X_{2,i}, M_2; Y_i | M_1, M_1', X_{1,i}, Y^{i-1}) \quad (147)$$

$$\leq \sum_{i=1}^n I(X_{2,i}, M_1, M_1', M_2, Y^{i-1}; Y_i | X_{1,i}) \quad (148)$$

$$\stackrel{(b)}{=} \sum_{i=1}^n I(X_{2,i}; Y_i | X_{1,i}) + I(M_1, M_1', M_2, Y^{i-1}; Y_i | X_{1,i}, X_{2,i}) \quad (149)$$

$$\stackrel{(c)}{=} \sum_{i=1}^n I(X_{2,i}; Y_i | X_{1,i}) \quad (150)$$

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 $(M_1, M_1', M_2, Y^{i-1}) \ominus (X_{1,i}, X_{2,i}) \ominus Y_i$. Finally, for the sum rate we have

$$n(R_1 + R_2) - n\delta_n \leq \sum_{i=1}^n I(M_1, M_2; Y_i | Y^{i-1}) \quad (151)$$

$$\begin{aligned} &\leq \sum_{i=1}^n I(M_1, M_2, Y^{i-1}; Y_i) \quad (152) \\ &= \sum_{i=1}^n I(M_1, M_2; Y_i) + I(Y^{i-1}; Y_i | M_1, M_2) \end{aligned} \quad (153)$$

where the last equality follows from the chain rule. However, we already saw that (recall (143)):

$$I(Y^{i-1}; Y_i | M_1, M_2) \leq I(X_1^{i-1}; Y_i | M_1, X_{2,i}), \quad (154)$$

and thus

$$\begin{aligned} n(R_1 + R_2) - n\delta_n &\leq \sum_{i=1}^n I(M_1, M_2; Y_i) \\ &\quad + I(X_1^{i-1}; Y_i | M_1, X_{2,i}) \end{aligned} \quad (155)$$

$$\stackrel{(a)}{=} \sum_{i=1}^n I(M_1, M_2, X_{2,i}; Y_i) + I(X_1^{i-1}; Y_i | M_1, X_{2,i}) \quad (156)$$

$$\stackrel{(b)}{=} \sum_{i=1}^n I(M_1, X_{2,i}; Y_i) + I(X_1^{i-1}; Y_i | M_1, X_{2,i}) \quad (157)$$

$$= \sum_{i=1}^n I(M_1, X_1^{i-1}, X_{2,i}; Y_i) \quad (158)$$

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 $I(M_1, M_2, X_{2,i}; Y_i) = I(M_1, X_{2,i}; Y_i) + I(M_2; Y_i | M_1, X_{2,i})$ and that $M_2 \ominus (M_1, X_{2,i}) \ominus Y_i$.

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

$$nR_1' - n\delta_n \leq I(M_1'; Y^{n''} | M_1) \quad (159)$$

$$= I(M_1, M_1'; Y^{n''} | M_1) \quad (160)$$

$$\stackrel{(a)}{=} I(M_1, M_1', X_1^n; Y^{n''} | M_1) \quad (161)$$

$$\stackrel{(b)}{=} I(X_1^n; Y^{n''} | M_1) + I(M_1, M_1'; Y^{n''} | M_1, X_1^n) \quad (162)$$

$$\stackrel{(c)}{=} I(X_1^n; Y^{n''} | M_1) \quad (163)$$

$$\leq H(X_1^n | M_1) \quad (164)$$

$$\stackrel{(d)}{=} \sum_{i=1}^n H(X_{1,i} | M_1, X_1^{i-1}) \quad (165)$$

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 $(M_1, M_1') \ominus X_1^n \ominus Y^{n''}$ (see, Lemma 3), and (d) is due to the entropy chain rule. Next, for R_2'' we have:

$$nR_2'' - n\delta_n \leq I(M_2''; Y^{n''} | M_1, M_1') \quad (166)$$

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

$$\stackrel{(a)}{=} \sum_{i=1}^n I(M_2''; Y_i'' | Y^{i-1''}, M_1, M_1', X_1^{i-1}, X_{1,i}) \quad (168)$$

$$\leq \sum_{i=1}^n I(Y^{i-1''}, M'_1, M''_2; Y_i'' | M_1, X_1^{i-1}, X_{1,i}) \quad (169)$$

$$\stackrel{(b)}{=} \sum_{i=1}^n I(Y^{i-1''}, M'_1, M''_2, X''_{2,i}; Y_i'' | M_1, X_1^{i-1}, X_{1,i}) \quad (170)$$

$$\stackrel{(c)}{=} \sum_{i=1}^n I(X''_{2,i}; Y_i'' | M_1, X_1^{i-1}, X_{1,i}) \quad (171)$$

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 chain rule for mutual information and the Markov chain $(M_1, X_1^{i-1}, Y^{i-1''}, M'_1, M''_2) \ominus (X_{1,i}, X''_{2,i}) \ominus Y_i''$. Finally, for the sum rate $R_1 + R'_1 + R''_2$, we have:

$$n(R_1 + R'_1 + R''_2) - n\delta_n \leq I(M_1, M'_1, M''_2; Y^{n''}) \quad (172)$$

$$\leq \sum_{i=1}^n I(X_{1,i}, X''_{2,i}; Y_i''). \quad (173)$$

So, hitherto we have that:

$$n(R_1 - \delta_n) \leq \sum_{i=1}^n I(M_1, X_1^{i-1}; Y_i | X_{2,i}) \quad (174a)$$

$$n(R_2 - \delta_n) \leq \sum_{i=1}^n I(X_{2,i}; Y_i | X_{1,i}) \quad (174b)$$

$$n(R_1 + R_2 - \delta_n) \leq \sum_{i=1}^n I(M_1, X_1^{i-1}, X_{2,i}; Y_i) \quad (174c)$$

$$n(R'_1 - \delta_n) \leq \sum_{i=1}^n H(X_{1,i} | M_1, X_1^{i-1}) \quad (174d)$$

$$n(R''_2 - \delta_n) \leq \sum_{i=1}^n I(X''_{2,i}; Y_i'' | M_1, X_1^{i-1}, X_{1,i}) \quad (174e)$$

$$n(R'_1 + R''_2 - \delta_n) \leq \sum_{i=1}^n H(X_{1,i} | M_1, X_1^{i-1}) + I(X''_{2,i}; Y_i'' | M_1, X_1^{i-1}, X_{1,i}) \quad (174f)$$

$$n(R_1 + R'_1 + R''_2 - \delta_n) \leq \sum_{i=1}^n I(X_{1,i}, X''_{2,i}; Y_i''). \quad (174g)$$

We are now in a position to define our auxiliary RV. From (174a)-(174g), letting $V_i \triangleq (M_1, X_1^{i-1})$, and thus preserving the Markov chain induced by \mathcal{P} , we have that

$$n(R_1 - \delta_n) \leq \sum_{i=1}^n I(V_i; Y_i | X_{2,i}) \quad (175a)$$

$$n(R_2 - \delta_n) \leq \sum_{i=1}^n I(X_{2,i}; Y_i | X_{1,i}) \quad (175b)$$

$$n(R_1 + R_2 - \delta_n) \leq \sum_{i=1}^n I(V_i, X_{2,i}; Y_i) \quad (175c)$$

$$n(R'_1 - \delta_n) \leq \sum_{i=1}^n H(X_{1,i} | V_i) \quad (175d)$$

$$n(R''_2 - \delta_n) \leq \sum_{i=1}^n I(X''_{2,i}; Y_i'' | V_i, X_{1,i}) \quad (175e)$$

$$n(R'_1 + R''_2 - \delta_n) \leq \sum_{i=1}^n H(X_{1,i} | V_i) + I(X''_{2,i}; Y_i'' | V_i, X_{1,i}) \quad (175f)$$

$$n(R_1 + R'_1 + R''_2 - \delta_n) \leq \sum_{i=1}^n I(X_{1,i}, X''_{2,i}; Y_i''). \quad (175g)$$

Using the standard time-sharing argument as in [22, Ch. 14.3], one can rewrite (175) by introducing an appropriate time-sharing random variable. Therefore, if $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$, the convex hull of this region can be shown to be equivalent to the convex hull of the region in (35).

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, (37)). 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, (171)):

$$nR''_2 - n\delta_n \leq I(M''_2; Y^{n''} | M_1, M'_1) \quad (176)$$

$$= \sum_{i=1}^n I(M''_2; Y_i'' | Y^{i-1''}, M_1, M'_1) \quad (177)$$

$$\stackrel{(a)}{=} \sum_{i=1}^n I(M''_2; Y_i'' | Y^{i-1''}, M_1, M'_1, X_1^n) \quad (178)$$

$$\leq \sum_{i=1}^n I(M_1, X_1^{n/i}, Y^{i-1''}, M'_1, M''_2; Y_i'' | X_{1,i}) \quad (179)$$

$$\stackrel{(b)}{=} \sum_{i=1}^n I(M_1, X_1^{n/i}, Y^{i-1''}, M'_1, M''_2, X''_{2,i}; Y_i'' | X_{1,i}) \quad (180)$$

$$\stackrel{(c)}{=} \sum_{i=1}^n I(X''_{2,i}; Y_i'' | X_{1,i}) \quad (181)$$

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 $(M_1, X_1^{n/i}, Y^{i-1''}, M'_1, M''_2) \ominus (X_{1,i}, X''_{2,i}) \ominus Y_i''$, 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) $M_2 \ominus (M_1, X_{2,i}) \ominus Y_i$
- 2) $(M_2, X_2^{i-1}) \ominus (M_1, X_1^{i-1}, X_{2,i}) \ominus Y_i$
- 3) $Y^{i-1} \ominus (X_1^{i-1}, X_2^{i-1}) \ominus Y_i$
- 4) $Y^{i-1} \ominus (X_1^{i-1}, X_2^{i-1}, M_1, M_2) \ominus Y_i$
- 5) $Y^{i-1} \ominus (X_1^{i-1}, X_2^i, M_1, M_2) \ominus Y_i$
- 6) $(M_1, M'_1) \ominus X_1^n \ominus Y^{n''}$

Proof of Lemma 3: First, recall that:

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

Thus, the first item of Lemma 3 follows from:

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

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

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

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,

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

$$= \sum_{x_{1,i}} P_{Y_i|M_1, X_1^i, X_2^i} \times P_{X_{1,i}|M_1, X_1^{i-1}, X_2^i} \quad (\text{A.6})$$

$$= P_{Y_i|M_1, X_1^{i-1}, X_2^i}. \quad (\text{A.7})$$

Next, the third item is true because:

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

$$= \sum_{x_{1,i}, x_{2,i}} P_{Y_i|X_1^i, X_2^i} P_{X_{1,i}, X_{2,i}|X_1^{i-1}, X_2^{i-1}} \quad (\text{A.9})$$

$$= \sum_{x_{1,i}, x_{2,i}} P_{X_{1,i}, X_{2,i}, Y_i|X_1^{i-1}, X_2^{i-1}} \quad (\text{A.10})$$

$$= P_{Y_i|X_1^{i-1}, X_2^{i-1}} \quad (\text{A.11})$$

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

$$P_{Y_i|X_1^{i-1}, X_2^i, Y^{i-1}, M_1, M_2} = \sum_{x_{1,i}} P_{Y_i|X_1^{i-1}, X_2^i, Y^{i-1}, X_{1,i}, M_1, M_2} \times P_{X_{1,i}|X_1^{i-1}, X_2^i, Y^{i-1}, M_1, M_2} \quad (\text{A.12})$$

$$= \sum_{x_{1,i}} P_{Y_i|X_1^i, X_2^i, M_1, M_2} P_{X_{1,i}|X_1^{i-1}, X_2^i, M_1, M_2} \quad (\text{A.13})$$

$$= \sum_{x_{1,i}} P_{X_{1,i}, Y_i|X_1^{i-1}, X_2^i, M_1, M_2} \quad (\text{A.14})$$

$$= P_{Y_i|X_1^{i-1}, X_2^i, M_1, M_2} \quad (\text{A.15})$$

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:

$$P_{Y^{n''}|X_1^n, M_1, M_1'} = \sum_{x_2^{n''}} P_{Y^{n''}|X_1^n, X_2^{n''}, M_1, M_1'} P_{X_2^{n''}|X_1^n, M_1, M_1'} \quad (\text{A.16})$$

$$= \sum_{x_2^{n''}} P_{Y^{n''}|X_1^n, X_2^{n''}} P_{X_2^{n''}|X_1^n} \quad (\text{A.17})$$

$$= \sum_{x_{1,i}} P_{Y^{n''}, X_2^{n''}|X_1^n} \quad (\text{A.18})$$

$$= P_{Y^{n''}|X_1^n}. \quad (\text{A.19})$$

□

APPENDIX B PROOF OF LEMMA 1

Proof: In the following, we upper bound each constraint in (39), and show that that the upper bounds can be achieved by taking $V = X_1$. We have:

$$R_1 \leq I(V; Y|X_2) \quad (\text{B.1})$$

$$\leq I(V, X_1; Y|X_2) \quad (\text{B.2})$$

$$= I(X_1; Y|X_2), \quad (\text{B.3})$$

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

$$R_2 \leq I(X_2; Y|V) \quad (\text{B.4})$$

$$= H(X_2|V) - H(X_2|V, Y) \quad (\text{B.5})$$

$$\leq H(X_2|X_1) - H(X_2|X_1, Y) \quad (\text{B.6})$$

$$= I(X_2; Y|X_1) \quad (\text{B.7})$$

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

$$H(X_2|X_1, Y) = H(X_2|X_1, V, Y) \quad (\text{B.8})$$

$$\leq H(X_2|V, Y) \quad (\text{B.9})$$

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

$$P_{X_2, V|X_1, Y} = \frac{P_{X_1 X_2 Y} P_{V|X_1, X_2, Y}}{P_{X_1, Y}} \quad (\text{B.10})$$

$$= P_{X_2|X_1, Y} P_{V|X_1, X_2, Y} \quad (\text{B.11})$$

$$= P_{X_2|X_1, Y} P_{V|X_1, X_2} \quad (\text{B.12})$$

$$= P_{X_2|X_1, Y} P_{V|X_1} \quad (\text{B.13})$$

$$= P_{X_2|X_1, Y} P_{V|X_1, Y} \quad (\text{B.14})$$

where the third and last equalities follow from the relations $V \ominus (X_1, X_2) \ominus Y$ and $V \ominus X_1 \ominus Y$, respectively, which are true due to (31). For the sum rate, we have:

$$R_1 + R_2 \leq I(V, X_2; Y) \quad (\text{B.15})$$

$$\leq I(V, X_1, X_2; Y) \quad (\text{B.16})$$

$$= I(X_1, X_2; Y) \quad (\text{B.17})$$

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

$$R_2'' \leq I(X_2''; Y'' | X_1, V) \quad (\text{B.18})$$

$$= H(Y'' | X_1, V) - H(Y'' | X_1, V, X_2'') \quad (\text{B.19})$$

$$\leq H(Y'' | X_1) - H(Y'' | X_1, X_2'') \quad (\text{B.20})$$

$$= I(X_2''; Y'' | X_1) \quad (\text{B.21})$$

where the inequality follows from the fact that conditioning reduces entropy, and the relation $V \circlearrowleft (X_1, X_2'') \circlearrowleft Y''$. 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 $V = X$. \square

APPENDIX C

THE CAPACITY REGION IN EXAMPLE 3

First, note that for $i \in \{0, 1\}$ and $j \in \{0, 1, 2, 3\}$:

$$P_{Y_1}(0) = P_{X_1}(1 - q) + \bar{P}_{X_1}q, \quad (\text{C.1})$$

$$\begin{aligned} \nu_i &\triangleq \Pr\{X_1 = 0 | Y_1 = i\} \\ &= \frac{P_{X_1}(1 - q)^{1-i}q^i}{P_{X_1}(1 - q)^{1-i}q^i + \bar{P}_{X_1}(1 - q)^i q^{1-i}}, \end{aligned} \quad (\text{C.2})$$

$$\begin{aligned} \lambda_{ij} &\triangleq \Pr\{Y_2 = 0 | Y_1 = i, X_2 = j\} \\ &= \nu_i \left[\delta(j) + \frac{1}{2} [\delta(j - 2) + \delta(j - 3)] \right] \\ &\quad + \bar{\nu}_i \left[\delta(j - 2) + \frac{1}{2} [\delta(j) + \delta(j - 1)] \right], \end{aligned} \quad (\text{C.3})$$

and

$$\Pr\{Y_2 = 0 | Y_1 = i\} = \sum_{j=0}^3 p_j \lambda_{ij}. \quad (\text{C.4})$$

Then, it is easy to check that:

$$H(Y_2 | X_2, Y_1) = \sum_{i,j} P_{Y_1}(i) p_j \mathcal{H}_2(\lambda_{ij}), \quad (\text{C.5})$$

$$H(Y_2 | Y_1) = \sum_{i=0}^1 P_{Y_1}(i) \mathcal{H}_2 \left(\sum_j \lambda_{ij} \right), \quad (\text{C.6})$$

$$\begin{aligned} H(Y_2 | X_1, Y_1) &= H(Y_2 | X_1) \\ &= P_{X_1} \mathcal{H}_2 \left(p_0 + \frac{1}{2} (p_2 + p_3) \right) \\ &\quad + \bar{P}_{X_1} \mathcal{H}_2 \left(p_2 + \frac{1}{2} (p_0 + p_1) \right). \end{aligned} \quad (\text{C.7})$$

Using the above results and (42), we have have:

$$\begin{aligned} R_1 &\leq I(X_1; Y_1, Y_2 | X_2) \\ &= H(Y_1, Y_2 | X_2) - H(Y_1, Y_2 | X_1, X_2) \\ &= H(Y_2) + H(Y_2 | X_2, Y_1) - H(Y_1 | X_1) - H(Y_2 | X_1, X_2) \\ &= \mathcal{H}_2(P_{X_1} * q) - \mathcal{H}_2(q) + H(Y_2 | X_2, Y_1) \\ &\quad - P_{X_1}(P_2 + P_3) - \bar{P}_{X_1}(P_0 + P_1) \\ &\triangleq \mathcal{R}_1, \end{aligned} \quad (\text{C.8})$$

and

$$R_2 \leq H(Y_2 | X_1, Y_1) - P_{X_1}(P_2 + P_3) - \bar{P}_{X_1}(P_0 + P_1). \quad (\text{C.9})$$

For the sum rate, we get:

$$\begin{aligned} R_1 + R_2 &\leq I(X_1, X_2; Y_1, Y_2) \\ &= I(X_1; Y_1, Y_2 | X_2) + I(X_2; Y_1, Y_2) \\ &= \mathcal{R}_1 + H(Y_2 | Y_1) - H(Y_2 | X_2, Y_1). \end{aligned} \quad (\text{C.10})$$

Regarding R_2'' , choosing the distribution $P_{X_2'' | X_1}$ as in (46)-(47), we readily get that

$$R_2'' \leq 1, \quad (\text{C.11})$$

and

$$\begin{aligned} R_1 + R_2'' &\leq I(X_1, X_2''; Y_1, Y_2) \\ &= I(X_1; Y_1, Y_2) + I(X_2''; Y_1, Y_2 | X_1) \\ &= I(X_1; Y_1) + I(X_1; Y_2 | Y_1) + 1 \\ &= 1 + \mathcal{H}_2(P_{X_1} * q) - \mathcal{H}_2(q). \end{aligned} \quad (\text{C.12})$$

Therefore, we have obtain that the capacity region in Example 3 is:

$$R_1 \leq \mathcal{R}_1, \quad (\text{C.13a})$$

$$R_2 \leq H(Y_2 | X_1, Y_1) - P_{X_1}(P_2 + P_3) - \bar{P}_{X_1}(P_0 + P_1), \quad (\text{C.13b})$$

$$R_1 + R_2 \leq \mathcal{R}_1 + H(Y_2 | Y_1) - H(Y_2 | X_2, Y_1), \quad (\text{C.13c})$$

$$R_2'' \leq 1, \quad (\text{C.13d})$$

$$R_1 + R_2'' \leq 1 + \mathcal{H}_2(P_{X_1} * q) - \mathcal{H}_2(q), \quad (\text{C.13e})$$

where $H(Y_2 | X_1, Y_1)$, $H(Y_2 | Y_1)$, and $H(Y_2 | X_2, Y_1)$, are given in (C.5)-(C.7).

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] A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 851-857, November 1982.
- [5] F.-W. Fu and R. W. Yeung, "On the rate-distortion region for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 2012-2021, July 2002.
- [6] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 269-275, March 1991.
- [7] B. Rimoldi, "Successive refinement of information: Characterization of the achievable rates," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 253-259, January 1994.
- [8] L. Dikstein, H. Permuter, and Y. Steinberg, "On State-Dependent Degraded Broadcast Channels With Cooperation," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2308-2323, May 2016.
- [9] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 6, pp. 727-734, November 1985.

- [10] 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.
- [11] 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.
- [12] Y. Liang and V. V. Veeravalli, "Cooperative relay broadcast channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 900-1028, Mar. 2007.
- [13] 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.
- [14] 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.
- [15] Y. Steinberg, "Channels with cooperation links that may be absent," in *IEEE International Symposium on Information Theory Proceedings (ISIT 2014)*, Honolulu, HI, 2014.
- [16] W. Huleihel and Y. Steinberg, "Multiple Access Channel with Unreliable Cribbing," in *IEEE International Symposium on Information Theory Proceedings (ISIT 2016)*, Barcelona, 2016.
- [17] H. Asnani and H. H. Permuter, "Multiple-Access Channel With Partial and Controlled Cribbing Encoders," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2252-2266, Apr. 2013.
- [18] R. Kolte, A. Ozgur, H. Permuter, "Cooperative Binning for Semi-deterministic Channels," *Submitted to IEEE Trans. Inf. Theory*, Mar. 2016
- [19] 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.
- [20] 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.
- [21] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
- [22] T. M. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.