Capacity Region of Cooperative Multiple-Access Channel With States

Abdellatif Zaidi, Pablo Piantanida, and Shlomo Shamai (Shitz), *Fellow, IEEE*

***Abstract*—We consider a two-user state-dependent multiaccess channel in which the states of the channel are known noncausally to one of the encoders and only strictly causally to the other encoder. Both encoders transmit a common message and, in addition, the encoder that knows the states noncausally transmits an individual message. We find explicit characterizations of the capacity region of this communication model in both discrete memoryless and memoryless Gaussian cases. In particular, the capacity region analysis demonstrates the utility of the knowledge of the states only strictly causally at the encoder that sends only the common message in general. More specifically, in the discrete memoryless setting, we show that such a knowledge is beneficial and increases the capacity region in general. In the Gaussian setting, we show that such a knowledge does not help, and the capacity is same as if the states were completely unknown at the encoder that sends only the common message. Furthermore, we also study the special case in which the two encoders transmit only the common message and show that the knowledge of the states only strictly causally at the encoder that sends only the common message is not beneficial in this case, in both discrete memoryless and memoryless Gaussian settings. The analysis also reveals optimal ways of exploiting the knowledge of the state only strictly causally at the encoder that sends only the common message when such a knowledge is beneficial. The encoders collaborate to convey to the decoder a lossy version of the state, in addition to transmitting the information messages through a generalized Gel'fand–Pinsker binning. Particularly important in this problem are the questions of 1) optimal ways of performing the state compression and 2) whether or not the compression indices should be decoded uniquely. By developing two optimal coding schemes that perform this state compression differently, we show that when used as parts of appropriately tuned encoding and decoding processes, both compression à-la noisy network coding by Lim *et al.* or the quantize-map-and-forward by Avestimeher *et al.*, i.e., with no binning, and compression using Wyner–Ziv binning are optimal. The scheme that uses Wyner–Ziv binning shares elements with Cover and El Gamal original compress-and-forward, but differs from it mainly in that backward decoding is employed instead of forward decoding and the compression indices are not decoded uniquely. Finally, by exploring the properties of our outer bound, we show that, although not required in general, the compression indices can in fact be decoded uniquely essentially without altering the capacity region, but at the expense of larger alphabets sizes for the auxiliary random variables.**

***Index Terms*—Capacity, channel state information, multiaccess channels, noisy network coding, Wyner-Ziv binning.**

Manuscript received January 14, 2012; revised November 07, 2012; accepted April 20, 2013. Date of publication July 09, 2013; date of current version September 11, 2013. This work was supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications (NEWCOM#). This paper was presented in part at the 2011 and 2012 IEEE International Symposia on Information Theory.

A. Zaidi is with the Université Paris-Est Marne La Vallée, 77454 Marne la Vallée Cedex 2, France (e-mail: abdellatif.zaidi@univ-mlv.fr).

P. Piantanida is with the Department of Telecommunications, SUPELEC, 91190 Gif-sur-Yvette, France (e-mail: pablo.piantanida@supelec.fr).

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: sshlomo@ee.technion.ac.il).

Communicated by S. Diggavi, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2013.2265553

I. Introduction

THE study of channels that are controlled by random states has spurred much interest, due to its importance from both information-theoretic and communications aspects. For example, state-dependent channels may model communication in random fading environments [1] or in the presence of interference imposed by users in broadcast scenarios. The channel states may be known in a strictly causal, causal, or noncausal manner, to all or only a subset of the encoders. For a transmission of length n, let $S^n = (S_1, S_2, \ldots, S_n)$ denote the state sequence, with S_i representing the channel state affecting the channel at time or block i. For the transmission in block i, the state sequence is known noncausally if it is known entirely before the beginning of the transmission. It is known causally if it is known up to and including time i, and it is known strictly causally if it is known only up to time $i-1$. The way the channel state information is utilized and influences capacity depends also on which of the encoders(s) and decoder(s) are aware of it. In single-user channels, the concept of channel state available at only the transmitter dates back to Shannon [2] for the causal channel state case, and to Gel'fand and Pinsker [3] for the noncausal channel state case. In multiuser environments, a growing body of work studies multiuser state-dependent models. Recent advances in this regard can be found in [4]–[27], and many other works. For a comprehensive review of state-dependent channels and related work, the reader may refer to [4].

There is a connection between the role of states known strictly causally at an encoder and that of output feedback given to that encoder. In single-user channels, it is now well known that strictly causal feedback does not increase the capacity [28]. In multiuser channels or networks, however, the situation changes drastically, and output feedback can be beneficial—but its role is still highly misunderstood. One has a similar picture with strictly causal states at the encoder. In single-user channels, independent and identically distributed (i.i.d.) states available only in a strictly causal manner at the encoder have no effect on the capacity. In multiuser channels or networks, however, like feedback, strictly causal states in general increase the capacity.

Advances in the study of the effect of strictly causal states in multiuser channels are rather very recent and concern mainly multiple-access scenarios. In [15], Lapidoth and Steinberg study a two-encoder multiple-access channel (MAC) with independent messages and states known causally or strictly causally at the encoders. They show that the strictly causal state sequence can be beneficial, in the sense that it increases the capacity for this model. This result is reminiscent of Dueck's proof [29] that feedback can increase the capacity region of some broadcast channels. In accordance with [29], the main idea of the achievability result in [15] is a block Markov coding scheme in which the two users collaborate to describe the state to the decoder by sending cooperatively a compressed version of it. As noticed in [15], although some nonzero rate that otherwise could be used to transmit pure information is spent in describing the state to the decoder, the net effect can be an increase in the capacity. In [16], they show that strictly causal state information is beneficial even if the channel is controlled by two independent states each known to one encoder strictly causally. In this case, each encoder can help the other encoder transmit at a higher rate by sending a compressed version of its state to the decoder. In [18], Li *et al.* improve the results of [15] and [16] and extend them to the case of multiple encoders. The achievability results in [18] are inspired by the noisy network coding scheme by Lim *et al.* [30] and, unlike [15], [16], do not use Wyner–Ziv binning [31] for the compression of the state. In a very recent contribution [32], Lapidoth and Steinberg derive a new inner bound on the capacity region for the case of a single state governing the MAC. They also prove that the inner bound of [18] for the case of two independent states each known strictly causally to one encoder can indeed be strictly better than the lower bound of [15] and [16]—a result which is conjectured previously by Li *et al.* in [18].

The noisy network coding scheme by Lim *et al.* [30] extends the results on coding for deterministic networks and wireless Gaussian relay networks by Avestimeher *et al.* [33]. In particular, it extends the insights of 1) quantization only, 2) joint decoding of the message and quantization bits, and 3) repetitive encoding of messages at the source, that were developed originally by Avestimeher *et al.* in [33] for Gaussian relay networks in a scheme that they called "quantize-map and forward," to discrete memoryless networks. In [30], the authors also simplify the proofs and generalize the results to multiple multicast sessions. For Gaussian relay networks, the coding scheme of [33] has also been extended to lattice vector quantizers in [34].

A. Studied Model

In this paper, which generalizes former conference versions [35], [36], we study a two-user state-dependent MAC with the channel states known noncausally at one encoder and only strictly causally at the other encoder. The decoder is not aware of the channel states. As shown in Fig. 1, both encoders transmit a common message and, in addition, the encoder that knows the states noncausally transmits an individual message. This model generalizes one whose capacity region is established in [5] and in which the encoder that sends only the common message does not know the states at all. More precisely, let W_c and W_1 denote the common message and the individual message to be transmitted in, say, n uses of the channel; and $S^n = (S_1, \ldots, S_n)$ denote the state sequence affecting the channel during this time. At time i, Encoder 1 knows the complete sequence $S^n = (S_1, \ldots, S_{i-1}, S_i, \ldots, S_n)$ and sends $X_{1i} = \phi_1(W_c, W_1, S^n)$, and Encoder 2 knows *only* $S^{i-1} = (S_1, \ldots, S_{i-1})$ and sends $X_{2i} = \phi_{2,i}(W_c, S^{i-1})$—the functions ϕ_1 and $\phi_{2,i}$ are some encoding functions. In this paper, we study the capacity region of this state-dependent MAC model. As our analysis will show, this requires, among others, understanding the role of the strictly causal part of the state that is revealed to Encoder 2.

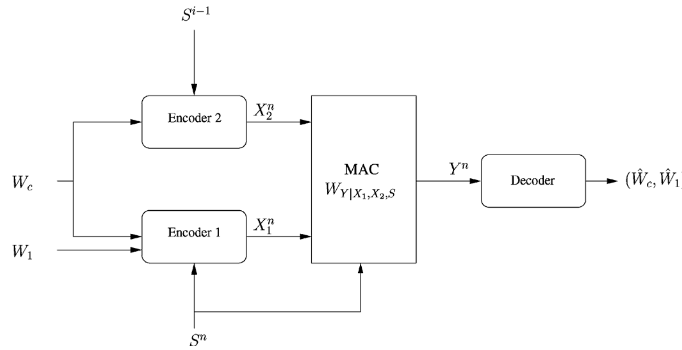

Fig. 1. State-dependent MAC with degraded message sets and states known noncausally at the encoder that sends both messages and only strictly causally at the other encoder.

From an application viewpoint, the state in the model of Fig. 1 can, for example, represent another message, not related to the system, and known beforehand to Encoder 1 (who, say, monitors the backhaul). However, this message is only received (essentially noiselessly, due to proximity), by Encoder 2, who does not know the codebook, and hence cannot decode that message.

B. Main Contributions

In the discrete memoryless case, we characterize the capacity region for the general finite-alphabet case with a single-letter expression. The proof of the achievability part is based on a block-Markov coding scheme in which the two encoders collaborate to convey a lossy version of the state to the decoder, in the spirit of [15], [16], and [32], in addition to a generalized Gel'fand–Pinsker binning for the transmission of the information messages [3]. From the angle of the state compression, coding schemes that perform the state compression for our model tie with very recent works on compressions in compress-and-forward-type relaying networks [30], [33], [37]–[39]. We first develop a coding scheme in which the state compression is performed à-la Lim *et al.* noisy network coding [30] and Avestimeher *et al.* quantize-map-and-forward [33], and show that it is optimal, i.e., achieves an outer bound that we establish for the studied model. In this coding scheme, unlike [15], [16], [32] where every information message is divided into blocks and different submessages are sent over these blocks and then decoded one at a time using the same codebook as in the original compress-and-forward scheme by Cover and El Gamal [40], here the *entire* common message and the *entire* individual message are transmitted over *all* blocks using codebooks that are generated independently, one for each block, and the decoding is performed simultaneously using all blocks as in the noisy network coding scheme of [30] or the quantize-map-and-forward scheme of [33]. Also, like [30] and [33], at each block, the compression index of the state of the previous block is sent using standard rate distortion, not Wyner–Ziv binning. At the end of the transmission, the receiver uses the outputs of all blocks to perform simultaneous decoding of the information common and individual messages, without uniquely decoding the compression indices. From this angle, our coding scheme connects more with [18], than with [15], [16], and [32].

Two of the most important features of our coding scheme are 1) standard compression without Wyner–Ziv binning and 2) nonexplicit decoding of the compression indices. Investigating whether these features are pivotal for optimality in our problem, as argued in [30] for some related models, we also explore binning-based compressions. We show that the capacity region of our model can also be achieved using an alternate coding scheme in which the state compression is realized using Wyner–Ziv binning. The employed optimal alternate coding scheme shares elements with Cover and El Gamal compress-and-forward [40], but differs from it in two aspects: 1) backward decoding is utilized instead of the forward decoding of [40], and 2) unlike [40], the compression indices are not decoded uniquely. Decoding backwardly instead of forwardly seems essential for the optimality of this alternate coding scheme here. At this level, we note that the fact that backward decoding with nonunique decoding of the compression indices is beneficial has also been observed independently in [41] in the context of unicast relay networks and in [42] for a fading relay network. Next, by exploring our outer bound further, we show that, although not required, one can modify this coding scheme in a manner to get the compression indices decoded at the receiver essentially without altering the capacity region but at the expense of larger alphabets sizes of the involved auxiliary random variables. The decoding of the compression indices introduces an additional rate constraint, but we show that this constraint is satisfied by the auxiliary random variables of the outer bound.

The single-letter characterization of the capacity region of our model remains intact if one allows feedback to the encoder that sends both messages. Also, the capacity region of our model contains that of the model of [5] in which the encoder that sends only the common message is unaware of the channel states, and this shows that revealing the states even only strictly causally to this encoder potentially increases the capacity region. Next, by investigating a discrete memoryless example, we show that this inclusion can be strict, thus demonstrating the utility of conveying a compressed version of the state to the decoder cooperatively by the encoders.

We also specialize our results to the case in which the two encoders send only the common message. We refer to the capacity in this case as *common-message* capacity. We show that, when one of the two encoders is informed noncausally, the knowledge of the states only strictly causally at the other encoder does not increase the common-message capacity. It should be noted that this result is not a direct consequence of that feedback does not increase the capacity in a MAC in which the encoders send only a common message, and our converse proof is needed here.

Next, we consider the memoryless Gaussian setting in which the channel state and the noise are additive and Gaussian. We establish an operative outer bound on the achievable rate pairs. Then, we show that this outer bound is achievable, yielding a closed-form expression of the capacity region. The resulting capacity region coincides with that of the model of [5] in which the encoder that sends only the common message is completely unaware of the states, thus demonstrating that, by opposition to the discrete memoryless case, revealing the states strictly causally to this encoder is not beneficial in the Gaussian case, in the sense that it does not increase the capacity region.

Finally, we note that in contrast to the related MAC models in [5] and [7], our converse proofs in this paper do not follow directly from the converse part proof of the capacity formula for the standard Gel'fand–Pinsker channel [3]. This is because, at time i, the encoder that transmits only the common message sends inputs which are function of not only that message, but also the observed past state sequence.

C. Outline and Notation

An outline of the remainder of this paper is as follows. Section II describes in more detail the communication model that we consider in this study. Section III provides the capacity region of the discrete memoryless model. In this section, we also establish an alternative outer bound on the capacity region that will turn to be useful in the Gaussian case, provide an example demonstrating the utility of revealing the states only strictly causally to the encoder that sends only the common message, and derive the common-message capacity. Section IV characterizes the capacity region as well as the common-message capacity of the Gaussian model. Finally, Section V concludes this paper.

We use the following notations throughout the paper. Uppercase letters are used to denote random variables, e.g., X; lowercase letters are used to denote realizations of random variables, e.g., x; and calligraphic letters designate alphabets, i.e., $\mathcal{X}$. The probability distribution of a random variable X is denoted by $P_X(x)$. Sometimes, for convenience, we write it as P_X. We use the notation $\mathbb{E}_X[\cdot]$ to denote the expectation of random variable X. A probability distribution of a random variable Y given X is denoted by $P_{Y|X}$. The set of probability distributions defined on an alphabet $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. The cardinality of a set $\mathcal{X}$ is denoted by $|\mathcal{X}|$. For convenience, the length n vector x^n will occasionally be denoted in boldface notation $\mathbf{x}$. The Gaussian distribution with mean μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$. For integers $i \leq j$, we define $[i:j] := \{i, i+1, \ldots, j\}$. Finally, throughout the paper, logarithms are taken to base 2, and the complement to unity of a scalar $u \in [0,1]$ is denoted by $\bar{u}$, i.e., $\bar{u} = 1 - u$.

II. System Model and Definitions

We consider a stationary memoryless state-dependent MAC $W_{Y|X_1,X_2,S}$ whose output $Y \in \mathcal{Y}$ is controlled by the channel inputs $X_1 \in \mathcal{X}_1$ and $X_2 \in \mathcal{X}_2$ from the encoders and the channel state $S \in \mathcal{S}$ which is drawn according to a memoryless probability law Q_S. We assume that the channel state S^n is known noncausally at Encoder 1, i.e., beforehand, at the beginning of the transmission block. Encoder 2 knows the channel states only strictly causally; that is, at time i, it knows the states only up to time $i-1$, $S^{i-1} = (S_1, \ldots, S_{i-1})$.

Encoder 2 wants to send a common message W_c and Encoder 1 wants to send an independent individual message W_1 along with the common message W_c. We assume that the common message W_c and the individual message W_1 are independent random variables drawn uniformly from the sets $\mathcal{W}_c = \{1, \ldots, M_c\}$ and $\mathcal{W}_1 = \{1, \ldots, M_1\}$, respectively.

The sequences X_1^n and X_2^n from the encoders are sent across a state-dependent MAC modeled as a memoryless conditional probability distribution $W_{Y|X_1,X_2,S}$. The joint probability mass function on $\mathcal{W}_c\times\mathcal{W}_1\times\mathcal{S}^n\times\mathcal{X}_1^n\times\mathcal{X}_2^n\times\mathcal{Y}^n$ is given by (1). The receiver guesses the pair $(\hat{W}_c, \hat{W}_1)$ from the channel output Y^n.

Definition 1: For positive integers n, M_c, and M_1, an (M_c, M_1, n, ϵ) code for the MAC with states known noncausally at one encoder and only strictly causally at the other encoder consists of a mapping

$$\phi_1 : \mathcal{W}_c\times\mathcal{W}_1\times\mathcal{S}^n \longrightarrow \mathcal{X}_1^n \tag{2}$$

at Encoder 1, a sequence of mappings

$$\phi_{2,i} : \mathcal{W}_c\times\mathcal{S}^{i-1} \longrightarrow \mathcal{X}_2, \quad i = 1,\dots,n \tag{3}$$

at Encoder 2, and a decoder map

$$\psi : \mathcal{Y}^n \longrightarrow \mathcal{W}_c\times\mathcal{W}_1 \tag{4}$$

such that the average probability of error is bounded by ϵ,

$$P_e^n = \mathbb{E}_{S^n}\left[\Pr\big(\psi(Y^n) \neq (W_c, W_1)|S^n = s^n\big)\right] \leq \epsilon. \tag{5}$$

The rate of the common message and the rate of the individual message are defined as

$$R_c = \frac{1}{n}\log M_c \qquad \text{and} \qquad R_1 = \frac{1}{n}\log M_1, \tag{6}$$

respectively.

A rate pair (R_c, R_1) is said to be achievable if for every $\epsilon > 0$, there exists an $(2^{nR_c}, 2^{nR_1}, n, \epsilon)$ code for the channel $W_{Y|X_1,X_2,S}$. The capacity region of the considered state-dependent MAC is defined as the closure of the set of achievable rate pairs.

III. Discrete Memoryless Case

In this section, it is assumed that the alphabets $\mathcal{S}, \mathcal{X}_1, \mathcal{X}_2$ are finite.

A. Capacity Region

Let $\mathcal{P}$ stand for the collection of all random variables (S, U, V, X_1, X_2, Y) such that U, V, X_1, and X_2 take values in finite alphabets $\mathcal{U}$, $\mathcal{V}$, $\mathcal{X}_1$, and $\mathcal{X}_2$, respectively, and

$$\begin{aligned}&P_{S,U,V,X_1,X_2,Y}(s,u,v,x_1,x_2,y)\\&= P_{S,U,V,X_1X_2}(s,u,v,x_1,x_2)W_{Y|X_1,X_2,S}(y|x_1,x_2,s) \quad (7a)\\&P_{S,U,V,X_1,X_2}(s,u,v,x_1,x_2)\\&=Q_S(s)P_{X_2}(x_2)P_{V|S,X_2}(v|s,x_2)P_{U,X_1|S,V,X_2}(u,x_1|s,v,x_2) \quad (7b)\\&\sum_{u,v,x_1,x_2} P_{S,U,V,X_1,X_2}(s,u,v,x_1,x_2) = Q_S(s). \quad (7c)\end{aligned}$$

The relations in (7) imply that $(U, V) \leftrightarrow (S, X_1, X_2) \leftrightarrow Y$ is a Markov chain, and X_2 is independent of S.

Define $\mathcal{C}$ to be the set of all rate pairs (R_c, R_1) such that

$$\begin{aligned}R_1 &\leq I(U;Y|V,X_2) - I(U;S|V,X_2)\\R_c + R_1 &\leq I(U,V,X_2;Y) - I(U,V,X_2;S)\\&\qquad \text{for some } (S,U,V,X_1,X_2,Y) \in \mathcal{P}.\end{aligned} \tag{8}$$

The following proposition states some properties of $\mathcal{C}$.

Proposition 1:

1. The set $\mathcal{C}$ is convex.
2. To exhaust $\mathcal{C}$, it is enough to restrict $\mathcal{V}$ and $\mathcal{U}$ to satisfy

$$|\mathcal{V}| \leq |\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 1 \quad (9a)$$
$$|\mathcal{U}| \leq \Big(|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 1\Big)|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2|. \quad (9b)$$

Proof: The proof of Proposition 1 appears in Appendix A.

As stated in the following theorem, the set $\mathcal{C}$ characterizes the capacity region of the state-dependent discrete memoryless MAC model that we study.

Theorem 1: The capacity region of the MAC with states known only strictly causally at the encoder that sends the common message and noncausally at the encoder that sends both messages is given by $\mathcal{C}$.

Proof: An outline proof of the coding scheme that we use for the direct part will follow. The associated error analysis and the proof of the converse appear in Appendix B.

Theorem 1 continues to hold if in (7) we replace $P_{U,X_1|S,V,X_2}$ by $P_{U,X_1|S,V}$. Also, it should be noted that setting $V = \emptyset$ in

$$\begin{aligned}P(w_c,w_1,s^n,x_1^n,x_2^n,y^n) &= P(w_c)P(w_1)\\&\times\prod_{i=1}^{n} Q_S(s_i)P(x_{1,i}|w_c,w_1,s^n)P(x_{2,i}|w_c,s^{i-1})W_{Y|X_1,X_2,S}(y_i|x_{1,i},x_{2,i},s_i)\end{aligned} \tag{1}$$

(8), the capacity region $\mathcal{C}$ reduces to the union of all rate pairs (R_c, R_1) satisfying

$$\begin{aligned} R_1 &\leq I(U;Y|X_2) - I(U;S|X_2) \\ R_c + R_1 &\leq I(U,X_2;Y) - I(U,X_2;S) \end{aligned} \tag{10}$$

for some measure on $\mathcal{S}\times\mathcal{U}\times\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{Y}$ of the form

$$P_{S,U,X_1,X_2,Y} = Q_S P_{X_2} P_{U,X_1|S,X_2} W_{Y|X_1,X_2,S}. \tag{11}$$

Let $\mathcal{C}'$ denote the region defined by (10) and (11) in the remaining of this paper. It has been shown in [5] that the region $\mathcal{C}'$ is the capacity region of the MAC model of Fig. 1 but with the states completely unknown at Encoder 2, i.e., while the encoding at Encoder 1 is given by (2), the encoding at Encoder 2 is defined by the mapping

$$\phi_2 : \mathcal{W}_c \longrightarrow \mathcal{X}_2^n. \tag{12}$$

Observing that $\mathcal{C}' \subseteq \mathcal{C}$ shows that the knowledge of the states only strictly causally at Encoder 2 in our model in general increases the capacity region. In Section III-B, we will show that the inclusion can be *strict*, i.e., $\mathcal{C}' \subsetneq \mathcal{C}$.

Furthermore, one can easily check that in the case of a channel that does not depend on the states, i.e., $W_{Y|X_1,X_2,S} = W_{Y|X_1,X_2}$, the capacity region $\mathcal{C}$ reduces to the closure of the union of all rate pairs (R_c, R_1) satisfying

$$\begin{aligned} R_1 &\leq I(X_1;Y|Z,X_2) \\ R_c + R_1 &\leq I(X_1,X_2;Y) \end{aligned} \tag{13}$$

for some

$$P_{Z,X_1,X_2,Y} = P_Z P_{X_1|Z} P_{X_2|Z} W_{Y|X_1,X_2}. \tag{14}$$

Also, it is noted that Theorem 1 remains intact if we allow feedback to Encoder 1, i.e., before producing the ith channel input symbol, Encoder 1 also observes the past channel output sequence Y^{i-1}. That is, the encoding at Encoder 2 is still given by (3) and that at Encoder 1 is replaced by a sequence of mappings $\{\phi_{1,i}\}_{i=1}^n$, with

$$\phi_{1,i} : \mathcal{W}_c\times\mathcal{W}_1\times\mathcal{S}^n\times\mathcal{Y}^{i-1} \longrightarrow \mathcal{X}_1. \tag{15}$$

We now turn to the proof of achievability of Theorem 1. The following remark is useful for a better understanding of the coding scheme that we use to establish the achievability of Theorem 1.

Remark 1: The proof of achievability of Theorem 1 is based on a block-Markov coding scheme in which a lossy version of the state is conveyed to the decoder, in the spirit of [15], [16], and [32], in addition to a generalized Gel'fand–Pinsker binning for the transmission of the information messages [3]. However, unlike [15], [16], and [32] where Wyner–Ziv compression [31] is utilized for the transmission of the lossy version of the state, here, inspired by the noisy network coding scheme of [30] and the quantize-map-and-forward scheme of [33], at each block, the compression index of the state of the previous block is sent using standard rate distortion, not Wyner–Ziv binning. Also, unlike [15], [16], and [32] where every information message is divided into blocks and different submessages are sent over these blocks and then decoded one at a time using the same codebook as in the original compress-and-forward scheme by Cover and El Gamal [40], here the *entire* common message and the *entire* individual message are transmitted over *all* blocks using codebooks that are generated independently, one for each block, and the decoding is performed simultaneously using all blocks as in [30] and [33]. At the end of the transmission, the receiver uses the outputs of all blocks to perform simultaneous decoding of the information common and individual messages, without uniquely decoding the compression indices. □

Proof of Achievability: The transmission takes place in B blocks. The common message W_c and the individual message W_1 are sent over *all* blocks. We thus have $B_{W_c} = nBR_c$, $B_{W_1} = nBR_1$, $N = nB$, $R_{W_c} = B_{W_c}/N = R_c$, and $R_{W_1} = B_{W_1}/N = R_1$, where B_{W_c} is the number of common message bits, B_{W_1} is the number of individual message bits, N is the number of channel uses, and R_{W_c} and R_{W_1} are the overall rates of the common and individual messages, respectively.

Codebook Generation: Fix a measure $P_{S,U,V,X_1,X_2,Y} \in \mathcal{P}$. Fix $\epsilon > 0$, $\eta_c > 0$, $\eta_1 > 0$, $\hat{\eta} > 0$, $\delta > 1$ and denote $M_c = 2^{nB[R_c - \eta_c\epsilon]}$, $M_1 = 2^{nB[R_1 - \eta_1\epsilon]}$, $\hat{M} = 2^{n[\hat{R} + \hat{\eta}\epsilon]}$ and $J = 2^{n[I(U;S|V,X_2)+\delta\epsilon]}$.

We randomly and independently generate a codebook for each block.

1) For each block i, $i = 1, \ldots, B$, we generate $M_c\hat{M}$ i.i.d. codewords $\mathbf{x}_{2,i}(w_c, t_i')$ indexed by $w_c = 1, \ldots, M_c$, $t_i' = 1, \ldots, \hat{M}$, each with i.i.d. components drawn according to P_{X_2}.
2) For each block i, for each codeword $\mathbf{x}_{2,i}(w_c, t_i')$, we generate $\hat{M}$ i.i.d. codewords $\mathbf{v}_i(w_c, t_i', t_i)$ indexed by $t_i = 1, \ldots, \hat{M}$, each with i.i.d. components drawn according to $P_{V|X_2}$.
3) For each block i, for each pair of codewords $(\mathbf{x}_{2,i}(w_c, t_i'), \mathbf{v}_i(w_c, t_i', t_i))$, we generate a collection of JM_1 i.i.d. codewords $\{\mathbf{u}_i(w_c, t_i', t_i, w_1, j_i)\}$ indexed by $w_1 = 1, \ldots, M_1$, $j_i = 1, \ldots, J$, each with i.i.d. components draw according to $P_{U|V,X_2}$.

Encoding: Suppose that a common message $W_c = w_c$ and an individual message $W_1 = w_1$ are to be transmitted. As we mentioned previously, w_c and w_1 will be sent over *all* blocks. We denote by $\mathbf{s}[i]$ the state affecting the channel in block i, $i = 1, \ldots, B$. For convenience, we let $\mathbf{s}[0] = \emptyset$ and $t_{-1} = t_0 = 1$ (a default value). The encoding at the beginning of block i, $i = 1, \ldots, B$, is as follows.

Encoder 2, which has learned the state sequence $\mathbf{s}[i-1]$, knows t_{i-2} and looks for a compression index $t_{i-1} \in [1 : \hat{M}]$ such that $\mathbf{v}_{i-1}(w_c, t_{i-2}, t_{i-1})$ is strongly jointly typical with $\mathbf{s}[i-1]$ and $\mathbf{x}_{2,i-1}(w_c, t_{i-2})$. If there is no such index or the observed state $\mathbf{s}[i-1]$ is not typical, t_{i-1} is set to 1 and an error is declared. If there is more than one such index t_{i-1}, choose the smallest. Encoder 2 then transmits the vector $\mathbf{x}_{2,i}(w_c, t_{i-1})$.

Encoder 1 obtains $\mathbf{x}_{2,i}(w_c, t_{i-1})$ similarly. It then finds the smallest compression index $t_i \in [1 : \hat{M}]$ such that

$\mathbf{v}_i(w_c, t_{i-i}, t_i)$ is strongly jointly typical with $\mathbf{s}[i]$ and $\mathbf{x}_{2,i}(w_c, t_{i-1})$. Again, if there is no such index or the observed state $\mathbf{s}[i]$ is not typical, t_i is set to 1 and an error is declared. Next, Encoder 1 looks for the smallest j_i such that $\mathbf{u}_i(w_c, t_{i-1}, t_i, w_1, j_i)$ is jointly typical with $\mathbf{s}[i]$ given $(\mathbf{x}_{2,i}(w_c, t_{i-1}), \mathbf{v}_i(w_c, t_{i-1}, t_i))$. Denote this j_i by $j_i^\star = j(\mathbf{s}[i], w_c, t_{i-1}, t_i, w_1)$. If such $j_i^\star$ is not found, an error is declared and $j(\mathbf{s}[i], w_c, t_{i-1}, t_i, w_1)$ is set to $j_i = J$. Encoder 1 then transmits a vector $\mathbf{x}_1[i]$ which is drawn i.i.d. conditionally given $\mathbf{u}_i(w_c, t_{i-1}, t_i, w_1, j_i^\star)$, $\mathbf{s}[i]$, $\mathbf{v}_i(w_c, t_{i-1}, t_i)$ and $\mathbf{x}_{2,i}(w_c, t_{i-1})$ [using the conditional measure $P_{X_1|U,S,V,X_2}$ induced by (7)].

Decoding: At the end of the transmission, the decoder has collected all the blocks of channel outputs $\mathbf{y}[1], \ldots, \mathbf{y}[B]$.

Step (a): The decoder estimates message w_c using all blocks $i = 1, \ldots, B$, i.e., simultaneous decoding. It declares that $\hat{w}_c$ is sent if there exist $t^B = (t_1, \ldots, t_B) \in [1 : \hat{M}]^B$, $w_1 \in [1 : M_1]$ and $j^B = (j_1, \ldots, j_B) \in [1 : J]^B$ such that $\mathbf{x}_{2,i}(\hat{w}_c, t_{i-1})$, $\mathbf{u}_i(\hat{w}_c, t_{i-1}, t_i, w_1, j_i)$, $\mathbf{v}_i(\hat{w}_c, t_{i-1}, t_i)$ and $\mathbf{y}[i]$ are jointly typical for all $i = 1, \ldots, B$. One can show that the decoder obtains the correct w_c as long as n and B are large and

$$R_c + R_1 \leq I(U, V, X_2; Y) - I(U, V, X_2; S). \tag{16}$$

Step (b): Next, the decoder estimates message w_1 using again all blocks $i = 1, \ldots, B$, i.e., simultaneous decoding. It declares that $\hat{w}_1$ is sent if there exist $t^B = (t_1, \ldots, t_B) \in [1 : \hat{M}]^B$, $j^B = (j_1, \ldots, j_B) \in [1 : J]^B$ such that $\mathbf{x}_{2,i}(\hat{w}_c, t_{i-1})$, $\mathbf{u}_i(\hat{w}_c, t_{i-1}, t_i, \hat{w}_1, j_i)$, $\mathbf{v}_i(\hat{w}_c, t_{i-1}, t_i)$ and $\mathbf{y}[i]$ are jointly typical for all $i = 1, \ldots, B$. One can show that the decoder obtains the correct w_1 as long as n and B are large and

$$R_1 \leq I(U; Y|V, X_2) - I(U; S|V, X_2) \tag{17a}$$
$$R_1 \leq I(U, V, X_2; Y) - I(U, V, X_2; S). \tag{17b}$$

□

Remark 2: In the coding scheme of Theorem 1, the same message is sent over all blocks, i.e., message repetitive encoding, and the decoding is performed jointly using all blocks. One can modify this coding scheme in such a way that every message is divided into blocks and different submessages are sent over these blocks, and the decoder utilizes step-by-step backward decoding. The modified scheme achieves the same rate region as that of the coding scheme of Theorem 1. This is in accordance with the observation made in the parallel and independent work [39] that "short"-message encoding combined with backward decoding performs the same rates as noisy network coding and quantize-map-and-forward.

In the coding scheme of Theorem 1, the state compression is standard, i.e., uses no Wyner–Ziv binning. Although of no benefit in the case of one relay, together with repetitive encoding and joint nonunique decoding, this was shown to be essential in achieving rates that are strictly larger than those offered by schemes based on Cover and El Gamal classic compress-and-forward scheme [40] for certain networks with multiple relays in [30] and [33]. That is, the coding schemes of [30] and [33] outperform Cover and El Gamal classic compress-and-forward for some multirelay networks. One can wonder whether the same holds for our model, i.e., whether schemes based on Cover and El Gamal classic compress-and-forward, i.e., block Markov encoding combined with Wyner–Ziv binning, fall short of achieving optimality for our model. In this paper, we show that the capacity region $\mathcal{C}$ as given by (8) can be achieved alternatively with a coding scheme that we obtain by building upon and modifying Cover and El Gamal original compress-and-forward scheme. The modification consists essentially in 1) decoding block-by-block backwardly instead of block-by-block forwardly and 2) nonunique decoding of the compression indices. (In fact, by investigating more closely the converse proof of Theorem 1, we will show later that 2) can be relaxed essentially without altering the capacity region). The following theorem states the result.

Theorem 2: For the state-dependent MAC model that we study, there exists an optimal coding scheme that uses Wyner–Ziv binning for the state compression. That is, the capacity region $\mathcal{C}$ given by (8) can also be achieved using a coding scheme in which the state compression is performed using Wyner–Ziv binning.

Proof: The achievability proof of Theorem 2 is based on a block-Markovian coding scheme that combines carefully Gel'fand–Pinsker binning and Wyner–Ziv binning, and utilizes backward decoding with nonunique decoding of the compression indices. The complete proof of Theorem 2 is given in Appendix C.

As we mentioned previously, the coding scheme of Theorem 2 shares elements with Cover and El Gamal original compress-and-forward [40, Th. 7], but differs from it mainly in two aspects. First, it uses backward decoding instead of the forward decoding of [40], and, second, unlike [40], it does not require unique decoding of the compression indices. The second aspect is essential for getting the *same* rate expression as in (8), with no additional constraints. However, as we will see shortly in the corollary that will follow, one can modify the coding scheme of Theorem 2 in a way to get the compression indices decoded uniquely and *still* get the capacity region, at the expense of slightly larger $|\mathcal{V}|$ and larger $|\mathcal{U}|$. The key element is the observation that the constraint introduced by getting the compression index decoded, i.e., (see Appendix D)

$$I(V; S|X_2) - I(V; Y|X_2) \leq I(X_2; Y), \tag{18}$$

or, equivalently,

$$I(V, X_2; Y) - I(V, X_2; S) \geq 0, \tag{19}$$

is also *implicit* in the converse proof of Theorem 1. That is, the auxiliary random variables U and V of the converse proof of Theorem 1 in Appendix B satisfy (19).

Corollary 1: The coding scheme of Theorem 2 can be modified in a way to get the compression index decoded. The resulting coding scheme is optimal and achieves an equivalent characterization of the capacity region of the model that we study given by the set of all rate pairs (R_c, R_1) such that

$$\begin{aligned} R_1 &\leq I(U; Y|V, X_2) - I(U; S|V, X_2) \\ R_c + R_1 &\leq I(U, V, X_2; Y) - I(U, V, X_2; S) \end{aligned} \tag{20}$$

for some measure $(S, U, V, X_1, X_2, Y) \in \mathcal{P}$ and satisfying

$$I(V, X_2; Y) - I(V, X_2; S) \geq 0, \tag{21}$$

where the auxiliary random variables V and U have their alphabets bounded as

$$|\mathcal{V}| \leq |\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 2 \tag{22a}$$
$$|\mathcal{U}| \leq \Big(|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2| + 2\Big)|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2|. \tag{22b}$$

Proof: The coding scheme that we use for the proof of Corollary 1 is very similar to that of Theorem 2, but with unique decoding of the compression indices. The details of the proof are given in Appendix D.

We now establish an alternative outer bound on the capacity region of the DM MAC model that we study. This outer bound will turn out to be useful in the proof of the converse part of the coding theorem for the Gaussian case in Section IV since, as it will be shown, it is also achievable in that case.

Theorem 3: The capacity region of the MAC with states known noncausally at the encoder that sends both messages and only strictly causally at the other encoder is contained in the closure of the set of all rate pairs (R_c, R_1) satisfying

$$\begin{aligned} R_1 &\leq I(X_1; Y|S, X_2) \\ R_c + R_1 &\leq I(X_1, X_2; Y|S) - I(X_2; S|Y), \end{aligned} \tag{23}$$

for some probability distribution of the form

$$P_{S,X_1,X_2,Y} = Q_S P_{X_2} P_{X_1|X_2,S} W_{Y|X_1,X_2,S}. \tag{24}$$

Proof: The proof of Theorem 3 appears in Appendix E.

Remark 3: In [5], the authors use an extension of the converse part of the proof of the standard Gel'fand–Pinsker capacity to establish a converse proof for the model with states S^n known noncausally at Encoder 1 and no states at all at Encoder 2. Then, they show that their outer bound, which involves an auxiliary random variable, is itself contained in the region defined by (23). In Appendix E, we provide a direct proof that the region defined by (23) is an outer bound on the capacity region of the more general model that we study here. Our converse proof accounts also for the availability of the states at Encoder 2 in a strictly causal manner. □

B. Example

In Section III-A, we have shown that the capacity region $\mathcal{C}$ of the model of Fig. 1 is potentially larger than that, $\mathcal{C}'$, of the same model but with Encoder 2 being totally unaware of the states, i.e., $\mathcal{C}' \subseteq \mathcal{C}$. In this section, we show that this inclusion can be *strict*, i.e., $\mathcal{C}' \subsetneq \mathcal{C}$.

We use $h(\alpha)$ to denote the entropy of a Bernoulli (α) source, i.e.,

$$h(\alpha) = -\alpha \log(\alpha) - (1-\alpha)\log(1-\alpha) \tag{25}$$

and $p * q$ to denote the binary convolution, i.e.,

$$p * q = p(1-q) + q(1-p). \tag{26}$$

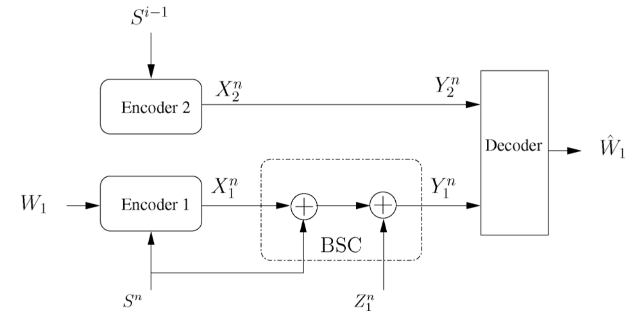

Fig. 2. Binary state-dependent MAC example with two output components, $Y^n = (Y_1^n, Y_2^n)$, with $Y_1^n = X_1^n + S^n + Z_1^n$ and $Y_2^n = X_2^n$.

Consider the binary memoryless MAC shown in Fig. 2. Here, all the random variables are binary $\{0, 1\}$. The channel has two output components, i.e., $Y^n = (Y_1^n, Y_2^n)$. The component Y_2^n is deterministic, $Y_2^n = X_2^n$, and the component $Y_1^n = X_1^n + S^n + Z_1^n$, where the addition is modulo 2. Encoder 2 knows the states only strictly causally and has no message to transmit. Encoder 1 knows the states noncausally and transmits an individual message W_1. The state and noise vectors are independent and memoryless, with the state process S_i, $i \geq 1$, and the noise process $Z_{1,i}$, $i \geq 1$, assumed to be Bernoulli $(\frac{1}{2})$ and Bernoulli (p) processes, respectively. The vectors X_1^n and X_2^n are the channel inputs, subjected to the constraints

$$\sum_{i=1}^{n} X_{1,i} \leq nq_1 \quad \text{and} \quad \sum_{i=1}^{n} X_{2,i} \leq nq_2, \quad q_2 \geq 1/2. \tag{27}$$

For this example, as we will show shortly, the strictly causal knowledge of the states at Encoder 2 *does* help, and in fact, Encoder 1 can transmit at rates that are larger than the standard Gel'fand–Pinsker $I(U; Y_1) - I(U; S)$ which would be the capacity had Encoder 2 been of no help.

Claim 1: The capacity of the state-dependent binary memoryless MAC shown in Fig. 2 is given by

$$C_B = \max_{p(x_1|s)} I(X_1; Y_1|S). \tag{28}$$

Proof:

1) The achievability follows from Theorem 1, as follows. Set $R_c = 0$ and $V = S$, $U = X_1$, $Y_2 = X_2$ with X_2 independent of (S, X_1) in Theorem 1. Evaluating the first inequality, we obtain

$$\begin{aligned} R_1 &\leq I(U; Y|V, X_2) - I(U; S|V, X_2) && (29) \\ &= I(X_1; Y_1, X_2|S, X_2) && (30) \\ &= I(X_1; Y_1|S, X_2) && (31) \\ &= I(X_1, X_2; Y_1|S) - I(X_2; Y_1|S) && (32) \\ &= I(X_1; Y_1|S) + I(X_2; Y_1|X_1, S) - I(X_2; Y_1|S) && (33) \\ &= I(X_1; Y_1|S) - I(X_2; Y_1|S) && (34) \\ &= I(X_1; Y_1|S), && (35) \end{aligned}$$

where (34) follows since $X_2 = Y_2$ and $Y_2 \leftrightarrow (X_1, S) \leftrightarrow Y_1$ is a Markov chain, and the last equality follows by the Markov relation $X_2 \leftrightarrow S \leftrightarrow Y_1$ for this example.

Evaluating the second inequality, we obtain

$$R_1 \leq I(U,V,X_2;Y) - I(U,V,X_2;S) \quad (36)$$
$$= I(X_1,S;Y_1,X_2) + H(X_2|X_1,S) - H(S) \quad (37)$$
$$= I(X_1,S;Y_1)+I(X_1,S;X_2|Y_1)+H(X_2|X_1,S)-H(S) \quad (38)$$
$$= I(X_1,S;Y_1) + H(X_2|Y_1) - H(X_2|X_1,S,Y_1) + H(X_2|X_1,S) - H(S) \quad (39)$$
$$= I(X_1;Y_1|S) + I(S;Y_1) + H(X_2|Y_1) - H(S) \quad (40)$$
$$= I(X_1;Y_1|S) + H(X_2|Y_1) - H(S|Y_1) \quad (41)$$
$$= I(X_1;Y_1|S) + H(Y_1|X_2) - H(Y_1|S)+H(X_2)-H(S) \quad (42)$$
$$= I(X_1;Y_1|S) + I(S;Y_1) + H(X_2) - H(S) \quad (43)$$

where (40) follows since X_2 is independent of (X_1, S, Y_1). Now, observe that with the choice $X_2 \sim \text{Bernoulli}(\frac{1}{2})$ independent of (S, X_1), we have $H(X_2) = H(S) = 1$ and, so, the right-hand side (RHS) of (43) is larger than the RHS of (35). This shows the achievability of the rate $R_1 = I(X_1;Y_1|S)$.

2) The converse follows straightforwardly by specializing Theorem 2 (or the cut-set upper bound) to this example,

$$R \leq I(X_1;Y|X_2,S) \quad (44)$$
$$= I(X_1;Y_1|X_2,S) \quad (45)$$
$$= H(Y_1|X_2,S) - H(Y_1|X_1,X_2,S) \quad (46)$$
$$\leq H(Y_1|S) - H(Y_1|X_1,X_2,S) \quad (47)$$
$$\leq H(Y_1|S) - H(Y_1|X_1,S) \quad (48)$$
$$= I(X_1;Y_1|S), \quad (49)$$

where (47) holds since conditioning reduces entropy, and (48) holds by the Markov relation $X_2 \leftrightarrow (X_1, S) \leftrightarrow Y_1$.

Claim 2: The capacity of the state-dependent binary memoryless MAC shown in Fig. 2 satisfies

$$C_B = h(p * q_1) - h(p) > \max_{p(u,x_1|s)} I(U;Y_1) - I(U;S). \quad (50)$$

Proof: Claim 2 is a simple consequence of Claim 1 and known results on the capacity of the binary dirty paper channel (see, for example, [43] and references therein). More specifically, the capacity C_B in Claim 1 is that of a point-to-point state-dependent additive binary channel with a Bernoulli $(\frac{1}{2})$ state known at both transmitter and receiver ends, a Bernoulli (p) noise representing the binary symmetric channel and average input constraint q_1 at the transmitter. Thus, an explicit characterization of C_B is given by [43]

$$C_B = h(p * q_1) - h(p). \quad (51)$$

Let now R_{GP} be the maximum achievable rate had the strictly causal part S^{i-1} of the state been of no utility, or equivalently, had Encoder 2 been of no help. R_{GP} is the capacity of a binary dirty paper channel given by [43]

$$R_{\text{GP}} = \max_{p(u,x_1|s)} I(U;Y_1) - I(U;S) = \left\{ \begin{array}{ll} G(q_1), & \text{if} p^\star \leq q_1 \leq \frac{1}{2} \\ q_1 \log(\frac{1-p^\star}{p^\star}), & \text{if} 0 \leq q_1 \leq p^\star \end{array} \right\} \quad (52)$$

where $p^\star = 1 - 2^{-h(p)}$ and the function $G(q)$, defined for $q \in [0, 1/2]$, is given by

$$G(q) = \begin{cases} h(q) - h(p), & \text{if } p \leq q \leq \frac{1}{2} \\ 0, & \text{if } 0 \leq q \leq p \end{cases}. \quad (53)$$

Observing that $h(p * q_1) > h(q_1)$ for all $0 < q_1 < 1/2$, it is easy to see that $C_B > R_{\text{GP}}$.

Remark 4: In this example, the encoder that knows the states only strictly causally simply conveys these states to the receiver, noiselessly. The receiver then becomes aware of the channel states fully (since the delay in learning these states at the decoder has no impact on the capacity). This explains why Encoder 1 can transmit at rates that can be strictly larger than the standard Gel'fand–Pinker rate (52); and in fact achieves the capacity (50) of a state-dependent additive binary channel with the states known at both transmitter and receiver ends. □

C. Common-Message Capacity

In this section, we study the important case in which the two encoders transmit only the common message, i.e., $R_1 = 0$. The following corollary characterizes the capacity in this case, to which we refer as *common-message capacity*.

Corollary 2: The common message capacity, C, of the MAC with common message and states known noncausally at one encoder and strictly causally at the other encoder is given by

$$C = \max I(K, X_2; Y) - I(K, X_2; S) \quad (54)$$

where the maximization is over joint measures $P_{S,K,X_1,X_2,Y}$ of the form

$$P_{S,K,X_1,X_2,Y} = Q_S P_{X_2} P_{K,X_1|S,X_2}. \quad (55)$$

Proof: The proof of Corollary 2 appears in Appendix F.

Remark 5: The common-message capacity of our model in Corollary 2 coincides with the common-message of the model with the state sequence S^n known noncausally at Encoder 1 and not at all at Encoder 2 [5]. That is, C can also be obtained by relaxing the constraint on R_1 in the region $\mathcal{C}'$ defined by (10) and (11). This shows that the knowledge of the states at Encoder 2 only strictly causally does not increase the common-message capacity. We should, however, note that this result is not a direct consequence of that in a MAC a state that is known only strictly causally at *all* encoders does not increase the capacity; and, so, the converse proof is needed here. □

IV. Memoryless Gaussian Case

In this section, we consider a two-user state-dependent Gaussian MAC in which the channel states and the noise are additive and Gaussian.

A. Channel Model

As in Section II, we assume that Encoder 1 knows the channel states noncausally and Encoder 2 knows the channel states strictly causally. The two encoders send some common message W_c, and, in addition, Encoder 1 sends an individual message W_1. At time instant i, the channel output Y_i is related to channel inputs $X_{1,i}$ and $X_{2,i}$ from the two encoders, the channel state S_i and the noise Z_i by

$$Y_i = X_{1,i} + X_{2,i} + S_i + Z_i, \tag{56}$$

where S_i and Z_i are zero-mean Gaussian random variables with variance Q and N, respectively. The random variables S_i and Z_i at time instant $i \in \{1, \dots, n\}$ are mutually independent, and independent from (S_j, Z_j) for $j \neq i$. Also, at time i, the input $X_{2,i}$ is independent from the state S_i.

We consider the individual power constraints on the transmitted power

$$\sum_{i=1}^{n} X_{1,i}^2 \leq nP_1, \sum_{i=1}^{n} X_{2,i}^2 \leq nP_2. \tag{57}$$

The definition of a code for this channel is the same as given in Section II, with the additional power constraints (57).

B. Capacity Region

The following theorem characterizes the capacity region of the studied Gaussian model.

Theorem 4: The capacity region of the Gaussian model (56) is given by the set of all the rate pairs (R_c, R_1) satisfying

$$\begin{aligned} R_1 &\leq \frac{1}{2}\log\Big(1 + \frac{P_1(1-\rho_{12}^2-\rho_{2s}^2)}{N}\Big) \\ R_c + R_1 & \\ &\leq \frac{1}{2}\log\Big(1 + \frac{(\sqrt{P_2}+\rho_{12}\sqrt{P_1})^2}{P_1(1-\rho_{12}^2-\rho_{1s}^2)+(\sqrt{Q}+\rho_{1s}\sqrt{P_1})^2+N}\Big) \\ &\quad + \frac{1}{2}\log\Big(1 + \frac{P_1(1-\rho_{12}^2-\rho_{1s}^2)}{N}\Big), \end{aligned} \tag{58}$$

where the maximization is over $\rho_{12} \in [0,1]$, $\rho_{1s} \in [-1,0]$ such that

$$\rho_{12}^2 + \rho_{1s}^2 \leq 1. \tag{59}$$

Proof: An outline proof of Theorem 4 is given in Appendix G.

Remark 6: The capacity region of our model in Theorem 4 coincides with that of the model (56) but with the state sequence S^n known noncausally at Encoder 1 and not all at Encoder 2 [5, Th. 7]. Then, an implication of Theorem 4 is that it is *optimal* for our model to just ignore the states S^{i-1} that are known at Encoder 2 and use the coding scheme of [5]. That is, the availability of the states only strictly causally at the encoder that sends only the common message in our model does not increase the capacity region any further. While one could expect some utility of the collaborative transmission of a lossy version of the state to the decoder as in the memoryless discrete setup (and also in the Gaussian setups of [15], [16], and [18]), a direct consequence of our converse proof is that this would be of no help, in the sense that it would not result in better transmission rates. This can be interpreted as follows. As can be seen from the proof of Theorem 1, the joint transmission of the state to the decoder aims at equipping it with an estimate of this state. This state estimate is then utilized as decoder side information for the decoding of the information messages. In the discrete memoryless case, this can be beneficial, in general, for the transmission of the private message, not the common message, as we already mentioned. In the Gaussian case, however, for the transmission of the private message, Encoder 1 knows the state noncausally, and therefore, it can cancel its effect completely using a variation of the standard dirty paper scheme [44], with no need to diminishing its effect via the joint transmission of the compressed version of the state. □

The following corollary follows straightforwardly from Theorem 4.

Corollary 3: The common message capacity, C_{G}, of the Gaussian model (56) is given by

$$\begin{aligned} C_{\mathrm{G}} = \max & \frac{1}{2}\log\Big(1+\frac{(\sqrt{P_2}+\rho_{12}\sqrt{P_1})^2}{P_1(1-\rho_{12}^2-\rho_{1s}^2)+(\sqrt{Q}+\rho_{1s}\sqrt{P_1})^2+N}\Big) \\ &+ \frac{1}{2}\log\Big(1 + \frac{P_1(1-\rho_{12}^2-\rho_{1s}^2)}{N}\Big), \end{aligned} \tag{60}$$

where the maximization is over $\rho_{12} \in [0,1]$, $\rho_{1s} \in [-1,0]$ such that

$$\rho_{12}^2 + \rho_{1s}^2 \leq 1. \tag{61}$$

V. Conclusion

In this paper, we consider a state-dependent MAC with the channel state available noncausally at one of the encoders and only strictly causally at the other encoder. The decoder is not aware of the channel state. Both encoders transmit a common message and, in addition, Encoder 1, the encoder that knows the state noncausally, transmits an individual message. We study the capacity region of this communication model. The analysis also helps understanding the utility of revealing the state only strictly causally to the encoder that sends only the common message as well as optimal compressions to perform it.

In the discrete memoryless case, we characterize the capacity region of this model with a single-letter expression. In particular, the analysis reveals optimal ways of exploiting the knowledge of the state only strictly causally at the encoder that sends only the common message. The encoders collaborate to convey to the decoder a lossy version of the state, in addition to transmitting the information messages through a generalized Gel'fand–Pinsker binning. Particularly important in this problem are the questions of 1) optimal ways of performing the state compression, and 2) whether or not the compression indices should be decoded uniquely. We develop two optimal coding schemes that perform the state compression differently. The first coding scheme is à-la noisy network coding by Lim *et al.* or the quantize-map-and-forward by Avestimeher *et al.*, i.e., with no binning and nonunique decoding of the compression indices. The second coding scheme employs Wyner–Ziv binning with backward decoding and nonunique decoding of the compression indices. We note that backward decoding and nonunique decoding seem to be key elements for the optimality of the Wyner–Ziv-based coding scheme. Next, by exploiting our outer bound and the involved auxiliary variables specifically, we show that, although not required in general, for our specific model, the compression indices can in fact be decoded uniquely essentially without altering the capacity region but at the expense of larger alphabets sizes for the auxiliary random variables.

The capacity region contains that of the model of [5], and this shows that revealing the state even only strictly causally to the encoder that sends only the common message is beneficial and enlarges the capacity region in general. Furthermore, by investigating a discrete memoryless example, we show that this inclusion can be strict, thus demonstrating the utility of conveying a compressed version of the state to the decoder cooperatively by the encoders.

We also specialize our results to the case in which the two encoders send only the common message. We characterize the common-message capacity and show that knowing the states only strictly causally at one of the encoders is not beneficial in this case.

Furthermore, we also study the memoryless Gaussian setting in which the channel state and the noise are additive and Gaussian. In this case, we establish an operative outer bound on the achievable rate pairs and then show that this outer bound is achievable, thus yielding a closed-form expression of the capacity region. Unlike the discrete memoryless case, we show that the knowledge of the states only strictly causally at the encoder that sends only the common message does not increase the capacity region in this case.

Appendix A
Proof of Proposition 1

Part 1: To prove the convexity of the region, we use a standard argument. We introduce a time-sharing random variable T and define the joint distribution

$$P_{T,S,U,V,X_1,X_2,Y}(t,s,u,v,x_1,x_2,y) = P_{T,S,U,V,X_1,X_2}(t,s,u,v,x_1,x_2)W_{Y|X_1,X_2,S}(y|x_1,x_2,s) \tag{A-1}$$

$$\sum_{u,v,x_1,x_2} P_{T,S,U,V,X_1,X_2}(t,s,u,v,x_1,x_2) = P_T(t)Q_S(s). \tag{A-2}$$

Let now (R_c^T, R_1^T) be the common and individual rates resulting from time sharing. Then,

$$\begin{aligned}
R_1^T &\le I(U;Y|V,X_2,T) - I(U;S|V,X_2,T) && \text{(A-3)}\\
&= I(U;Y|\tilde{V},X_2) - I(U;S|\tilde{V},X_2) && \text{(A-4)}\\
R_c^T + R_1^T &\le I(U,V,X_2;Y|T) - I(U,V,X_2;S|T) && \text{(A-5)}\\
&= I(U,V,X_2;Y|T) - I(U,V,X_2,T;S) && \text{(A-6)}\\
&\le I(U,V,X_2,T;Y) - I(U,V,X_2,T;S) && \text{(A-7)}\\
&= I(U,\tilde{V},X_2;Y) - I(U,\tilde{V},X_2;S), && \text{(A-8)}
\end{aligned}$$

where $\tilde{V} := (V,T)$. That is, the time sharing random variable T is incorporated into the auxiliary random variable V. This shows that time sharing cannot yield rate pairs that are not included in $\mathcal{C}$ and, hence, $\mathcal{C}$ is convex.

Part 2: To prove that the region $\mathcal{C}$ is not altered if one restricts the random variables U and V to have their alphabets restricted as indicated in (9), we invoke the support lemma [46, p. 310]. Fix a distribution $\mu \in \mathcal{P}$ of (S,U,V,X_1,X_2,Y) and, without loss of generality, let us denote the product set $\mathcal{S}\times\mathcal{X}_1\times\mathcal{X}_2 = \{1,\ldots,m\}$, $m = |\mathcal{S}\times\mathcal{X}_1\times\mathcal{X}_2|$.

To prove the bound (9a) on $|\mathcal{V}|$, note that we have

$$\begin{aligned}
&I_\mu(U;Y|V,X_2) - I_\mu(U;S|V,X_2)\\
&\quad= I_\mu(U,X_2;Y|V) - I_\mu(X_2;Y|V)\\
&\qquad - I_\mu(U,X_2;S|V) + I_\mu(X_2;S|V)\\
&\quad= H_\mu(U,X_2,S|V) - H_\mu(U,X_2,Y|V)\\
&\qquad - H_\mu(X_2,S|V) + H_\mu(X_2,Y|V) && \text{(A-9)}
\end{aligned}$$

and

$$\begin{aligned}
&I_\mu(U,V,X_2;Y) - I_\mu(U,V,X_2;S)\\
&\quad= I_\mu(U,X_2;Y|V) - I_\mu(U,X_2;S|V) + I_\mu(V;Y) - I_\mu(V;S)\\
&\quad= H_\mu(U,X_2,S|V) - H_\mu(U,X_2,Y|V) + H_\mu(Y) - H_\mu(S).
\end{aligned} \tag{A-10}$$

Hence, it suffices to show that the following functionals of $\mu(S,U,V,X_1,X_2,Y)$:

$$r_i(\mu) = \mu(s,x,x'), \quad i = 1,\ldots,m-1 \tag{A-11a}$$

$$r_m(\mu) = \int_v d_\mu(v)[H_\mu(U,X_2,S|v) - H_\mu(U,X_2,Y|v) - H_\mu(X_2,S|v) + H_\mu(X_2,Y|v)] \tag{A-11b}$$

$$r_{m+1}(\mu) = \int_v d_\mu(v)[H_\mu(U,X_2,S|v) - H_\mu(U,X_2,Y|v)] \tag{A-11c}$$

can be preserved with another measure $\mu' \in \mathcal{P}$. Observing that there is a total of $\Big(|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2|+1\Big)$ functionals in (A-11), this is ensured by a standard application of the support lemma, and this shows that the alphabet of the auxiliary random variable V can be restricted as indicated in (9a) without altering the region $\mathcal{C}$.

Once the alphabet of V is fixed, we apply similar arguments to bound the alphabet of U, where this time $(|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2|+1)|\mathcal{S}||\mathcal{X}_1||\mathcal{X}_2|-1$ functionals must be satisfied in order to preserve the joint distribution of (S,V,X_1,X_2), and one more functional to preserve

$$\begin{aligned}
&I_\mu(U;Y|V,X_2) - I_\mu(U;S|V,X_2)\\
&\quad= H_\mu(Y,V,X_2) - H_\mu(S,V,X_2) + H_\mu(S,V,X_2|U)\\
&\qquad - H_\mu(Y,V,X_2|U) && \text{(A-12a)}\\
&I_\mu(U,V,X_2;Y) - I_\mu(U,V,X_2;S)\\
&\quad= H_\mu(Y) - H_\mu(S) + H_\mu(S,V,X_2|U) - H_\mu(Y,V,X_2|U). && \text{(A-12b)}
\end{aligned}$$

This shows that the alphabet of the auxiliary random variable U can be restricted as indicated in (9b) without altering the region $\mathcal{C}$, and completes the proof of Proposition 1.

APPENDIX B
PROOF OF THEORM 1

Throughout this section, we denote the set of strongly jointly ϵ-typical sequences [45, Ch. 14.2] with respect to the distribution $P_{X,Y}$ as $\mathcal{T}_\epsilon^n(P_{X,Y})$.

A. Direct Part of Theorem 1

To bound the probability of error, we assume without loss of generality that the compression indices are all equal to unity, i.e., $t_1 = t_2 = \cdots = t_B = 1$.

We examine the probability of error associated with each of the encoding and decoding procedures. The events E_1, E_2, and E_3 correspond to encoding errors, and the events E_4, E_5, E_6, and E_7 correspond to decoding errors.

- Let $E_1 = \cup_{i=1}^B E_{1i}$ where E_{1i} is the event that, for the encoding in block i, there is no covering codeword $\mathbf{v}_{i-1}(w_c, t_{i-2}, t_{i-1})$ strongly jointly typical with $\mathbf{s}[i-1]$ given $\mathbf{x}_{2,i-1}(w_c, t_{i-2})$, i.e.,

$$E_1 = \bigcup_{i=1}^B \Big\{ \nexists\ t_{i-1} \in [1 : \hat{M}] \text{ s.t.} : \Big(\mathbf{v}_{i-1}(w_c, t_{i-2}, t_{i-1}), \mathbf{s}[i-1], \mathbf{x}_{2,i-1}(w_c, t_{i-2}) \Big) \in \mathcal{T}_\epsilon^n(P_{V,S,X_2}) \Big\}. \quad \text{(B-1)}$$

For $i \in [1 : B]$, the probability that $(\mathbf{s}[i-1], \mathbf{x}_{2,i-1}(w_c, t_{i-2}))$ is not jointly typical goes to zero as $n \to \infty$, by the asymptotic equipartition property [45, p. 384]. Then, for $(\mathbf{s}[i-1], \mathbf{x}_{2,i-1}(w_c, t_{i-2}))$ jointly typical, the covering lemma [47, Lecture Note 3] ensures that the probability that there is no $t_{i-1} \in [1 : \hat{M}]$ such that $(\mathbf{v}_{i-1}(w_c, t_{i-2}, t_{i-1}), \mathbf{s}[i-1])$ is strongly jointly typical given $\mathbf{x}_{2,i-1}(w_c, t_{i-2})$ is exponentially small for large n provided that the number of covering codewords $\mathbf{v}_{i-1}$ is greater than $2^{nI(V;S|X_2)}$, i.e.,

$$\hat{R} > I(V; S|X_2). \quad \text{(B-2)}$$

Thus, if (B-2) holds, $\Pr(E_{1i}) \to 0 \quad \text{as} \quad n \to \infty$ and, so, by the union of bound over the B blocks, $\Pr(E_1) \to 0 \quad \text{as} \quad n \to \infty$.

- Let $E_2 = \cup_{i=1}^B E_{2i}$ where E_{2i} is the event that, for the encoding in block i, Encoder 1 can find no covering codeword $\mathbf{v}_i(w_c, t_{i-1}, t_i)$ strongly jointly typical with $\mathbf{s}[i]$ given $\mathbf{x}_{2,i}(w_c, t_{i-1})$. Similarly to the event E_1, it is easy to see that $\Pr(E_2|E_1^c) \to 0 \quad \text{as} \quad n \to \infty$ if (B-2) is true.
- Let $E_3 = \cup_{i=1}^B E_{3i}$ where E_{3i} is the event that, for the encoding in block i, there is no sequence $\mathbf{u}_i(w_c, t_{i-1}, t_i, w_1, j_i)$ jointly typical with $\mathbf{s}[i]$ given $\mathbf{x}_{2,i}(w_c, t_{i-1})$ and $\mathbf{v}_i(w_c, t_{i-1}, t_i)$, i.e.,

$$E_3 = \bigcup_{i=1}^B \Big\{ \nexists\ j_i \in [1 : J] \text{ s.t} : \Big(\mathbf{u}_i(w_c, t_{i-1}, t_i, w_1, j_i), \mathbf{s}[i], \mathbf{v}_i(w_c, t_{i-1}, t_i), \mathbf{x}_{2,i}(w_c, t_{i-1}) \Big) \in \mathcal{T}_\epsilon^n(P_{U,S,V,X_2}) \Big\}. \text{(B-3)}$$

To bound the probability of the event E_{3i}, we use a standard argument [3]. More specifically, conditioned on $E_{1,i}^c$ and $E_{2,i}^c$, the complement events of E_{1i} and E_{2i}, respectively, we have that the state $\mathbf{s}[i]$ is jointly typical with $(\mathbf{x}_{2,i}(w_c, t_{i-1}), \mathbf{v}_i(w_c, t_{i-1}, t_i))$. Then, for $\mathbf{u}_i(w_c, t_{i-1}, t_i, w_1, j_i)$ generated independently of $\mathbf{s}[i]$ given $\mathbf{x}_{2,i}(w_c, t_{i-1})$ and $\mathbf{v}_i(w_c, t_{i-1}, t_i)$, with i.i.d. components drawn according to $P_{U|V,X_2}$, the probability that $\mathbf{u}_i(w_c, t_{i-1}, t_i, w_1, j_i)$ is jointly typical with $\mathbf{s}[i]$ given $\mathbf{x}_{2,i}(w_c, t_{i-1})$ and $\mathbf{v}_i(w_c, t_{i-1}, t_i)$ is greater than $(1-\epsilon)2^{-n(I(U;S|V,X_2)+\epsilon)}$ for sufficiently large n. There is a total of J such $\mathbf{u}_i$'s in each bin. Conditioned on E_{1i}^c and E_{2i}^c, the probability of the event E_{3i}, the probability that there is no such $\mathbf{u}_i$, is therefore bounded as

$$\Pr(E_{3i}|E_{1i}^c, E_{2i}^c) \leq [1 - (1-\epsilon)2^{-n(I(U;S|V,X_2)+\epsilon)}]^J. \quad \text{(B-4)}$$

Taking the logarithm on both sides of (B-4) and substituting J, we obtain that $\ln(\Pr(E_{3i}|E_{1i}^c, E_{2i}^c)) \leq -(1-\epsilon)2^{n(\delta-1)\epsilon}$. Thus, $\Pr(E_{3i}|E_{1i}^c, E_{2i}^c) \to 0 \quad \text{as} \quad n \to \infty$ and, so, by the union bound, $\Pr(E_3|E_1^c, E_2^c) \to 0 \quad \text{as} \quad n \to \infty$.

- For the decoding of the common message w_c at the receiver, let $E_4 = \cup_{i=1}^B E_{4i}$ where E_{4i} is the event that $\Big(\mathbf{x}_{2,i}(w_c, t_{i-1}),\ \mathbf{u}_i(w_c, t_{i-1}, t_i, w_1, j_i^\star),\ \mathbf{v}_i(w_c, t_{i-1}, t_i),\ \mathbf{y}[i] \Big)$ is not jointly typical, i.e.,

$$E_4 = \bigcup_{i=1}^B \Big\{ \Big(\mathbf{x}_{2,i}(w_c, t_{i-1}), \mathbf{u}_i(w_c, t_{i-1}, t_i, w_1, j_i^\star), \mathbf{v}_i(w_c, t_{i-1}, t_i), \mathbf{y}[i] \Big) \notin \mathcal{T}_\epsilon^n(P_{X_2,U,V,Y}) \Big\}. \quad \text{(B-5)}$$

Conditioned on E_{1i}^c, E_{2i}^c, and E_{3i}^c, the vectors $\mathbf{s}[i]$, $\mathbf{x}_{2,i}(w_c, t_{i-1}), \mathbf{v}_i(w_c, t_{i-1}, t_i)$ and $\mathbf{u}_i(w_c, t_{i-1}, t_i, w_1, j_i^\star)$ are jointly typical and with $\mathbf{x}_1[i]$. Then, conditioned on E_{1i}^c, E_{2i}^c, and E_{3i}^c, the vectors $\mathbf{s}[i]$, $\mathbf{x}_{2,i}(w_c, t_{i-1})$, $\mathbf{v}_i(w_c, t_{i-1}, t_i)$, $\mathbf{u}_i(w_c, t_{i-1}, t_i, w_1, j_i^\star)$ and $\mathbf{y}[i]$ are jointly typical by the Markov lemma [45, p. 436], i.e., $\Pr(E_{4i}|E_{1i}^c, E_{2i}^c, E_{3i}^c) \to 0 \quad \text{as} \quad n \to \infty$. Thus, by the union bound over the B blocks, $\Pr(E_4|E_1^c, E_2^c, E_3^c) \to 0 \quad \text{as} \quad n \to \infty$.

- For the decoding of the common message w_c at the receiver, let E_5 be the event that $\mathbf{x}_{2,i}(w_c', t_{i-1})$, $\mathbf{u}_i(w_c', t_{i-1}, t_i, w_1, j_i)$, $\mathbf{v}_i(w_c', t_{i-1}, t_i)$, and $\mathbf{y}[i]$ are jointly typical for all $i = 1, \ldots, B$ and some $w_c' \in [1 : M_c]$, $w_1 \in [1 : M_1]$, $t^B = (t_1, \ldots, t_B) \in [1 : \hat{M}]^B$ and $j^B = (j_1, \ldots, j_B) \in [1 : J]^B$ such that $w_c' \neq w_c$, i.e.,

$$E_5 = \Big\{ \exists\ w_c' \in [1 : M_c], w_1 \in [1 : M_1], t^B = (t_1, \ldots, t_B) \in [1 : \hat{M}]^B, j^B \in [1 : J]^B \text{s.t.} : w_c' \neq w_c, \bigcap_{i=1}^B \Big\{ \Big(\mathbf{x}_{2,i}(w_c', t_{i-1}), \mathbf{u}_i(w_c', t_{i-1}, t_i, w_1, j_i), \mathbf{v}_i(w_c', t_{i-1}, t_i), \mathbf{y}[i] \Big) \in \mathcal{T}_\epsilon^n(P_{X_2,U,V,Y}) \Big\} \Big\}. \quad \text{(B-6)}$$

To bound the probability of the event E_5, define the following event for given $w_c' \in [1 : M_c]$, $w_1 \in [1 : M_1]$, $(t_{i-1}, t_i) \in [1 : \hat{M}]^2$ and $j_i \in [1 : J]$ such that $w_c' \neq w_c$:

$$E_{5i}(w_c', t_{i-1}, t_i, w_1, j_i) = \Big\{ \Big(\mathbf{x}_{2,i}(w_c', t_{i-1}), \mathbf{u}_i(w_c', t_{i-1}, t_i, w_1, j_i), \mathbf{v}_i(w_c', t_{i-1}, t_i), \mathbf{y}[i] \Big) \in \mathcal{T}_\epsilon^n(P_{X_2,U,V,Y}) \Big\}.$$

Note that for $w_c' \neq w_c$ the vectors $\mathbf{x}_{2,i}(w_c', t_{i-1})$, $\mathbf{v}_i(w_c', t_{i-1}, t_i)$ and $\mathbf{u}_i(w_c', t_{i-1}, t_i, w_1, j_i)$ are generated independently of $\mathbf{y}[i]$. Hence, by the joint typicality lemma [47, Lecture Note 2], we get

$$\Pr\Big(E_{5i}(w_c', t_{i-1}, t_i, w_1, j_i)|E_1^c, E_2^c, E_3^c, E_4^c\Big) \leq 2^{-n[I(U,V,X_2;Y)-\epsilon]}. \quad \text{(B-7)}$$

Then, conditioned on the events E_1^c, E_2^c, E_3^c, and E_4^c, the probability of the event E_5 can be bounded as given by (B-8) at the bottom of the page.

The RHS of (B-8) tends to zero as $n \to \infty$ if

$$R_c + R_1 \leq \frac{B-1}{B}\big(I(U,V,X_2;Y) - I(U;S|V,X_2) - \hat{R}\big) - \frac{\hat{R}}{B} - \frac{I(U;S|V,X_2)}{B}. \quad \text{(B-9)}$$

Finally, using (B-2) to eliminate $\hat{R}$ from (B-9) and taking $B \to \infty$, we get $\Pr(E_5|E_1^c, E_2^c, E_3^c, E_4^c) \to 0$ as long as

$$\begin{aligned} R_c + R_1 &\leq I(U,V,X_2;Y) - I(U,V;S|X_2) \\ &= I(U,V,X_2;Y) - I(U,V,X_2;S), \quad \text{(B-10)} \end{aligned}$$

where the last equality follows since X_2 and S are independent.

- For the decoding of the individual message w_1 at the receiver, let $E_6 = \cup_{i=1}^B E_{6i}$, where E_{6i} is the event that $\mathbf{x}_{2,i}(w_c, t_{i-1})$, $\mathbf{v}_i(w_c, t_{i-1}, t_i)$, $\mathbf{u}_i(w_c, t_{i-1}, t_i, w_1, j_i^\star)$, and $\mathbf{y}[i]$ are not jointly typical, i.e.,

$$E_{6i} = \Big\{\Big(\mathbf{x}_{2,i}(w_c, t_{i-1}), \mathbf{v}_i(w_c, t_{i-1}, t_i), \mathbf{u}_i(w_c, t_{i-1}, t_i, w_1, j_i^\star), \mathbf{y}[i]\Big) \notin \mathcal{T}_\epsilon^n(P_{X_2,V,U,Y})\Big\}. \quad \text{(B-11)}$$

From our analysis of the probability of the error event E_4, it is easy to see that, conditioned on E_1^c, E_2^c and E_3^c, the event E_{6i} has exponentially small probability. Thus, by the union bound over the B blocks, $\Pr(E_6|E_1^c, E_2^c, E_3^c) \longrightarrow 0$ as $n \longrightarrow \infty$, where $E_6 = \cup_{i=1}^B E_{6i}$.

- For the decoding of the individual message w_1 at the receiver, let E_7 be the event that $\mathbf{x}_{2,i}(w_c, t_{i-1})$, $\mathbf{u}_i(w_c, t_{i-1}, t_i, w_1', j_i)$, $\mathbf{v}_i(w_c, t_{i-1}, t_i)$, and $\mathbf{y}[i]$ are jointly typical for all $i = 1, \ldots, B$ and some $w_1' \in [1:M_1]$, $t^B = (t_1, \ldots, t_B) \in [1:\hat{M}]^B$ and $j^B = (j_1, \ldots, j_B) \in [1:J]^B$ such that $w_1' \neq w_1$, i.e., as given by Eq.(B-12) at the bottom of the next page.

To bound the probability of the event E_7, define the following event for given $w_1' \in [1:M_1]$, $(t_{i-1}, t_i) \in [1:\hat{M}]^2$ and $j_i \in [1:J]$,

$$E_{7i}(w_1', t_{i-1}, t_i, j_i) = \Big\{\Big(\mathbf{x}_{2,i}(w_c, t_{i-1}), \mathbf{u}_i(w_c, t_{i-1}, t_i, w_1', j_i), \mathbf{v}_i(w_c, t_{i-1}, t_i), \mathbf{y}[i]\Big) \in \mathcal{T}_\epsilon^n(P_{X_2,U,V,Y})\Big\}.$$

$$\begin{aligned}
\Pr(E_5|E_1^c, E_2^c, E_3^c, E_4^c) &= \Pr\Big(\bigcup_{w_c' \neq w_c} \bigcup_{w_1 \in [1:M_1]} \bigcup_{t^B \in [1:\hat{M}]^B} \bigcup_{j^B \in [1:J]^B} \bigcap_{i=1}^B E_{5i}(w_c', t_{i-1}, t_i, w_1, j_i)|E_1^c, E_2^c, E_3^c, E_4^c\Big) \\
&\overset{(a)}{\leq} \sum_{w_c' \neq w_c} \sum_{w_1 \in [1:M_1]} \sum_{t^B \in [1:\hat{M}]^B} \sum_{j^B \in [1:J]^B} \Pr\Big(\bigcap_{i=1}^B E_{5i}(w_c', t_{i-1}, t_i, w_1, j_i)|E_1^c, E_2^c, E_3^c, E_4^c\Big) \\
&\overset{(b)}{=} \sum_{w_c' \neq w_c} \sum_{w_1 \in [1:M_1]} \sum_{t^B \in [1:\hat{M}]^B} \sum_{j^B \in [1:J]^B} \prod_{i=1}^B \Pr\Big(E_{5i}(w_c', t_{i-1}, t_i, w_1, j_i)|E_1^c, E_2^c, E_3^c, E_4^c\Big) \\
&\leq \sum_{w_c' \neq w_c} \sum_{w_1 \in [1:M_1]} \sum_{t^B \in [1:\hat{M}]^B} \sum_{j^B \in [1:J]^B} \prod_{i=2}^B \Pr\Big(E_{5i}(w_c', t_{i-1}, t_i, w_1, j_i)|E_1^c, E_2^c, E_3^c, E_4^c\Big) \\
&\overset{(c)}{\leq} \sum_{w_c' \neq w_c} \sum_{w_1 \in [1:M_1]} \sum_{t^B \in [1:\hat{M}]^B} \sum_{j^B \in [1:J]^B} \prod_{i=2}^B 2^{-n\big[I(U,V,X_2;Y)-\epsilon\big]} \\
&= \sum_{w_c' \neq w_c} \sum_{w_1 \in [1:M_1]} \sum_{t_B \in [1:\hat{M}]} \sum_{j_B \in [1:J]} 2^{n(B-1)\big[\hat{R}+\hat{\eta}\epsilon\big]} 2^{-n(B-1)\big[I(U,V,X_2;Y)-I(U;S|V,X_2)-(\delta+1)\epsilon\big]} \\
&\leq M_c M_1 \hat{M} J 2^{-n(B-1)\big[I(U,V,X_2;Y)-I(U;S|V,X_2)-\hat{R}-(\hat{\eta}+\delta+1)\epsilon\big]} \\
&= 2^{-nB\Big[\frac{B-1}{B}\big(I(U,V,X_2;Y)-I(U;S|V,X_2)-\hat{R}\big)-(R_c+R_1)-\frac{\hat{R}}{B}-\frac{I(U;S|V,X_2)}{B}+\big(\eta_c+\eta_1-\hat{\eta}-\delta-\frac{B-1}{B}\big)\epsilon\Big]} \quad \text{(B-8)}
\end{aligned}$$

where (a) follows by the union bound; (b) follows since the codebook is generated independently for each block $i \in [1:B]$ and the channel is memoryless; and (c) follows by (B-7).

Then, the probability of the event E_7 given by (B-12) can be bounded as given by (B-13) at the bottom of this page. For $w_1' \neq w_1$, the probability of the event $E_{7i}(w_1', t_{i-1}, t_i, j_i)$ conditioned on $E_1^c, E_2^c, E_3^c, E_4^c, E_5^c, E_6^c$ can be bounded as follows, depending on the values of t_{i-1} and t_i:

i) if $t_{i-1} \neq 1$ then $\Big(\mathbf{u}_i(w_c, t_{i-1}, t_i, w_1', j_i), \mathbf{x}_{2,i}(w_c, t_{i-1}), \mathbf{v}_i(w_c, t_{i-1}, t_i)\Big)$ is generated independently of the output vector $\mathbf{y}[i]$ irrespective to the value of t_i, and so, by the joint typicality lemma [47, Lecture Note 2]

$$\Pr\Big(E_{7i}(w_1', t_{i-1}, t_i, j_i) | \cap_{k=1}^{6} E_k^c\Big) \leq 2^{-n[I(U,V,X_2;Y)-\epsilon]}. \tag{B-14}$$

ii) if $t_{i-1} = 1$ and $t_i \neq 1$, then $\Big(\mathbf{u}_i(w_c, t_{i-1}, t_i, w_1', j_i), \mathbf{v}_i(w_c, t_{i-1}, t_i)\Big)$ is generated independently of the output vector $\mathbf{y}[i]$ conditionally on $\mathbf{x}_{2,i}(w_c, t_{i-1})$, and hence

$$\Pr\Big(E_{7i}(w_1', t_{i-1}, t_i, j_i) | \cap_{k=1}^{6} E_k^c\Big) \leq 2^{-n[I(U,V;Y|X_2)-\epsilon]}. \tag{B-15}$$

iii) if $t_{i-1} = 1$ and $t_i = 1$, then $\mathbf{u}_i(w_c, t_{i-1}, t_i, w_1', j_i)$ is generated independently of the output vector $\mathbf{y}[i]$ conditionally on $\mathbf{x}_{2,i}(w_c, t_{i-1})$ and $\mathbf{v}_i(w_c, t_{i-1}, t_i)$, and hence

$$\Pr\Big(E_{7i}(w_1', t_{i-1}, t_i, j_i) | \cap_{k=1}^{6} E_k^c\Big) \leq 2^{-n[I(U;Y|V,X_2)-\epsilon]}. \tag{B-16}$$

Now, note that since $I(U, V; Y|X_2) \geq I(U; Y|V, X_2)$, if $w_1' \neq w_1$ and $t_{i-1} = 1$ the following holds irrespective to the value of t_i:

$$\Pr\Big(E_{7i}(w_1', t_{i-1}, t_i, j_i) | \cap_{k=1}^{6} E_k^c\Big) \leq 2^{-n[I(U;Y|V,X_2)-\epsilon]}. \tag{B-17}$$

Let $I_1 := I(U; Y|V, X_2)$ and $I_2 := I(U, V, X_2; Y)$. If the sequence $(t_1, \ldots, t_{B-1})$ has k ones, we have

$$\prod_{i=2}^{B} \Pr\Big(E_{7i}(w_1', t_{i-1}, t_i, j_i) | \cap_{k=1}^{6} E_k^c\Big) \leq 2^{-n[kI_1+(B-1-k)I_2-(B-1)\epsilon]}. \tag{B-18}$$

Continuing from (B-13), we then bound the probability of the event E_7 as given by (B-19) at the bottom of the next page.

The RHS of (B-19) tends to zero as $n \to \infty$ if

$$R_1 \leq \frac{B-1}{B}\Big(\min\big(I_1 - I(U; S|V, X_2), I_2 - \hat{R} - I(U; S|V, X_2)\big)\Big) - \frac{\hat{R}}{B} - \frac{I(U; S|V, X_2)}{B}. \tag{B-20}$$

Finally, using (B-2) to eliminate $\hat{R}$ from (B-20) and taking $B \to \infty$, we get $\Pr(E_7 | \cap_{k=1}^{6} E_k^c) \to 0$ as long as

$$R_1 \leq I_1 - I(U; S|V, X_2) = I(U; Y|V, X_2) - I(U; S|V, X_2) \tag{B-21}$$

and

$$R_1 \leq I_2 - I(V; S|X_2) - I(U; S|V, X_2) = I(U, V, X_2; Y) - I(U, V, X_2; S). \tag{B-22}$$

$$E_7 = \Big\{\exists\, w_1' \in [1 : M_1], t^B = (t_1, \ldots, t_B) \in [1 : \hat{M}]^B, j^B \in [1 : J]^B \text{s.t. } : w_1' \neq w_1, \bigcap_{i=1}^{B} \Big\{\Big(\mathbf{x}_{2,i}(w_c, t_{i-1}), \mathbf{u}_i(w_c, t_{i-1}, t_i, w_1', j_i), \mathbf{v}_i(w_c, t_{i-1}, t_i), \mathbf{y}[i]\Big) \in \mathcal{T}_\epsilon^n(P_{X_2,U,V,Y})\Big\}\Big\} \tag{B-12}$$

$$\begin{aligned}
\Pr(E_7 | E_1^c, E_2^c, E_3^c, E_4^c, E_5^c, E_6^c) &= \Pr\Big(\bigcup_{w_1' \neq w_1} \bigcup_{t^B \in [1:\hat{M}]^B} \bigcup_{j^B \in [1:J]^B} \bigcap_{i=1}^{B} E_{7i}(w_1', t_{i-1}, t_i, j_i) | \cap_{k=1}^{6} E_k^c\Big) \\
&\overset{(d)}{\leq} \sum_{w_1' \neq w_1} \sum_{t^B \in [1:\hat{M}]^B} \sum_{j^B \in [1:J]^B} \Pr\Big(\bigcap_{i=1}^{B} E_{7i}(w_1', t_{i-1}, t_i, j_i) | \cap_{k=1}^{6} E_k^c\Big) \\
&\overset{(e)}{=} \sum_{w_1' \neq w_1} \sum_{t^B \in [1:\hat{M}]^B} \sum_{j^B \in [1:J]^B} \prod_{i=1}^{B} \Pr\Big(E_{7i}(w_1', t_{i-1}, t_i, j_i) | \cap_{k=1}^{6} E_k^c\Big) \\
&\leq \sum_{w_1' \neq w_1} \sum_{t^B \in [1:\hat{M}]^B} \sum_{j^B \in [1:J]^B} \prod_{i=2}^{B} \Pr\Big(E_{7i}(w_1', t_{i-1}, t_i, j_i) | \cap_{k=1}^{6} E_k^c\Big)
\end{aligned} \tag{B-13}$$

where (d) follows by the union bound and (e) follows since the codebook is generated independently for each block $i \in [1 : B]$ and the channel is memoryless.

Finally, noting that the condition (B-22) is redundant as $R_c \geq 0$ in (B-10), we obtain that the probability of error tends to zero as $n \to \infty$ and $B \to \infty$ if

$$R_1 \leq I(U;Y|V,X_2) - I(U;S|V,X_2) \quad \text{(B-23a)}$$
$$R_c + R_1 \leq I(U,V,X_2;Y) - I(U,V,X_2;S). \quad \text{(B-23b)}$$

This completes the proof of achievability.

B. Converse Part of Theorem 1

We prove that for any (M_c, M_1, n, ϵ) code consisting of a mapping $\phi_1 : \mathcal{W}_c \times \mathcal{W}_1 \times \mathcal{S}^n \longrightarrow \mathcal{X}_1^n$ at Encoder 1, a sequence of mappings $\phi_{2,i} : \mathcal{W}_c \times \mathcal{S}^{i-1} \longrightarrow \mathcal{X}_2$, $i = 1, \ldots, n$, at Encoder 2, and a mapping $\psi : \mathcal{Y}^n \longrightarrow \mathcal{W}_c \times \mathcal{W}_1$ at the decoder with average error probability $P_e^n \to 0$ as $n \to 0$ and rates $R_c = n^{-1} \log_2 M_c$ and $R_1 = n^{-1} \log_2 M_1$, there exist random variables $(V, U, X_1, X_2) \in \mathcal{V} \times \mathcal{U} \times \mathcal{X}_1 \times \mathcal{X}_2$ with U and V satisfying (9) such that the joint distribution P_{S,V,U,X_1,X_2} is of the form

$$P_{S,V,U,X_1,X_2} = Q_S P_{X_2} P_{V|S,X_2} P_{U,X_1|V,S,X_2}, \quad \text{(B-24)}$$

the marginal distribution of S is $Q_S(s)$, i.e.,

$$\sum_{v,u,x_1,x_2} P_{S,V,U,X_1,X_2}(s,v,u,x_1,x_2) = Q_S(s) \quad \text{(B-25)}$$

and the rate pair (R_c, R_1) satisfies (8).

Define the random variables

$$\begin{aligned} \bar{V}_i &= (W_c, S^{i-1}, Y_{i+1}^n) \\ \bar{U}_i &= (W_1, \bar{V}_i). \end{aligned} \quad \text{(B-26)}$$

Observe that the random variables so defined satisfy

$$(S_i, \bar{U}_i, \bar{V}_i, X_{1,i}, X_{2,i}, Y_i) \in \mathcal{P}, \quad \forall i \in \{1, \ldots, n\}. \quad \text{(B-27)}$$

We first prove the following auxiliary result.

Lemma 1: The following inequalities hold:

$$I(W_1; Y^n|W_c) - I(W_1; S^n|W_c) \leq \sum_{i=1}^{n} I(\bar{U}_i; Y_i|\bar{V}_i, X_{2,i}) - I(\bar{U}_i; S_i|\bar{V}_i, X_{2,i}) \quad \text{(B-28)}$$

$$I(W_c, W_1; Y^n) - I(W_c, W_1; S^n) \leq \sum_{i=1}^{n} I(\bar{U}_i, \bar{V}_i, X_{2i}; Y_i) - I(\bar{U}_i, \bar{V}_i; S_i|X_{2i}). \quad \text{(B-29)}$$

$$\begin{aligned}
\Pr(E_7|\cap_{k=1}^6 E_k^c) &\leq \sum_{w_1' \neq w_1} \sum_{t^B \in [1:\hat{M}]^B} \sum_{j^B \in [1:J]^B} \prod_{i=2}^{B} \Pr\Big(E_{7i}(w_1', t_{i-1}, t_i, j_i)|\cap_{k=1}^6 E_k^c\Big) \\
&= \sum_{w_1' \neq w_1} \sum_{t_B \in [1:\hat{M}]} \sum_{j^B \in [1:J]^B} \sum_{t^{B-1} \in [1:\hat{M}]^{B-1}} \prod_{i=2}^{B} \Pr\Big(E_{7i}(w_1', t_{i-1}, t_i, j_i)|\cap_{k=1}^6 E_k^c\Big) \\
&\leq \sum_{w_1' \neq w_1} \sum_{t_B \in [1:\hat{M}]} \sum_{j^B \in [1:J]^B} \sum_{k=0}^{B-1} \binom{B-1}{k} 2^{n(B-1-k)\left[\hat{R}+\hat{\eta}\epsilon\right]} 2^{-n\left[kI_1+(B-1-k)I_2-(B-1)\epsilon\right]} \\
&= \sum_{w_1' \neq w_1} \sum_{t_B \in [1:\hat{M}]} \sum_{j_B \in [1:J]} \sum_{j^{B-1} \in [1:J]^{B-1}} \sum_{k=0}^{B-1} \binom{B-1}{k} 2^{-n\left[kI_1+(B-1-k)(I_2-\hat{R})-(B-1-k)\hat{\eta}\epsilon-(B-1)\epsilon\right]} \\
&= \sum_{w_1' \neq w_1} \sum_{t_B \in [1:\hat{M}]} \sum_{j_B \in [1:J]} \sum_{k=0}^{B-1} \binom{B-1}{k} 2^{n(B-1)\left[I(U;S|V,X_2)+\delta\epsilon\right]} 2^{-n\left[kI_1+(B-1-k)(I_2-\hat{R})-(B-1)(\hat{\eta}+1)\epsilon\right]} \\
&= \sum_{w_1' \neq w_1} \sum_{t_B \in [1:\hat{M}]} \sum_{j_B \in [1:J]} \sum_{k=0}^{B-1} \binom{B-1}{k} 2^{-n\left[k\left(I_1-I(U;S|V,X_2)\right)+(B-1-k)\left(I_2-\hat{R}-I(U;S|V,X_2)\right)-(B-1)(\hat{\eta}+\delta+1)\epsilon\right]} \\
&\leq \sum_{w_1' \neq w_1} \sum_{t_B \in [1:\hat{M}]} \sum_{j_B \in [1:J]} \sum_{k=0}^{B-1} \binom{B-1}{k} 2^{-n\left[(B-1)\min\left(I_1-I(U;S|V,X_2), I_2-\hat{R}-I(U;S|V,X_2)\right)-(B-1)(\hat{\eta}+\delta+1)\epsilon\right]} \\
&\leq M_1 \hat{M} J 2^B 2^{-n\left[(B-1)\min\left(I_1-I(U;S|V,X_2), I_2-\hat{R}-I(U;S|V,X_2)\right)-(B-1)(\hat{\eta}+\delta+1)\epsilon\right]} \\
&= 2^{-nB\left[\frac{B-1}{B}\min\left(I_1-I(U;S|V,X_2), I_2-\hat{R}-I(U;S|V,X_2)\right)-R_1-\frac{\hat{R}}{B}-\frac{I(U;S|V,X_2)}{B}-\frac{1}{n}+\left(\eta_1-\frac{\hat{\eta}}{B}-\frac{\delta}{B}-\frac{(B-1)(\hat{\eta}+\delta+1)}{B}\right)\epsilon\right]} \\
&= 2^{-nB\left[\frac{B-1}{B}\min\left(I_1-I(U;S|V,X_2), I_2-\hat{R}-I(U;S|V,X_2)\right)-R_1-\frac{\hat{R}}{B}-\frac{I(U;S|V,X_2)}{B}-\frac{1}{n}+\left(\eta_1-\delta-\hat{\eta}-\frac{B-1}{B}\right)\epsilon\right]}
\end{aligned} \quad \text{(B-19)}$$

Proof:

i) We show the first inequality in the lemma as follows:

$$I(W_1;Y^n|W_c) - I(W_1;S^n|W_c) \tag{B-30}$$
$$= \sum_{i=1}^{n} I(W_1;Y_i|W_c,Y_{i+1}^n) - I(W_1;S_i|W_c,S^{i-1}) \tag{B-31}$$
$$= \sum_{i=1}^{n} I(W_1,S^{i-1};Y_i|W_c,Y_{i+1}^n) - I(S^{i-1};Y_i|W_c,W_1,Y_{i+1}^n) - I(W_1;S_i|W_c,S^{i-1}) \tag{B-32}$$
$$= \sum_{i=1}^{n} I(W_1,S^{i-1};Y_i|W_c,Y_{i+1}^n) - I(W_1;S_i|W_c,S^{i-1}) - \sum_{i=1}^{n} I(S^{i-1};Y_i|W_c,W_1,Y_{i+1}^n) \tag{B-33}$$
$$\overset{(a)}{=} \sum_{i=1}^{n} I(W_1,S^{i-1};Y_i|W_c,Y_{i+1}^n) - I(W_1;S_i|W_c,S^{i-1}) - \sum_{i=1}^{n} I(S_i;Y_{i+1}^n|W_c,W_1,S^{i-1}) \tag{B-34}$$
$$= \sum_{i=1}^{n} I(W_1,S^{i-1};Y_i|W_c,Y_{i+1}^n) - I(S_i;W_1,Y_{i+1}^n|W_c,S^{i-1}) \tag{B-35}$$
$$= \sum_{i=1}^{n} I(W_1;Y_i|W_c,S^{i-1},Y_{i+1}^n) + I(S^{i-1};Y_i|W_c,Y_{i+1}^n) - I(S_i;Y_{i+1}^n|W_c,S^{i-1}) - I(S_i;W_1|W_c,S^{i-1},Y_{i+1}^n) \tag{B-36}$$
$$= \sum_{i=1}^{n} I(W_1;Y_i|W_c,S^{i-1},Y_{i+1}^n) - I(S_i;W_1|W_c,S^{i-1},Y_{i+1}^n) + \sum_{i=1}^{n} I(S^{i-1};Y_i|W_c,Y_{i+1}^n) - \sum_{i=1}^{n} I(S_i;Y_{i+1}^n|W_c,S^{i-1}) \tag{B-37}$$
$$\overset{(b)}{=} \sum_{i=1}^{n} I(W_1;Y_i|W_c,S^{i-1},Y_{i+1}^n) - I(S_i;W_1|W_c,S^{i-1},Y_{i+1}^n) \tag{B-38}$$
$$\overset{(c)}{=} \sum_{i=1}^{n} I(W_1;Y_i|W_c,S^{i-1},Y_{i+1}^n,X_{2,i}) - I(S_i;W_1|W_c,S^{i-1},Y_{i+1}^n,X_{2,i}) \tag{B-39}$$
$$\overset{(d)}{=} \sum_{i=1}^{n} I(\bar{U}_i;Y_i|\bar{V}_i,X_{2,i}) - I(\bar{U}_i;S_i|\bar{V}_i,X_{2,i}) \tag{B-40}$$

where (a) and (b) follow from Csiszár and Körner's Sum Identities[48]

$$\sum_{i=1}^{n} I(S^{i-1};Y_i|W_c,W_1,Y_{i+1}^n) = \sum_{i=1}^{n} I(S_i;Y_{i+1}^n|W_c,W_1,S^{i-1}) \tag{B-41}$$
$$\sum_{i=1}^{n} I(S^{i-1};Y_i|W_c,Y_{i+1}^n) = \sum_{i=1}^{n} I(S_i;Y_{i+1}^n|W_c,S^{i-1}) \tag{B-42}$$

(c) follows from the fact that X_{2i} is a deterministic function of (W_c, S^{i-1}), and (d) follows by the definition of the random variables $\bar{U}_i$ and $\bar{V}_i$ in (B-26).

ii) Similarly, we show the second inequality in the lemma as follows:

$$I(W_c,W_1;Y^n) - I(W_c,W_1;S^n) \tag{B-43}$$
$$= \sum_{i=1}^{n} I(W_c,W_1;Y_i|Y_{i+1}^n) - I(W_c,W_1;S_i|S^{i-1}) \tag{B-44}$$
$$= \sum_{i=1}^{n} I(W_c,W_1,S^{i-1};Y_i|Y_{i+1}^n) - I(S^{i-1};Y_i|W_c,W_1,Y_{i+1}^n) - I(W_c,W_1;S_i|S^{i-1}) \tag{B-45}$$
$$= \sum_{i=1}^{n} I(W_c,W_1,S^{i-1};Y_i|Y_{i+1}^n) - I(W_c,W_1;S_i|S^{i-1}) - \sum_{i=1}^{n} I(S^{i-1};Y_i|W_c,W_1,Y_{i+1}^n) \tag{B-46}$$
$$\overset{(e)}{=} \sum_{i=1}^{n} I(W_c,W_1,S^{i-1};Y_i|Y_{i+1}^n) - I(W_c,W_1;S_i|S^{i-1}) - \sum_{i=1}^{n} I(Y_{i+1}^n;S_i|W_c,W_1,S^{i-1}) \tag{B-47}$$
$$= \sum_{i=1}^{n} I(W_c,W_1,S^{i-1};Y_i|Y_{i+1}^n) - I(W_c,W_1,Y_{i+1}^n;S_i|S^{i-1}) \tag{B-48}$$
$$= \sum_{i=1}^{n} I(W_c,W_1,S^{i-1};Y_i|Y_{i+1}^n) - H(S_i|S^{i-1}) + H(S_i|W_c,W_1,S^{i-1},Y_{i+1}^n) \tag{B-49}$$
$$\overset{(f)}{=} \sum_{i=1}^{n} I(W_c,W_1,S^{i-1};Y_i|Y_{i+1}^n) - H(S_i) + H(S_i|W_c,W_1,S^{i-1},Y_{i+1}^n) \tag{B-50}$$
$$= \sum_{i=1}^{n} I(W_c,W_1,S^{i-1};Y_i|Y_{i+1}^n) - I(W_c,W_1,S^{i-1},Y_{i+1}^n;S_i) \tag{B-51}$$
$$\leq \sum_{i=1}^{n} I(W_c,W_1,S^{i-1},Y_{i+1}^n;Y_i) - I(W_c,W_1,S^{i-1},Y_{i+1}^n;S_i) \tag{B-52}$$
$$\overset{(g)}{=} \sum_{i=1}^{n} I(W_c,W_1,S^{i-1},Y_{i+1}^n,X_{2i};Y_i) - I(W_c,W_1,S^{i-1},Y_{i+1}^n,X_{2i};S_i) \tag{B-53}$$
$$\overset{(h)}{=} \sum_{i=1}^{n} I(\bar{U}_i,\bar{V}_i,X_{2i};Y_i) - I(\bar{U}_i,\bar{V}_i,X_{2i};S_i) \tag{B-54}$$

where (e) follows from Csiszár and Körner's Sum Identity (B-41), (f) follows from the fact that the state S^n is i.i.d., (g) follows from the fact that X_{2i} is a deterministic function of (W_c, S^{i-1}), and (h) follows by the definition of the random variables $\bar{U}_i$ and $\bar{V}_i$ in (B-26). ■

We continue the proof of the converse. The decoder map ψ recovers (W_c, W_1) from Y^n with vanishing average error probability P_e^n. By Fano's inequality, we have

$$H(W_c, W_1|Y^n) \le n\epsilon_n, \tag{B-55}$$

where $\epsilon_n \to 0$ as $P_e^n \to 0$.

We can bound the individual rate as

$$\begin{aligned} nR_1 &\le H(W_1|W_c) && \text{(B-56)}\\ &= I(W_1; Y^n|W_c) + H(W_1|Y^n, W_c) && \text{(B-57)}\\ &\overset{(i)}{\le} I(W_1; Y^n|W_c) + n\epsilon_n && \text{(B-58)}\\ &\overset{(j)}{=} I(W_1; Y^n|W_c) - I(W_1; S^n|W_c) + n\epsilon_n && \text{(B-59)}\\ &\overset{(k)}{=} \sum_{i=1}^{n} I(\bar{U}_i; Y_i|\bar{V}_i, X_{2,i}) - I(\bar{U}_i; S_i|\bar{V}_i, X_{2,i}) + n\epsilon_n && \text{(B-60)} \end{aligned}$$

where (i) follows by using (B-55) and the fact that $H(W_1|W_c, Y^n) \le H(W_c, W_1|Y^n)$, (j) follows from the fact that the messages are independent of each other and of the state sequence, and (k) follows by Lemma 1.

Similarly, we can bound the sum rate as

$$\begin{aligned} n(R_c + R_1) &\le H(W_c, W_1) && \text{(B-61)}\\ &= I(W_c, W_1; Y^n) + H(W_c, W_1|Y^n) && \text{(B-62)}\\ &\overset{(l)}{\le} I(W_c, W_1; Y^n) + n\epsilon_n && \text{(B-63)}\\ &\overset{(m)}{=} I(W_c, W_1; Y^n) - I(W_c, W_1; S^n) + n\epsilon_n && \text{(B-64)}\\ &\overset{(n)}{=} \sum_{i=1}^{n} I(\bar{U}_i, \bar{V}_i, X_{2i}; Y_i) - I(\bar{U}_i, \bar{V}_i, X_{2i}; S_i), && \text{(B-65)} \end{aligned}$$

where (l) follows by (B-55), (m) follows from the fact that the messages are independent of the state sequence, and (n) follows by Lemma 1.

From the above, we get that

$$\begin{aligned} R_1 &\le \frac{1}{n}\sum_{i=1}^{n} I(\bar{U}_i; Y_i|\bar{V}_i, X_{2,i}) - I(\bar{U}_i; S_i|\bar{V}_i, X_{2,i}) + \epsilon_n\\ R_c + R_1 &\le \frac{1}{n}\sum_{i=1}^{n} I(\bar{U}_i, \bar{V}_i, X_{2,i}; Y_i) - I(\bar{U}_i, \bar{V}_i, X_{2,i}; S_i) + \epsilon_n. \end{aligned} \tag{B-66}$$

The statement of the converse follows now by applying to (B-66) the standard time-sharing argument and taking the limits of large n. This is shown briefly here. We introduce a random variable T which is independent of S, and uniformly distributed over $\{1, \ldots, n\}$. Set $S = S_T$, $\bar{U} = \bar{U}_T$, $\bar{V} = \bar{V}_T$, $X_1 = X_{1,T}$, $X_2 = X_{2,T}$, and $Y = Y_T$. Then, considering the first bound in (B-66), we obtain

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} I(\bar{U}_i; Y_i|\bar{V}_i, X_{2,i}) - I(\bar{U}_i; S_i|\bar{V}_i, X_{2,i})\\ &\quad = I(\bar{U}; Y|\bar{V}, X_2, T) - I(\bar{U}; S|\bar{V}, X_2, T)\\ &\quad = I(\bar{U}, T; Y|\bar{V}, X_2, T) - I(\bar{U}, T; S|\bar{V}, X_2, T). \end{aligned} \tag{B-67}$$

Similarly, considering the second bound in (B-66), we obtain

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} I(\bar{U}_i, \bar{V}_i, X_{2,i}; Y_i) - I(\bar{U}_i, \bar{V}_i, X_{2,i}; S_i)\\ &\quad = I(\bar{U}, \bar{V}, X_2; Y|T) - I(\bar{U}, \bar{V}, X_2; S|T)\\ &\quad = I(T, \bar{U}, \bar{V}, X_2; Y) - I(T; Y) - I(T, \bar{U}, \bar{V}, X_2; S)\\ &\quad\quad + I(T; S)\\ &\quad \le I(T, \bar{U}, \bar{V}, X_2; Y) - I(T, \bar{U}, \bar{V}, X_2; S). \end{aligned} \tag{B-68}$$

The distribution on $(T, S, \bar{U}, \bar{V}, X_1, X_2, Y)$ from the given code is of the form

$$\begin{aligned} &P_{T,S,\bar{U},\bar{V},X_1,X_2,Y} =\\ &\quad Q_S P_T P_{X_2|T} P_{\bar{V}|X_2,S,T} P_{\bar{U},X_1|\bar{V},S,X_2,T} W_{Y|X_1,X_2,S}. \end{aligned} \tag{B-69}$$

Let us now define $U = (\bar{U}, T)$ and $V = (\bar{V}, T)$. Using (B-66)–(B-68), we then get

$$\begin{aligned} R_1 &\le I(U; Y|V, X_2) - I(U; S|V, X_2) + \epsilon_n\\ R_c + R_1 &\le I(U, V, X_2; Y) - I(U, V, X_2; S) + \epsilon_n, \end{aligned} \tag{B-70}$$

where the distribution on (S, U, V, X_1, X_2, Y), obtained by marginalizing (B-69) over the time sharing random variable T, satisfies $(S, U, V, X_1, X_2, Y) \in \mathcal{P}$.

So far we have shown that, for a given sequence of $(\epsilon_n, n, R_c, R_1)$, codes with ϵ_n going to zero as n goes to infinity, there exist random variables $(S, U, V, X_1, X_2, Y) \in \mathcal{P}$ such that the rate pair (R_c, R_1) essentially satisfies the inequalities in (8), i.e., $(R_c, R_1) \in \mathcal{C}$.

This completes the proof of the converse part and of Theorem 1.

Appendix C
Proof of Theorem 2

The transmission takes place in B blocks. The common message W_c is divided into $B-1$ blocks $w_{c,1}, \ldots, w_{c,B-1}$ of nR_c bits each, and the individual message W_1 is divided into $B-1$ blocks $w_{1,1}, \ldots, w_{1,B-1}$ of nR_1 bits each. For convenience, we let $w_{c,B} = w_{1,B} = 1$ (a default value). We thus have $B_{W_c} = n(B-1)R_c$, $B_{W_1} = n(B-1)R_1$, $N = nB$, $R_{W_c} = B_{W_c}/N = R_c \cdot (B-1)/B$, and $R_{W_1} = B_{W_1}/N = R_1 \cdot (B-1)/B$, where B_{W_c} is the number of common message bits, B_{W_1} is the number of individual message bits, N is the number of channel uses, and R_{W_c} and R_{W_1} are the overall rates of the common and individual messages, respectively. For fixed n, the average rate pair (R_{W_c}, R_{W_1}) over B blocks can be made as close to (R_c, R_1) as desired by making B large.

Codebook Generation: Fix a measure $P_{S,U,V,X_1,X_2,Y} \in \mathcal{P}$. Fix $\epsilon > 0$ and denote $M_c = 2^{n[R_c - \eta_c \epsilon]}$, $M_1 = 2^{n[R_1 - \eta_1 \epsilon]}$, $M_0 = 2^{n[R_0 + \eta_0 \epsilon]}$, $\hat{M} = 2^{n[\hat{R} + \hat{\eta}\epsilon]}$, $J = 2^{n[I(U;S|V,X_2) + \delta_U \epsilon]}$.

1) We generate $M_c M_0$ i.i.d. codewords $\mathbf{x}_2(w_c, s)$ indexed by $w_c = 1, \ldots, M_c$, $s = 1, \ldots, M_0$, each with i.i.d. components drawn according to P_{X_2}.
2) For each codeword $\mathbf{x}_2(w_c, s)$, we generate $\hat{M}$ i.i.d. codewords $\mathbf{v}(w_c, s, z)$ indexed by $z = 1, \ldots, \hat{M}$, each with i.i.d. components drawn according to $P_{V|X_2}$.
3) For each pair of codewords $(\mathbf{x}_2(w_c, s), \mathbf{v}(w_c, s, z))$, we generate a collection of JM_1 i.i.d. codewords

$\{\mathbf{u}(w_c, s, z, w_1, j)\}$ indexed by $w_1 = 1, \ldots, M_1$, $j = 1, \ldots, J$, each with i.i.d. components draw according to $P_{U|V,X_2}$.

4) Randomly partition the set $\{1, \ldots, \hat{M}\}$ into M_0 cells $\mathcal{C}_s$, $s \in [1, M_0]$.

Encoding: Suppose that a common message $W_c = w_c$ and an individual message $W_1 = w_1$ are to be transmitted. As we mentioned previously, message w_c is divided into $B-1$ blocks $w_{c,1}, \ldots, w_{c,B-1}$ and message w_1 is divided into $B-1$ blocks $w_{1,1}, \ldots, w_{1,B-1}$, with $(w_{c,i}, w_{1,i})$ the pair messages sent in block i. We denote by $\mathbf{s}[i]$ the channel state in block i, $i = 1, \ldots, B$. For convenience, we let $\mathbf{s}[0] = \phi$ and $z_0 = 1$ (a default value), and s_0 the index of the cell containing z_0, i.e., $z_0 \in \mathcal{C}_{s_0}$. The encoding at the beginning of the block i, $i = 1, \ldots, B$, is as follows.

Encoder 2, which has learned the state sequence $\mathbf{s}[i-1]$, knows s_{i-2} and looks for a compression index $z_{i-1} \in [1, \hat{M}]$ such that $\mathbf{v}(w_{c,i-1}, s_{i-2}, z_{i-1})$ is strongly jointly typical with $\mathbf{s}[i-1]$ and $\mathbf{x}_2(w_{c,i-1}, s_{i-2})$. If there is no such index or the observed state $\mathbf{s}[i-1]$ is not typical, z_{i-1} is set to 1 and an error is declared. If there is more than one such index z_{i-1}, choose the smallest. One can show that the probability of error of this event is arbitrarily small provided that n is large and

$$\hat{R} > I(V; S|X_2). \tag{C-1}$$

Encoder 2 then transmits the vector $\mathbf{x}_2(w_{c,i}, s_{i-1})$, where s_{i-1} is such that $z_{i-1} \in \mathcal{C}_{s_{i-1}}$.

Encoder 1 obtains $\mathbf{x}_2(w_{c,i}, s_{i-1})$ similarly. It then finds the smallest compression index $z_i \in [1, \hat{M}]$ such that $\mathbf{v}(w_{c,i}, s_{i-1}, z_i)$ is strongly jointly typical with $\mathbf{s}[i]$ and $\mathbf{x}_2(w_{c,i}, s_{i-1})$. Again, if there is no such index or the observed state $\mathbf{s}[i]$ is not typical, z_i is set to 1 and an error is declared. Let $s_i \in [1, M_0]$ such that $z_i \in \mathcal{C}_{s_i}$. Next, Encoder 1 looks for the smallest j_i such that $\mathbf{u}(w_{c,i}, s_{i-1}, z_i, w_{1,i}, j_i)$ is jointly typical with $\mathbf{s}[i]$, $\mathbf{x}_2(w_{c,i}, s_{i-1})$ and $\mathbf{v}(w_{c,i}, s_{i-1}, z_i)$. Denote this j_i by $j_i^\star = j(\mathbf{s}[i], w_{c,i}, s_{i-1}, z_i, w_{1,i})$. If such $j_i^\star$ is not found, an error is declared and $j(\mathbf{s}[i], w_{c,i}, s_{i-1}, z_i, w_{1,i})$ is set to $j_i = J$. Encoder 1 then transmits a vector $\mathbf{x}_1[i]$ which is drawn i.i.d. conditionally given $\mathbf{s}[i]$, $\mathbf{u}(w_{c,i}, s_{i-1}, z_i, w_{1,i}, j_i^\star)$, $\mathbf{v}(w_{c,i}, s_{i-1}, z_i)$ and $\mathbf{x}_2(w_{c,i}, s_{i-1})$ (using the conditional measure $P_{X_1|S,U,V,X_2}$ induced by $P_{S,U,V,X_1,X_2,Y} \in \mathcal{P}$).

Decoding: Let $\mathbf{y}[i]$ denote the information received at the receiver at block i, $i = 1, \ldots, B$. The receiver collects these information until the last block of transmission is completed. The decoder then performs Willem's backward decoding [49], by first decoding the pair $(w_{c,B-1}, w_{1,B-1})$ from $\mathbf{y}[B-1]$.

1) Decoding in Block $B-1$: The decoding of the pair $(w_{c,B-1}, w_{1,B-1})$ is performed in four steps, as follows.

Step (a): The decoder knows $w_{c,B} = 1$ and looks for the unique cell index $\hat{s}_{B-1}$ such that the vector $\mathbf{x}_2(w_{c,B}, \hat{s}_{B-1})$ is jointly typical with $\mathbf{y}[B]$. The decoding operation in this step incurs small probability of error as long as n is sufficiently large and

$$R_0 < I(X_2; Y). \tag{C-2}$$

Step (b): The decoder now knows $\hat{s}_{B-1}$ (i.e., the index of the cell in which the compression index z_{B-1} lies). It then decodes message $w_{c,B-1}$ by looking for the unique $\hat{w}_{c,B-1}$ such that $\mathbf{x}_2(\hat{w}_{c,B-1}, s_{B-2})$, $\mathbf{v}(\hat{w}_{c,B-1}, s_{B-2}, z_{B-1})$, $\mathbf{u}(\hat{w}_{c,B-1}, s_{B-2}, z_{B-1}, w_{1,B-1}, j_{B-1})$, and $\mathbf{y}[B-1]$ are jointly typical for some $s_{B-2} \in [1, M_0]$, $w_{1,B-1} \in [1, M_1]$, $j_{B-1} \in [1, J]$, and $z_{B-1} \in \mathcal{C}_{\hat{s}_{B-1}}$. One can show that the decoder obtains the correct $w_{c,B-1}$ as long as n and B are large and

$$R_0 + (\hat{R} - R_0) + R_c + R_1 \le I(U, V, X_2; Y) - I(U; S|V, X_2). \tag{C-3}$$

Step (c): The decoder knows $\hat{w}_{c,B-1}$ and can again obtain the correct s_{B-2} if n is large and (C-2) is true. This is accomplished by looking for the unique $\hat{s}_{B-2}$ such that the vector $\mathbf{x}_2(\hat{w}_{c,B-1}, \hat{s}_{B-2})$ is jointly typical with $\mathbf{y}[B-1]$.

Step (d): Finally, the decoder, which now knows message $\hat{w}_{c,B-1}$ and the cell index $\hat{s}_{B-2}$ (but not the exact compression index z_{B-1}), estimates $w_{1,B-1}$ using $\mathbf{y}[B-1]$. It declares that $\hat{w}_{1,B-1}$ was sent if there exists a unique $\hat{w}_{1,B-1}$ such that $\mathbf{x}_2(\hat{w}_{c,B-1}, \hat{s}_{B-2})$, $\mathbf{v}(\hat{w}_{c,B-1}, \hat{s}_{B-2}, z'_{B-1})$, $\mathbf{u}(\hat{w}_{c,B-1}, \hat{s}_{B-2}, z'_{B-1}, \hat{w}_{1,B-1}, j_{B-1})$, and $\mathbf{y}[B-1]$ are jointly typical for some $z'_{B-1} \in \mathcal{C}_{\hat{s}_{B-1}}$ and $j_{B-1} \in [1, J]$.

- If $z'_{B-1} = z_{B-1}$, the decoder finds the correct $w_{1,b-1}$ for sufficiently large n if

$$R_1 \le I(U; Y|V, X_2) - I(U; S|V, X_2). \tag{C-4}$$

- If $z'_{B-1} \neq z_{B-1}$, the decoder finds the correct $w_{1,b-1}$ for sufficiently large n if

$$(\hat{R} - R_0) + R_1 \le I(U, V; Y|X_2) - I(U; S|V, X_2). \tag{C-5}$$

2) Decoding in Block b, $b = B-1, B-2, \ldots, 2$: Next, for b ranging from $B-1$ to 2, the decoding of the pair $(w_{c,b-1}, w_{1,b-1})$ is performed similarly, in five steps, by using the information $\mathbf{y}[b]$ received in block b and the information $\mathbf{y}[b-1]$ received in block $b-1$. More specifically, this is done as follows.

Step (a): The decoder knows $w_{c,b}$ and looks for the unique cell index $\hat{s}_{b-1}$ such that the vector $\mathbf{x}_2(w_{c,b}, \hat{s}_{b-1})$ is jointly typical with $\mathbf{y}[b]$. The decoding error in this step is small for sufficiently large n if (C-2) is true.

Step (b): The decoder knows $\hat{s}_{b-1}$ and decodes message $w_{c,b-1}$ from $\mathbf{y}[b]$. It looks for the unique $\hat{w}_{c,b-1}$ such that $\mathbf{x}_2(\hat{w}_{c,b-1}, s_{b-2})$, $\mathbf{v}(\hat{w}_{c,b-1}, s_{b-2}, z_{b-1})$, $\mathbf{u}(\hat{w}_{c,b-1}, s_{b-2}, z_{b-1}, w_{1,b-1}, j_{b-1})$ and $\mathbf{y}[b-1]$ are jointly typical for some $s_{b-2} \in [1, M_0]$, $w_{1,b-1} \in [1, M_1]$, $j_{b-1} \in [1, J]$ and $z_{b-1} \in \mathcal{C}_{\hat{s}_{b-1}}$. One can show that the decoding error in this step is small for sufficiently large n if (C-3) is true.

Step (c): The decoder knows $\hat{w}_{c,b-1}$ and obtains $\hat{s}_{b-2}$ by looking for the unique $\hat{s}_{b-2}$ such that the vector $\mathbf{x}_2(\hat{w}_{c,b-1}, \hat{s}_{b-2})$ is jointly typical with $\mathbf{y}[b-1]$. For sufficiently large n, the decoder obtains the correct s_{b-2} with high probability if (C-2) is true.

Step (d): Finally, the decoder, which now knows message $\hat{w}_{c,b-1}$ and the cell index $\hat{s}_{b-2}$ (but not the exact compression index z_{b-1}), estimates message $w_{1,b-1}$ using $\mathbf{y}[b-1]$. It declares that $\hat{w}_{1,b-1}$ was sent if there exists a unique

$\hat{w}_{1,b-1}$ such that $\mathbf{x}_2(\hat{w}_{c,b-1}, \hat{s}_{b-2})$, $\mathbf{v}(\hat{w}_{c,b-1}, \hat{s}_{b-2}, z'_{b-1})$, $\mathbf{u}(\hat{w}_{c,b-1}, \hat{s}_{b-2}, z'_{b-1}, \hat{w}_{1,b-1}, j_{b-1})$, and $\mathbf{y}[b-1]$ are jointly typical for some $z'_{b-1} \in \mathcal{C}_{\hat{s}_{b-1}}$ and $j_{b-1} \in [1, J]$.

- If $z'_{b-1} = z_{b-1}$, the decoder finds the correct $w_{1,b-1}$ for sufficiently large n if (C-4) is true.
- If $z'_{b-1} \neq z_{b-1}$, the decoder finds the correct $w_{1,b-1}$ for sufficiently large n if (C-5) is true.

Fourier–Motzkin Elimination: From the above, we get that the error probability is small provided that n is large and

$$R_0 < I(X_2; Y) \quad \text{(C-6a)}$$
$$\hat{R} > I(V; S|X_2) \quad \text{(C-6b)}$$
$$R_1 \leq I(U; Y|V, X_2) - I(U; S|V, X_2) \quad \text{(C-6c)}$$
$$(\hat{R} - R_0) + R_1 \leq I(U, V; Y|X_2) - I(U; S|V, X_2) \quad \text{(C-6d)}$$
$$R_c + R_1 + \hat{R} \leq I(U, V, X_2; Y) - I(U; S|V, X_2). \quad \text{(C-6e)}$$

We now apply Fourier–Motzkin Elimination (FME) to project out R_0 and $\hat{R}$ from (C-6). Projecting out R_0 from (C-6), we get

$$\hat{R} > I(V; S|X_2) \quad \text{(C-7a)}$$
$$R_1 \leq I(U; Y|V, X_2) - I(U; S|V, X_2) \quad \text{(C-7b)}$$
$$\hat{R} + R_1 \leq I(U, V, X_2; Y) - I(U; S|V, X_2) \quad \text{(C-7c)}$$
$$R_c + R_1 + \hat{R} \leq I(U, V, X_2; Y) - I(U; S|V, X_2). \quad \text{(C-7d)}$$

Note that the inequality (C-7c) can be implied by (C-7d) since $R_c \geq 0$, and, so, is redundant in (C-7). Finally, projecting out $\hat{R}$ from the remaining system, we obtain

$$R_1 \leq I(U; Y|V, X_2) - I(U; S|V, X_2) \quad \text{(C-8)}$$
$$R_c + R_1 \leq I(U, V, X_2; Y) - I(U, V, X_2; S). \quad \text{(C-9)}$$

This completes the proof of Theorem 2.

Appendix D
Proof of Corollary 1

A. Converse Part

Investigating the proof of Theorem 1 in Appendix B, it can be seen that the auxiliary random variables U and V satisfy tacitly the condition

$$I(V, X_2; Y) - I(V, X_2; S) \geq 0. \quad \text{(D-1)}$$

This can be seen by noticing that (with the notation of Appendix B)

$$I(W_1; Y^n|W_c) = \sum_{i=1}^{n} I(\bar{U}_i; Y_i|\bar{V}_i, X_{2,i}) - I(\bar{U}_i; S_i|\bar{V}_i, X_{2,i}) \quad \text{(D-2)}$$

$$I(W_c, W_1; Y^n) \leq \sum_{i=1}^{n} I(\bar{U}_i, \bar{V}_i, X_{2i}; Y_i) - I(\bar{U}_i, \bar{V}_i, X_{2i}; S_i) \quad \text{(D-3)}$$

and then observing that $I(W_1; Y^n|W_c) \leq I(W_c, W_1; Y^n)$, which together yield

$$\sum_{i=1}^{n} I(\bar{V}_i, X_{2i}; Y_i) - I(\bar{V}_i, X_{2i}; S_i) \geq 0; \quad \text{(D-4)}$$

and, so, after standard single letterization, the condition (D-1).

B. Direct Part

The codebook generation and the encoding process remain exactly as in the proof of Theorem 2 in Appendix C. The decoding at the receiver is modified in a way to get the compression indices decoded uniquely, as follows (with the notation of Appendix C).

Decoding: Let $\mathbf{y}[i]$ denote the information received at the receiver at block i, $i = 1, \ldots, B$. The receiver collects these information until the last block of transmission is completed. The decoder then performs Willem's backward decoding [49], by first decoding the pair $(w_{c,B-1}, w_{1,B-1})$ from $\mathbf{y}[B-1]$.

1) Decoding in Block $B-1$: The decoding of the pair $(w_{c,B-1}, w_{1,B-1})$ is performed in five steps, as follows.

Step (a): The decoder knows $w_{c,B} = 1$ and looks for the unique cell index $\hat{s}_{B-1}$ such that the vector $\mathbf{x}_2(w_{c,B}, \hat{s}_{B-1})$ is jointly typical with $\mathbf{y}[B]$. This decoding operation incurs small probability of error as long as n is sufficiently large and

$$R_0 < I(X_2; Y). \quad \text{(D-5)}$$

Step (b): The decoder now knows $\hat{s}_{B-1}$ (i.e., the index of the cell in which the compression index z_{B-1} lies). It then decodes message $w_{c,B-1}$ by looking for the unique $\hat{w}_{c,B-1}$ such that $\mathbf{x}_2(\hat{w}_{c,B-1}, s_{B-2})$, $\mathbf{v}(\hat{w}_{c,B-1}, s_{B-2}, z_{B-1})$, $\mathbf{u}(\hat{w}_{c,B-1}, s_{B-2}, z_{B-1}, w_{1,B-1}, j_{B-1})$, and $\mathbf{y}[B-1]$ are jointly typical for some $s_{B-2} \in [1, M_0]$, $w_{1,B-1} \in [1, M_1]$, $j_{B-1} \in [1, J]$ and $z_{B-1} \in \mathcal{C}_{\hat{s}_{B-1}}$. One can show that the decoder obtains the correct $w_{c,B-1}$ as long as n and B are large and

$$R_0 + (\hat{R} - R_0) + R_c + R_1 \leq I(U, V, X_2; Y) - I(U; S|V, X_2). \quad \text{(D-6)}$$

Step (c): The decoder knows $\hat{w}_{c,B-1}$ and can again obtain the correct s_{B-2} if n is large and (D-5) is true. This is accomplished by looking for the unique $\hat{s}_{B-2}$ such that the vector $\mathbf{x}_2(\hat{w}_{c,B-1}, \hat{s}_{B-2})$ is jointly typical with $\mathbf{y}[B-1]$.

Step (d): The decoder calculates a set $\mathcal{L}(\mathbf{y}[B-1])$ of z_{B-1} such that $z_{B-1} \in \mathcal{L}(\mathbf{y}[B-1])$ if $\mathbf{v}(\hat{w}_{c,B-1}, \hat{s}_{B-2}, z_{B-1})$, $\mathbf{x}_2(\hat{w}_{c,B-1}, \hat{s}_{B-2})$, $\mathbf{y}[B-1]$ are jointly typical. It then declares that z_{B-1} was sent in block $B-1$ if

$$\hat{z}_{B-1} \in \mathcal{C}_{\hat{s}_{B-1}} \cap \mathcal{L}(\mathbf{y}[B-1]). \quad \text{(D-7)}$$

One can show that $\hat{z}_{B-1} = z_{B-1}$ with arbitrarily high probability provided that n is sufficiently large and

$$\hat{R} < I(V; Y|X_2) + R_0. \quad \text{(D-8)}$$

Step (e): Finally, the decoder, which now knows message $\hat{w}_{c,B-1}$, the cell index $\hat{s}_{B-2}$ and the compression index $z_{B-1} \in \mathcal{C}_{\hat{s}_{B-1}}$, estimates $w_{1,B-1}$ using $\mathbf{y}[B-1]$. It declares that $\hat{w}_{1,B-1}$ was sent if there exists a unique $\hat{w}_{1,B-1}$ such that $\mathbf{x}_2(\hat{w}_{c,B-1}, \hat{s}_{B-2})$, $\mathbf{v}(\hat{w}_{c,B-1}, \hat{s}_{B-2}, \hat{z}_{B-1})$, $\mathbf{u}(\hat{w}_{c,B-1}, \hat{s}_{B-2}, \hat{z}_{B-1}, \hat{w}_{1,B-1}, j_{B-1})$, and $\mathbf{y}[B-1]$ are jointly typical for some $j_{B-1} \in [1, J]$. One can show that the decoder obtains the correct $w_{1,B-1}$ as long as n is large and

$$R_1 \leq I(U;Y|V,X_2) - I(U;S|V,X_2). \tag{D-9}$$

2) Decoding in Block b, $b = B-1, B-2, \ldots, 2$: Next, for b ranging from $B-1$ to 2, the decoding of the pair $(w_{c,b-1}, w_{1,b-1})$ is performed similarly, in five steps, by using the information $\mathbf{y}[b]$ received in block b and the information $\mathbf{y}[b-1]$ received in block $b-1$. More specifically, this is done as follows.

Step (a): The decoder knows $w_{c,b}$ and looks for the unique cell index $\hat{s}_{b-1}$ such that the vector $\mathbf{x}_2(w_{c,b}, \hat{s}_{b-1})$ is jointly typical with $\mathbf{y}[b]$. The decoding error in this step is small for sufficiently large n if (D-5) is true.

Step (b): The decoder knows $\hat{s}_{b-1}$ and decodes message $w_{c,b-1}$ from $\mathbf{y}[b]$. It looks for the unique $\hat{w}_{c,b-1}$ such that $\mathbf{x}_2(\hat{w}_{c,b-1}, s_{b-2})$, $\mathbf{v}(\hat{w}_{c,b-1}, s_{b-2}, z_{b-1})$, $\mathbf{u}(\hat{w}_{c,b-1}, s_{b-2}, z_{b-1}, w_{1,b-1}, j_{b-1})$, and $\mathbf{y}[b-1]$ are jointly typical for some $s_{b-2} \in [1, M_0]$, $w_{1,b-1} \in [1, M_1]$, $j_{b-1} \in [1, J]$ and $z_{b-1} \in \mathcal{C}_{\hat{s}_{b-1}}$. One can show that the decoding error in this step is small for sufficiently large n if (D-6) is true.

Step (c): The decoder knows $\hat{w}_{c,b-1}$ and obtains $\hat{s}_{b-2}$ by looking for the unique $\hat{s}_{b-2}$ such that the vector $\mathbf{x}_2(\hat{w}_{c,b-1}, \hat{s}_{b-2})$ is jointly typical with $\mathbf{y}[b-1]$. For sufficiently large n, the decoder obtains the correct s_{b-2} with high probability if (D-5) is true.

Step (d): The decoder calculates a set $\mathcal{L}(\mathbf{y}[b-1])$ of z_{b-1} such that $z_{b-1} \in \mathcal{L}(\mathbf{y}[b-1])$ if $\mathbf{v}(\hat{w}_{c,b-1}, \hat{s}_{b-2}, z_{b-1})$, $\mathbf{x}_2(\hat{w}_{c,b-1}, \hat{s}_{b-2})$, $\mathbf{y}[b-1]$ are jointly typical. It then declares that z_{b-1} was sent in block $b-1$ if

$$\hat{z}_{b-1} \in \mathcal{C}_{\hat{s}_{b-1}} \cap \mathcal{L}(\mathbf{y}[b-1]). \tag{D-10}$$

One can show that, for large n, $\hat{z}_{b-1} = z_{b-1}$ with arbitrarily high probability provided that (D-8) is true.

Step (e): Finally, the decoder knows message $\hat{w}_{c,b-1}$, the cell index $\hat{s}_{b-2}$, and the compression index $z_{b-1} \in \mathcal{C}_{\hat{s}_{b-1}}$, and estimates $w_{1,b-1}$ using $\mathbf{y}[b-1]$. It declares that $\hat{w}_{1,b-1}$ was sent if there exists a unique $\hat{w}_{1,b-1}$ such that $\mathbf{x}_2(\hat{w}_{c,b-1}, \hat{s}_{b-2})$, $\mathbf{v}(\hat{w}_{c,b-1}, \hat{s}_{b-2}, \hat{z}_{b-1})$, $\mathbf{u}(\hat{w}_{c,b-1}, \hat{s}_{b-2}, \hat{z}_{b-1}, \hat{w}_{1,b-1}, j_{b-1})$, and $\mathbf{y}[b-1]$ are jointly typical for some $j_{b-1} \in [1, J]$. One can show that the decoding error in this step is small for sufficiently large n if (D-9) is true.

Fourier–Motzkin Elimination: From the above, we get that the error probability is small provided that n is large and

$$R_0 < I(X_2;Y) \tag{D-11a}$$
$$\hat{R} < I(V;Y|X_2) + R_0 \tag{D-11b}$$
$$\hat{R} > I(V;S|X_2) \tag{D-11c}$$
$$R_1 \leq I(U;Y|V,X_2) - I(U;S|V,X_2) \tag{D-11d}$$
$$R_c + R_1 + \hat{R} \leq I(U,V,X_2;Y) - I(U;S|V,X_2). \tag{D-11e}$$

Applying FME to project out $\hat{R}$ and R_0 from (D-11), we get

$$0 \leq I(V,X_2;Y) - I(V,X_2;S) \tag{D-12a}$$
$$R_1 \leq I(U;Y|V,X_2) - I(U;S|V,X_2) \tag{D-12b}$$
$$R_c + R_1 \leq I(U,V,X_2;Y) - I(U,V,X_2;S). \tag{D-12c}$$

C. Bounds on $|\mathcal{V}|$ and $|\mathcal{U}|$

It remains to show that the rate pair (20) is not altered if one restricts the random variables V and U to have their alphabet sizes limited as indicated in (22). This is done by a standard application of the support lemma [46, p. 310], essentially by following the lines in the proof of Theorem 1 in Appendix B and noticing that, this time, because of the additional nonnegativity constraint, one more functional needs to be preserved in bounding the cardinality of V,

$$I_\mu(V,X_2;Y) - I_\mu(V,X_2;S) = H_\mu(Y) - H_\mu(S) + H_\mu(X_2,S|V) - H_\mu(X_2,Y|V). \tag{D-13}$$

This concludes the proof of Corollary 1.

APPENDIX E
PROOF OF THEOREM 3

We prove that for any (M_c, M_1, n, ϵ) code consisting of a mapping $\phi_1 : \mathcal{W}_c \times \mathcal{W}_1 \times \mathcal{S}^n \longrightarrow \mathcal{X}_1^n$ at Encoder 1, a sequence of mappings $\phi_{2,i} : \mathcal{W}_c \times \mathcal{S}^{i-1} \longrightarrow \mathcal{X}_2, i = 1, \ldots, n$, at Encoder 2, and a mapping $\psi : \mathcal{Y}^n \longrightarrow \mathcal{W}_c \times \mathcal{W}_1$ at the decoder with average error probability $P_e^n \to 0$ as $n \to 0$ and rates $R_c = n^{-1} \log_2 M_c$ and $R_1 = n^{-1} \log_2 M_1$, the rate pair (R_c, R_1) must satisfy (23).

Fix n and consider a given code of block length n. The joint probability mass function on $\mathcal{W}_c \times \mathcal{W}_1 \times \mathcal{S}^n \times \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}^n$ is given by (E-1), shown at the bottom of the page, where

$$P(w_c, w_1, s^n, x_1^n, x_2^n, y^n) = P(w_c, w_1) \prod_{i=1}^{n} P(s_i) P(x_{1i}|w_c, w_1, s^n) P(x_{2i}|w_c, s^{i-1}) P(y_i|x_{1i}, x_{2i}, s_i), \tag{E-1}$$

$P(x_{1i}|w_c, w_1, s^n)$ is equal 1 if $x_{1i} = f_1(w_c, w_1, s^n)$ and 0 otherwise, and $P(x_{2i}|w_c, s^{i-1})$ is equal 1 if $x_{2i} = f_2(w_c, s^{i-1})$ and 0 otherwise.

The proof of the bound on R_1 follows trivially by revealing the state S^n to the decoder.

The proof of the bound on the sum rate $R_c + R_1$ is as follows. The decoder map ψ recovers (W_c, W_1) from Y^n with vanishing average error probability. By Fano's inequality, we have

$$H(W_c, W_1|Y^n) \leq n\epsilon_n, \tag{E-2}$$

where $\epsilon_n \to 0$ as $P_e^n \to 0$

$$\begin{aligned}
n(R_c + R_1) &= H(W_c, W_1) \\
&= I(W_c, W_1; Y^n) + H(W_c, W_1|Y^n) \\
&\overset{(a)}{\leq} I(W_c, W_1; Y^n) + n\epsilon_n \\
&= I(W_c, W_1, S^n; Y^n) - I(S^n; Y^n|W_c, W_1) + n\epsilon_n \\
&= \Big(\sum_{i=1}^{n} I(W_c, W_1, S^n; Y_i|Y^{i-1})\Big) \\
&\quad - H(S^n|W_c, W_1) + H(S^n|W_c, W_1, Y^n) + n\epsilon_n \\
&\overset{(b)}{=} \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - H(Y_i|W_c, W_1, S^n, Y^{i-1}) \\
&\quad - H(S_i) + H(S_i|W_c, W_1, Y^n, S^{i-1}) + n\epsilon_n \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} H(Y_i) - H(Y_i|X_{1,i}, X_{2,i}, S_i) - H(S_i) \\
&\quad + H(S_i|W_c, W_1, Y^n, S^{i-1}, X_{2,i}) + n\epsilon_n \\
&\overset{(d)}{\leq} \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}, S_i; Y_i) - H(S_i) + H(S_i|X_{2,i}, Y_i) \\
&\quad + n\epsilon_n \\
&= \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}, S_i; Y_i) - I(S_i; X_{2,i}, Y_i) + n\epsilon_n \\
&= \sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; Y_i|S_i) - I(S_i; X_{2,i}|Y_i) + n\epsilon_n,
\end{aligned} \tag{E-3}$$

where (a) follows from Fano's inequality, (b) follows from the fact that the state S^n is i.i.d. and is independent of the messages, (c) follows from $(W_c, W_1, S^n, Y^{i-1}) \leftrightarrow (X_{1,i}, X_{2,i}, S_i) \leftrightarrow Y_i$, and the fact that $X_{2,i}$ is a deterministic function of (W_c, S^{i-1}), and (d) follows from the fact that conditioning reduces entropy.

Finally, we obtain the desired bound from (E-3) by standard single letterization [46].

Appendix F
Proof of Corollary 2

Relaxing the constraint on R_1 in Theorem 1, we obtain

$$C = \max I(U, V, X_2; Y) - I(U, V, X_2; S) \tag{F-1}$$

where the maximization is over joint measures $P_{S,U,V,X_1,X_2,Y}$ of the form

$$P_{S,U,V,X_1,X_2,Y} = Q_S P_{X_2} P_{V|S,X_2} P_{U,X_1|S,V,X_2}. \tag{F-2}$$

The corollary then follows by substituting $K = (U, V)$, and noticing that the distribution on (S, K, X_1, X_2, Y) is given by

$$\begin{aligned}
P_{S,K,X_1,X_2,Y} &= P_{S,U,V,X_1,X_2,Y} && \text{(F-3)}\\
&= Q_S P_{X_2} P_{V|S,X_2} P_{U,X_1|S,V,X_2} && \text{(F-4)}\\
&= Q_S P_{X_2} P_{U,V|S,X_2} P_{X_1|S,U,V,X_2} && \text{(F-5)}\\
&= Q_S P_{X_2} P_{K|S,X_2} P_{X_1|S,K,X_2}. && \text{(F-6)}
\end{aligned}$$

Appendix G
Proof of Theorem 4

A. Direct Part

The achievability follows by ignoring the strictly causal part of the state at Encoder 2 and using the generalized dirty paper coding scheme of [5, Th. 7].

B. Converse Part

For the converse part, we use the outer bound of Theorem 3 for the discrete MAC which can be readily extended to memoryless channels with discrete time and continuous alphabets using standard techniques [50]. Then, we obtain an outer bound on the capacity region of the Gaussian MAC in terms of the closure of the convex hull of the set of rate pairs (R_c, R_1) satisfying

$$\begin{aligned}
R_1 &\leq I(X_1; Y|S, X_2), \\
R_c + R_1 &\leq I(X_1, X_2; Y|S) - I(X_2; S|Y),
\end{aligned} \tag{G-1}$$

for some probability distribution of the form $P_{S,X_1,X_2,Y} = Q_S P_{X_2} P_{X_1|X_2,S} W_{Y|X_1,X_2,S}$ such that $\mathbb{E}[X_1^2] \leq P_1$ and $\mathbb{E}[X_2^2] \leq P_2$. The rest of the converse proof follows by reasoning and using algebra similar to in the proofs of [5, Th. 7] and [11, Th. 4], and is omitted for brevity.

References

[1] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[2] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Development*, vol. 2, pp. 289–293, Oct. 1958.

[3] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Control Inf. Theory*, vol. 9, pp. 19–31, 1980.

[4] G. Keshet, Y. Steinberg, and N. Merhav, "Channel coding in the presence of side information: Subject review," *Found. Trends Commun. Inf. Theory*, 2008.

[5] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdù, "Cooperative multiple access encoding with states available at one transmitter," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4448–4469, Oct. 2008.

[6] S. Kotagiri and J. N. Laneman, "Multiaccess channels with state known to some encoders and independent messages," *EURASIP J. Wireless Commnun. Network.*, vol. 2008, p. 450680, 2008.

[7] A. Zaidi, S. Kotagiri, J. N. Laneman, and L. Vandendorpe, "Multiaccess channels with state known to one encoder: Another case of degraded message sets," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, Korea, Jun./Jul. 2009, pp. 2376–2380.

[8] A. Khisti, U. Erez, A. Lapidoth, and G. Wornell, "Carbon copying onto dirty paper," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1814–1827, May 2007.

[9] T. Philosoph, A. Khisti, U. Erez, and R. Zamir, "Lattice strategies for the dirty multiple access channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, Jun. 2007, pp. 386–390.

[10] A. Zaidi, S. Kotagiri, J. N. Laneman, and L. Vandendorpe, "Cooperative relaying with state at the relay," in *Proc. IEEE Inf. Theory Workshop*, Porto, Portugal, May 2008, pp. 139–143.

[11] A. Zaidi, S. Kotagiri, J. N. Laneman, and L. Vandendorpe, "Cooperative relaying with state available non-causally at the relay," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2272–2298, May 2010.

[12] A. Zaidi and L. Vandendorpe, "Lower bounds on the capacity of the relay channel with states at the source," *EURASIP J. Wireless Commnun. Network.*, vol. 2009, p. 634296, 2009.

[13] Y. Cemal and Y. Steinberg, "The multiple-access channel with partial state information at the encoders," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3992–4003, Nov. 2005.

[14] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2867–2877, Aug. 2005.

[15] A. Lapidoth and Y. Steinberg, "The multiple access channel with causal and strictly causal side information at the encoders," in *Proc. Int. Zurich Semin. Commun.*, Zurich, Switzerland, Mar. 2010, pp. 13–16.

[16] A. Lapidoth and Y. Steinberg, "The multiple access channel with two independent states each known causally at one encoder," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 2010, pp. 480–484.

[17] H. Permuter, S. Shamai (Shitz), and A. Somekh-Baruch, "Message and state cooperation in multiple access channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6379–6396, Oct. 2011.

[18] M. Li, O. Simeone, and A. Yener, "Leveraging strictly causal state information at the encoders for multiple access channels," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Jul./Aug. 2011, pp. 2806–2810.

[19] M. Li, O. Simeone, and A. Yener, Message and state cooperation in a relay channel when only the relay knows the state [Online]. Available: http://arxiv.org/abs/1102.0768 2011

[20] B. Akhbari, M. Mirmohseni, and M. R. Aref, "Compress-and-forward strategy for the relay channel with non-causal state information," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, Korea, Jun./Jul. 2009, pp. 1169–1173.

[21] M. N. Khormuji and M. Skoglund, "On cooperative downlink transmission with frequency reuse," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, Korea, Jun./Jul. 2009, pp. 849–853.

[22] G. Como and S. Yüksel, On the capacity of memoryless finite state multiple-access channels with asymmetric state information at the encoders [Online]. Available: http://arxiv.org/abs/1011.1012.1912 2011

[23] N. Şen, G. Como, S. Yüksel, and F. Alajaji, "On the capacity of memoryless finite-state multiple access channels with asymmetric noisy state information at the encoders," presented at the 49th Annu. Conf. Commun., Control, Comput., Monticello, IL, USA, Sep. 2011.

[24] R. Khosravi-Farsani and F. Marvasti, "Capacity bounds for multiuser channels with non-causal channel state information at the transmitters," in *Proc. IEEE Inf. Theory Workshop*, Paraty, Brasil, Oct. 2011, pp. 195–199.

[25] S. I. Bross and A. Lapidoth, "The state-dependent multiple-access channel with states available at a cribbing encoder," presented at the IEEE 26th Convent. Electr. Electron. Eng. Israel, Eilat, Israel, 2010.

[26] S. Jafar, "Capacity with causal and noncausal side-information: A unified view," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5468–5474, Dec. 2006.

[27] S. Sigurjonsson and Y. H. Kim, "On multiple user channels with state information at the transmitters," in *Proc. IEEE Int. Symp. Inf. Theory*, Sep. 2005, pp. 72–76.

[28] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, pp. 8–19, 1956.

[29] G. Dueck, "Partial feedback for two-way and broadcast channels," *Inf. Control*, vol. 46, pp. 1–15, 1980.

[30] S. H. Lim, Y.-H. Kim, A. E. Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[31] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[32] A. Lapidoth and Y. Steinberg, A note on multiple access channels with strictly causal state information Jun. 2011 [Online]. Available: http://arxiv.org/abs/1106.0380

[33] A. Avestimeher, S. Diggavi, and D. Tse, "Wireless network information flow: A determenistic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.

[34] A. Ozgur and S. N. Diggavi, "Approximately achieving Gaussian relay network capacity with lattice codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 2010, pp. 669–673.

[35] A. Zaidi, P. Piantanida, and S. Shamai (Shitz), "Multiple access channel with states known noncausally at one encoder and only strictly causally at the other encoder," in *Proc. IEEE Int. Symp. Inf. Theory*, Saint-Petersburg, Russia, 2011, pp. 2801–2805.

[36] A. Zaidi, P. Piantanida, and S. Shamai (Shitz), "Wyner-Ziv type versus noisy network coding for a state-dependent MAC," presented at the IEEE Int. Symp. Inf. Theory, Cambridge, MA, USA, Jul. 2012.

[37] X. Wu and L.-L. Xie, On the optimal compressions in the compress-and-forward relay schemes [Online]. Available: http://arxiv.org/abs/1009.5959 Feb. 2011

[38] P. Zhong, A. A. Haija, and M. Vu, On compress-and-forward without Wyner-Ziv binning for relay networks [Online]. Available: http://arxiv.org/abs/1111.2837 Nov. 2011

[39] G. Kramer and J. Hou, "On message lengths for noisy network coding," presented at the IEEE Inf. Theory Workshop, Paraty, Brasil, Oct. 2011.

[40] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.

[41] A. Raja and P. Viswanath, Compress-and-forward scheme for a relay network: Approximate optimality and connection to algebraic flows [Online]. Available: http://arxiv.org/abs/1012.0416v3 Jan. 2012

[42] M. Katz and S. Shamai (Shitz), "Cooperative schemes for a source and an occasional nearby relay in wireless networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5139–5160, Nov. 2009.

[43] S. S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1181–1203, May 2003.

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

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

[46] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. London, U.K.: Academic, 1981.

[47] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[48] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 339–348, 1978.

[49] F. M. J. Willems, "Information theoretical results for the discrete memoryless multiple access channel," Doctor in de Wetenschappen Proefschrift, Faculteit der Wetenschappen, Dept. Wiskunde, Katholieke Univ. Leuven, Leuven, Belgium, 1982.

[50] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

Abdellatif Zaidi received the B.S. degree in Electrical Engineering from Écaron;cole Nationale Supérieure de Techniques Avancés, ENSTA ParisTech, France in 2002 and the M. Sc. and Ph.D. degrees in Electrical Engineering from École Nationale Supérieure des Télécommunications, TELECOM ParisTech, Paris, France in 2002 and 2005, respectively.

From December 2002 to December 2005, he was with the Communications an Electronics Dept., TELECOM ParisTech, Paris, France and the Signals and Systems Lab., CNRS/Supélec, France pursuing his PhD degree. From May 2006 to September 2010, he was at École Polytechnique de Louvain, Université Catholique de Louvain, Belgium, working as a research assistant. Dr. Zaidi was "Research Visitor" at the University of Notre Dame, Indiana, USA, during fall 2007 and Spring 2008. He is now, an assistant professor at Universite Paris-Est Marne-La-Vallee, France. He is a member of Laboratoire d'Informatique Gaspard Monge (LIGM).

Dr. Zaidi's research interests cover a broad range of topics from network information theory and communication theory. Of particular interest are the problems of multiterminal information theory, Shannon theory, relaying and cooperation, network coding, physical layer security, source coding and interference mitigation in multiuser channels.

Pablo Piantanida received the B.Sc. and M.Sc degrees (with honors) in Electrical Engineering from the University of Buenos Aires (Argentina), in 2003, and the Ph.D. from the Paris-Sud University (France) in 2007. In 2006, he has been with the Department of Communications and Radio-Frequency Engineering at Vienna University of Technology (Austria). Since October 2007 he has joined in 2007 the Department of Telecommunications, SUPELEC, as an Assistant Professor in network information theory. His research interests include multi-terminal information theory, Shannon theory, cooperative communications, physical-layer security and coding theory for wireless applications.

Shlomo Shamai (Shitz) (S'80–M'82–SM'89–F'94) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1975, 1981 and 1986 respectively.

During 1975–1985, he was with the Communications Research Labs, in the capacity of a Senior Research Engineer. Since 1986 he is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is now a Technion Distinguished Professor, and holds theWilliam Fondiller Chair of Telecommunications. His research interests encompasses a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is an IEEE Fellow, a member of the Israeli Academy of Sciences and Humanities and a Foreign Associate of the US National Academy of Engineering. He is the recipient of the 2011 Claude E. Shannon Award.

Dr. Shamai (Shitz) is an IEEE Fellow, and the recipient of the 2011 Claude E. Shannon Award. He is the recipient of the 1999 van der Pol Gold Medal of the Union Radio Scientifique Internationale (URSI), and a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 joint IT/COM societies paper award, the 2007 IEEE Information Theory Society Paper Award, the 2009 European Commission FP7, Network of Excellence inWireless COMmunications (NEWCOM++) Best Paper Award, and the 2010 Thomson Reuters Award for International Excellence in Scientific Research. He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY, and has also served twice on the Board of Governors of the Information Theory Society. He is a member of the Executive Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY.