Cooperative Binning for Semideterministic Channels

Ritesh Kolte, Ayfer Özgür, Haim Permuter
rkolte@stanford.edu, aozgur@stanford.edu, haimp@bgu.ac.il

Abstract

The capacity regions of semideterministic multiuser channels, such as the semideterministic relay channel and the multiple access channel with partially cribbing encoders, have been characterized using the idea of partial-decode-forward. However, the requirement to explicitly decode part of the message at intermediate nodes can be restrictive in some settings; for example, when nodes have different side information regarding the state of the channel. In this paper, we generalize this scheme to cooperative-bin-forward by building on the observation that explicit recovering of part of the message is not needed to induce cooperation. Instead, encoders can bin their received signals and cooperatively forward the bin index to the decoder. The main advantage of this new scheme is illustrated by considering state-dependent extensions of the aforementioned semideterministic setups. While partial-decode-forward is not applicable in these new setups, cooperative-bin-forward continues to achieve capacity.

Index Terms

Cooperative-bin-forward, Cribbing, Multiple-access channel, Relay channel, State, Semideterministic

I. INTRODUCTION

The capacity region of the semideterministic relay channel, depicted in Figure 1, is characterized in [1] using the partial-decode-forward scheme. In this scheme, the source splits its message into two parts and encodes them using superposition coding. The relay decodes one part of the message, and maps this to a codeword to be transmitted in the next block. The codebooks at the source are generated conditioned on the relay’s transmission, which results in coherent transmissions from the source and the relay.

Consider now the extension of this model depicted in Figure 2, which corresponds to a state-dependent semideterministic relay channel where the state information is causally available only at the source and the destination. This model captures the natural cellular downlink scenario, in which training enables the source and the destination to learn the channel gain between them (state = channel gain), while a relay could be potentially available to assist the communication, e.g. a wifi access point. In this scenario, it is typically unrealistic to assume that the relay is also able to obtain timely information about the channel state between the source and the destination. In this case, requiring the relay to still decode part of the source message, without any state information, is unduly restrictive and to our knowledge the capacity remains unknown to date.

The main contribution of this paper is to develop a new scheme which we call cooperative-bin-forward. This new scheme does not require the relay to decode part of the message; instead, the relay simply bins its received signal and maps the bin-index to a codeword to be transmitted in the next block. As in partial-decode-forward, the codebooks at the source are generated conditioned on the relay’s transmission, resulting in coherent cooperation. This cooperative aspect of the scheme distinguishes it from bin-forward (a.k.a. hash-forward) that has been considered previously for primitive relay channels in [2]. For the vanilla semideterministic relay channel in Figure 1, cooperative-bin-forward recovers the capacity achieved by partial-decode-forward. However, while partial-decode-forward is not applicable for the state-dependent semideterministic relay channel in Figure 2, we show that cooperative-bin-forward continues to achieve capacity.
We next consider another setup where partial-decode-forward is known to be capacity achieving, the multiple-access channel (MAC) with strictly causal partial cribbing encoders, depicted in Figure 3. The MAC with cribbing has been introduced by Willems in [3] and its generalization to partial cribbing has been studied in [4]. Compared to the canonical MAC, transmitters here can overhear each other’s transmissions while simultaneously transmitting their own data. This possibility, increasingly enabled today by the development of full-duplex radios, is especially appealing since such overheard information can be exploited to induce cooperation among the transmitters by exploiting the natural broadcast nature of the wireless medium without requiring any dedicated resources. Partial cribbing refers to the assumption that the overheard signal is some deterministic function of the signal transmitted by the other transmitter, which allows to capture the signal degradation in the cribbing link via a simple model.\footnote{Earlier work \cite{4} has observed that even if a very coarsely quantized signal is overheard, it can still be sufficient to achieve rates that are close to the best rates achievable with the unrealistically perfect cribbing (overhearing via a noiseless link). Thus, we can in fact manually perform coarse quantization of the overheard signals to simplify operations without a significant loss in performance, while the coarseness simultaneously lends justification to modeling the output of the noisy overhearing channel as a deterministic function of the input.}

The MAC with partial cribbing can be regarded as a generalization of the semideterministic relay channel in the sense that when one of the transmitters does not have a message and does not have an outgoing cribbing link, the former reduces to an instance of the latter.

As in the semideterministic relay channel, partial-decode-forward achieves the capacity region of the MAC with partial cribbing \cite{4}. Here, we consider a natural extension of this setup with states, depicted in Figure 2.\footnote{This setup is more general than the setup in Figure 2 in the sense that there are two messages and two states, however it is also special in the sense that the partial cribbing links are of the form $z_1(X_{1i}, S_{1i})$ and $z_2(X_{2i}, S_{2i})$, instead of $z_1(X_{1i}, X_{2i}, S_{1i})$ and $z_2(X_{1i}, X_{2i}, S_{2i})$.} The most prevalent example of a situation captured by this model is cellular uplink. In cellular communication, training enables a transmitter to learn the channel gain between itself and the receiver, but assuming knowledge of the channel between any other transmitter and the receiver is unrealistic. Hence, the model in Figure 2 includes a state composed of two components, each known causally only to the corresponding transmitter. These two components are not necessarily independent. Since encoders do not share common state information, partial-decode-forward becomes too restrictive for this setting. Instead, an achievability scheme based on cooperative-bin-forward provides the capacity region.

Finally, motivated by the relay-without-delay channel considered in [5], we consider “without-delay” variations of the two state-dependent setups described above, that are depicted in Figure 5 and Figure 6 respectively. In these setups, the strict causality of one of the links is replaced by causality. In the former, which is the state-dependent semideterministic relay-without-delay channel, the transmission of the relay is allowed to depend on its past and current received signal. The capacity region for this setup without state is characterized in [5], using partial-decode-forward combined with instantaneous relaying (a.k.a. codetrees or Shannon strategies). The latter is a state-dependent multiple access channel with one strictly causal and one causal partial cribbing link. The capacity region for this setup without states is characterized in [5], again using partial-decode-forward combined with instantaneous relaying. We show that cooperative-bin-forward combined with instantaneous relaying achieves the capacity regions of these setups too, while partial-decode-forward suffers from the same shortcoming encountered for the previous two setups.

---

*Fig. 2: State-dependent Semideterministic Relay Channel with Causal State Information at Source and Destination*

*Fig. 3: Multiple Access Channel with Strictly-Causal Partial-Cribbing Encoders*
Related Work

We describe here some multiuser setups considered in literature that involve state-dependent channels and/or some form of cooperation, and how they relate to the setups considered in this paper.

Various cases of state-dependent relay channels have been considered in [6]–[12]. The achievability schemes in these works combine well-known block-Markov relaying ideas such as partial-decode-forward and compress-forward with Shannon strategies or (Gelfand-Pinsker) multicoding. A class of state-dependent orthogonal relay channels with state information only works combine well-known block-Markov relaying ideas such as partial-decode-forward and compress-forward with Shannon cooperation, and how they relate to the setups considered in this paper.

There has been interesting recent work on state-dependent multiple access channels where the state is only known to the There has been interesting recent work on state-dependent multiple access channels where the state is only known to the

\[ (S_1, S_2) \sim p_{S_1, S_2} \]

\[ \begin{align*}
 & M_1 \\
 \quad \text{Encoder 1} \\
 & x_1(M_1, S_1, Z_{i-1}^1) \\
 & z_1(X_1, S_1) \\
 & z_2(X_2, S_2) \\
 & p_{Y|X_1, X_2, S_1, S_2} \\
 & Y_1 \\
 \quad \text{Decoder} \\
 & \hat{m}_1(Y_n, S_n^1, S_n^2) \\
 & (S_1, S_2) \sim p_{S_1, S_2} \\
 & M_2 \\
 \quad \text{Encoder 2} \\
 & x_2(M_2, S_2, Z_{i-1}^1) \\
 & p_{Y|X_1, X_2, S_1, S_2} \\
 & Y_1 \\
 \quad \text{Decoder} \\
 & \hat{m}_2(Y_n, S_n^1, S_n^2) \\
 & (S_1, S_2) \sim p_{S_1, S_2} \\
 & \end{align*} \]

\[ \begin{align*}
 & M_1 \\
 \quad \text{Encoder 1} \\
 & x_1(M_1, S_1) \\
 & z_1(X_1, S_1) \\
 & z_2(X_2, S_2) \\
 & p_{Y|X_1, X_2, S_1, S_2} \\
 & Y_1 \\
 \quad \text{Decoder} \\
 & \hat{m}_1(Y_n, S_n^1, S_n^2) \\
 & \end{align*} \]

\[ \text{Related Work} \]

Given that the state is only known to the

\[ (S_1, S_2) \sim p_{S_1, S_2} \]

\[ \begin{align*}
 & M_1 \\
 \quad \text{Encoder 1} \\
 & x_1(M_1, S_1, Z_{i-1}^1) \\
 & z_1(X_1, S_1) \\
 & z_2(X_2, S_2) \\
 & p_{Y|X_1, X_2, S_1, S_2} \\
 & Y_1 \\
 \quad \text{Decoder} \\
 & \hat{m}_1(Y_n, S_n^1, S_n^2) \\
 & (S_1, S_2) \sim p_{S_1, S_2} \\
 & M_2 \\
 \quad \text{Encoder 2} \\
 & x_2(M_2, S_2, Z_{i-1}^1) \\
 & p_{Y|X_1, X_2, S_1, S_2} \\
 & Y_1 \\
 \quad \text{Decoder} \\
 & \hat{m}_2(Y_n, S_n^1, S_n^2) \\
 & (S_1, S_2) \sim p_{S_1, S_2} \\
 & \end{align*} \]
(additive interference composed of two components each known noncausally to one and only one encoder in Gaussian noise). For this case, a straightforward extension of Gelfand-Pinsker coding turns out to be highly suboptimal. Instead, a structured form of Gelfand-Pinsker coding using lattices is useful for achieving high rates, since it ensures that the overall interference at the destination concentrates on a small set. Finally, the case of common causal state information at the encoders was studied in [18], which provided an inner bound using Shannon strategies. Recall that the multiple access channels studied in this paper, Figure 4 and Figure 6, assume also that the destination has full state information, while the encoders cooperate via partial cribbing, thus the main challenge in this paper is to optimally establish cooperation via cribbing among encoders that have disparate state information, rather than handling the lack of state information at the destination.

A few works have considered state-dependent multiple access channels where the state information is available at the encoders as well as the decoder, also with no cooperation between the encoders. It was shown in [19]–[21] that optimal rates can be achieved by effectively treating the state components as time-sharing. These setups can be obtained as a special case of the setup in Figure 4 by setting the partial cribbing links to zero, though the optimal schemes presented previously for these special cases do not provide the necessary insights for establishing cooperation among the encoders. In particular, the causality of cribbing requires the scheme to be block-Markov, and the encoding operation across blocks needs to be such that the two state components can be effectively treated as time-sharing, on top of establishing cooperation inspite of the disparateness of the state components.

Cooperation in multiple access channels was studied by Willems in [22] and [3], wherein he introduced the notions of conferencing (orthogonal links) and cribbing respectively. In conferencing, dedicated orthogonal links are introduced for cooperation. Cribbing, in contrast, does not assume dedicated resources for cooperation. For example, cribbing can be thought of as exploiting the natural broadcasting nature of the wireless medium for cooperation. Decode-forward based schemes were proved to be optimal for multiple access channels with cribbing. However, the cribbing in [3] was assumed to be perfect (noiseless). To account for the fact that perfect cribbing is unrealistic, the notion of partial cribbing, as described in the introduction, was studied in [4], and a partial-decode-forward based scheme was shown to be optimal. As mentioned in the introduction, the fact that a part of the message needs to be explicitly decoded in partial-decode-forward, renders its straightforward extension inapplicable for our purpose.

There has also been interest in studying multiple access channels that include states together with some form of cooperation between the encoders, under various assumptions on the state information availability and the form of cooperation [23]–[27]. All these works assume that whenever the state information is available, it is available noncausally. The capacity regions for noncausal state information only at one encoder were provided in [23] and [29]. The former additionally needed to assume that the informed encoder also knows the other message, while the latter assumed instead that there is a strictly causal or causal perfect cribbing link from the uninformed encoder to the informed encoder. The achievability schemes in both works used Gelfand-Pinsker multicoding at the informed encoder conditioned on the additional information received in the form of message cognition or cribbing. The capacity region for the case of conferencing encoders when noncausal state information is available at all nodes, including the destination, was provided in [25] using the idea of double-binning. Achievable rate regions were derived in [26] and [27], where the former considered perfect cribbing among encoders with noncausal state information only at the encoders, while the latter replaced perfect cribbing by noisy cribbing.

**Organization**

The following section describes the models and notation. Section III contains the formal statements of the main results described in the introduction. A toy example is considered in Section IV for the purpose of explicitly illustrating the advantage of cooperative-bin-forward over partial-decode-forward. The following sections contain the proofs of the main results. We conclude by describing some open problems in section X.

II. System Models

As standard, capital letters denote random variables, small letters denote realizations, and calligraphic letters denote the alphabet of the corresponding random variable. The notation $\mathcal{T}_{\epsilon}^{(n)}$ stands for the $\epsilon$-strongly typical set of sequences for the random variables in context.

A. State-Dependent Semideterministic Relay Channels

The state-dependent semideterministic relay channel is depicted in Figure 2 and described by the pmf $p_{S}(s)\mid y(x, x_r, s)$ and $Z = z(X, X_r, S)$. The encoder and decoder have causal state information. So a $(n, 2^{nR}, \epsilon)$ code for the above channel consists of the source encoding, relay encoding and decoding functions:

$$x_i : [1 : 2^{nR}] \times S_i \rightarrow X, \quad 1 \leq i \leq n,$$

$$x_{r,i} : 2^{i-1} \rightarrow X_r, \quad 1 \leq i \leq n,$$

$$\hat{m} : 2^n S^n \rightarrow [1 : 2^{nR}]$$
such that
$$\Pr \{ \hat{m}(Y^n, S^n) \neq M \} \leq \epsilon,$$
where \(M \in [1 : 2^{nR}]\) denotes the transmitted message. A rate \(R\) is said to be achievable if for every \(\epsilon > 0\), there exists a \((n, 2^{nR}, \epsilon)\) code for sufficiently large \(n\). The capacity is defined to be the supremum of achievable rates.

The state-dependent semideterministic relay-without-delay channel is depicted in Figure 5 and described by the pmf \(p_{s_1,s_2}(s_1,s_2)p_{y|x,x_s}(y|x,x_s)\) and \(Z = z(X,S)\). The difference from the previous setup is that the relay encoding function is now allowed to depend also on \(Z\):
$$x_{r,i} : Z^i \rightarrow X_r, \quad 1 \leq i \leq n.$$  
Note that here we need to restrict \(Z\) to be \(z(X,S)\), instead of \(z(X,X_r,S)\).

### B. State-Dependent Multiple-Access Channels

The state-dependent multiple access channel with strictly-causal partial-cribbing encoders is depicted in Figure 4 and described by the pmf \(p_{s_1,s_2}(s_1,s_2)p_{y|x,x_s}(y|x,x_s)\) and \(Z_1 = z_1(X_1,S_1)\) and \(Z_2 = z_2(X_2,S_2)\). The encoders have causal knowledge of the corresponding state components, but no knowledge of the other state component. The decoder is assumed to know both the state components. A \((n,2^{nR_1},2^{nR_2},\epsilon)\) code for the above channel consists of the encoding and decoding functions:

\[
\begin{align*}
x_{1,i} & : [1 : 2^{nR_1}] \times S_1^n \times Z_2^{i-1} \rightarrow X_1, \quad 1 \leq i \leq n, \\
x_{2,i} & : [1 : 2^{nR_2}] \times S_2^n \times Z_1^{i-1} \rightarrow X_2, \quad 1 \leq i \leq n, \\
\hat{m}_1 & : Y^n \times S_1^n \times S_2^n \rightarrow [1 : 2^{nR_1}], \\
\hat{m}_2 & : Y^n \times S_1^n \times S_2^n \rightarrow [1 : 2^{nR_2}],
\end{align*}
\]

such that
$$\Pr \{ (\hat{m}_1(Y^n, S_1^n, S_2^n), \hat{m}_2(Y^n, S_1^n, S_2^n)) \neq (M_1, M_2) \} \leq \epsilon,$$
where \(M_1 \in [1 : 2^{nR_1}]\) and \(M_2 \in [1 : 2^{nR_2}]\) denote the transmitted messages. A rate pair \((R_1, R_2)\) is said to be achievable if for every \(\epsilon > 0\), there exists a \((n,2^{nR_1},2^{nR_2},\epsilon)\) code for sufficiently large \(n\). The capacity region is defined to be the closure of the achievable rate region.

The “without-delay” variation of this setup, also referred to as causal cribbing, is depicted in Figure 6 where one of the partial cribbing links is changed from strictly causal to causal. So, the only difference from the previous setting is that \(x_{2i}(M_2, S_2^i, Z_1^{i-1})\) is replaced by \(x_{2i}(M_2, S_2^i, Z_1^i)\).

### III. Main Results

In this section, we describe the capacity regions for all the setups described in the previous section. The proofs are presented in subsequent sections. The achievability parts of all the theorems are accomplished by building on the idea of cooperative-bin-forward.

The first result is taken from [1]. We restate it here and provide a proof of the achievability in Section V using cooperative-bin-forward. Due to the simplicity of the setup, it serves well to bring out the main idea of the new scheme, before we describe results for the more complicated setups.

**Theorem 1.** The capacity of the semideterministic relay channel, shown in Figure 1, is given by
$$C = \max_{p_{X,X_r}(x,x_r)} \min \left\{ I(X,X_r;Y), H(Z|X_r) + I(X;Y|X_r,Z) \right\}.$$  
(1)

The next result provides an expression for the capacity of the state-dependent semideterministic relay channel.

**Theorem 2.** The capacity of the state-dependent semideterministic relay channel, shown in Figure 2, is given by
$$C = \max_{p_{X,X_r}(x,x_r)p_{X|X_r,s}(x|x_r,s)} \min \left\{ I(X,X_r;Y|S), H(Z|S,X_r) + I(X;Y|S,X_r,Z) \right\}.$$  
(2)

One difference between the capacity expressions of Theorem 1 and Theorem 2 is that the mutual information and entropy terms involve a conditioning on \(S\). Such an expression would also characterize the capacity if the relay is provided with the state information, and it would be achievable by performing partial-decode-forward while treating the state as a time-sharing sequence. It is quite interesting then that the capacity expression remains the same even when the relay does not have state information. However, the limitation is reflected in the fact that the choice of pmf is restricted to be \(p_{X,X_r}(x_r)p_{X|X_r,S}(x|x_r,s)\), instead of \(p_{X,X_r,S}(x,x_r,s)\). So, the cost of not having state information at the relay is reflected entirely in the limited choice of pmf.
The following theorem states the capacity of the without-delay variation of the above case. The expression involves an auxiliary random variable, which allows the relay to perform instantaneous relaying on top of the binning. This can achieve maximal source-relay cooperation, as conveyed by the following theorem.

**Theorem 3.** The capacity of the state-dependent semideterministic relay-without-delay channel, shown in Figure 5, is given by

\[
C = \max_{p_U(u)p_{X_1}(x_1|u,s)p_{X_2}(x_2|u,s)} \min \left\{ I(U;X_1,Y|S), H(Z|U,S) + I(X_1;Y|U,Z,S) \right\},
\]

where \(|U| \leq |S|(|X_1|+|X_2|+1) + 2\).

The capacity region for the setup of Theorem 3 in the absence of states is characterized in [5, Proposition 7]. Setting \(S = \emptyset\) in Theorem 3 recovers this result. Note that the objective in (3) is the same as that in (2) with \(X_r\) being replaced by \(U\). However, the optimization in (3) is over a different domain since the dependence of \(X_r\) on \(Z\) can now be chosen and is not specified by the channel.

The next two theorems describe the capacity regions for the two multiple-access setups.

**Theorem 4.** The capacity region of the state-dependent multiple-access channel with partially cribbing encoders, shown in Figure 4, is given by the set of rate pairs \((R_1, R_2)\) satisfying

\[
\begin{align*}
R_1 &\leq I(X_1;Y|U, X_2, Z_1, S_1, S_2) + H(Z_1|U, S_1), \\
R_2 &\leq I(X_2;Y|U, X_1, Z_2, S_1, S_2) + H(Z_2|U, S_2), \\
R_1 + R_2 &\leq I(X_1, X_2;Y|U, Z_1, Z_2, S_1, S_2) + H(Z_1, Z_2|U, S_1, S_2), \\
R_1 + R_2 &\leq I(X_1, X_2;Y|S_1, S_2),
\end{align*}
\]

for pmf of the form

\[
p_{U|S_1,S_2}(u|x_1,s_1)p_{X_1|U,S_1,S_2}(x_1|u,s_1)p_{X_2|U,S_2}(x_2|u,s_2),
\]

with \(Z_1 = z_1(X_1, S_1)\) and \(Z_2 = z_2(X_2, S_2)\), with \(|U| \leq \min\{|S_1||S_2|(|X_1|+|X_2|+1) + 4, |S_1||S_2|(|X_1||X_2|+1) + 3\}\).

*Remark:* It can be shown that the set described in the above theorem is convex, so there is no need to introduce an additional auxiliary random variable for time-sharing.

As described earlier, the special case of no cribbing (obtained by setting \(Z_1 = 0\) and \(Z_2 = 0\)) has been considered in [19]–[21]. For this case, the last inequality becomes redundant and setting the auxiliary random variable \(U\) to be the time-sharing random variable in the statement of the above theorem is optimal. The resulting region recovers the results in the aforementioned papers. For the other extreme of constant states, i.e. \(S_1 = 0\) and \(S_2 = 0\), the capacity region in the above theorem recovers the result for strictly causal partial cribbing from [4].

**Theorem 5.** The capacity region of the state-dependent multiple-access channel with partially cribbing encoders in the presence of a causal cribbing link, shown in Figure 6, is given by the set of rate pairs \((R_1, R_2)\) satisfying

\[
\begin{align*}
R_1 &\leq I(X_1;Y|U, X_2, Z_1, S_1, S_2) + H(Z_1|U, S_1), \\
R_2 &\leq I(X_2;Y|U, X_1, Z_2, S_1, S_2) + H(Z_2|U, S_2, Z_1), \\
R_1 + R_2 &\leq I(X_1, X_2;Y|U, Z_1, Z_2, S_1, S_2) + H(Z_1, Z_2|U, S_1, S_2), \\
R_1 + R_2 &\leq I(X_1, X_2;Y|S_1, S_2),
\end{align*}
\]

for pmf of the form

\[
p_{U|S_1,S_2}(u|x_1,s_1)p_{X_1|U,S_1,S_2,Z_1}(x_1|u,s_1)p_{X_2|U,S_2}(x_2|u,s_2,z_1),
\]

with \(Z_1 = z_1(X_1, S_1)\) and \(Z_2 = z_2(X_2, S_2)\), with \(|U| \leq \min\{|S_1||S_2|(|Y|+1) + 4, |S_1||S_2|(|X_1||X_2|+1) + 3\}\).

Note that if \(p_{X_2|U,S_2}(x_2|u,s_2)\) in Theorem 5 is replaced by \(p_{X_2|U,S_1}(x_2|u,s_2)\), then the region becomes identical to that in Theorem 4. Setting \(S_1\) and \(S_2\) to be constant retrieves the result for causal partial cribbing from [4].

### IV. Illustrative Example

Consider the following special case of Figure 2.

Let the state \(S\) be the ternary random variable

\[
p_S(s) = \begin{cases} 
  \frac{p}{2}, & \text{if } s = 0, \\
  \frac{p}{2}, & \text{if } s = 1, \\
  1-p, & \text{if } s = 2.
\end{cases}
\]
where \( p < 1/2 \). The other variables are all binary. The channel \( z(X,S) \) is the memory with stuck-at faults channel considered in [28], while the channel \( p_{Y|X,X,S} \) is specialized to be a noiseless channel from \( X_r \) to \( Y \). Formally,

\[
z(X,S) = \begin{cases} 
0, & \text{if } S = 0, \\
1, & \text{if } S = 1, \\
x, & \text{if } S = 2, 
\end{cases}
\]

\[
Y = X_r.
\]

Recall that the source and the destination know the state information causally while the relay has no state information.

If the relay is required to decode the message, motivated by the optimality of decode-forward in the case of a line network with no state, the achievable rate is limited to be no more than the capacity of the memory with stuck-at faults channel when the state is known causally only to the source, which is \( 1 - H_2 \left( \frac{z}{2} \right) \). We point out that this cannot be improved by using partial-decode-forward, because the absence of a direct link between the source and destination means that any part of the message that is not decoded by the relay cannot be communicated to the destination in any manner. However, a higher rate can be achieved if the relay simply forwards its received signal, resulting in an effective channel between the source and the destination that is the memory with stuck-at faults channel with state known causally both to the source and the destination. The capacity of this channel is \( 1 - p \), which is achieved by multiplexing at the source and demultiplexing at the destination according to the observed state. Thus, a rate \( 1 - p \) which is higher than \( 1 - H_2 \left( \frac{z}{2} \right) \) can be achieved.

What if the channel from the relay to destination is not a noiseless bit-pipe, but a general noisy channel with capacity at least \( 1 - p \)? The rate \( 1 - p \) can still be achieved if the operation at the relay is changed from simply forwarding to randomly binning its received signal into \( \approx 2^{n(1-p)} \) bins and forwarding a codeword corresponding to the chosen bin. To recover the message, the destination can first decode the bin-index. Since the destination has state information, it can reconstruct the message that results in the received signal at the relay falling in the correct bin.

The above example serves to illustrate the limitation of partial-decode-forward when nodes have different side-information. This example did not require cooperative transmissions from the source and the relay, because the source transmission did not directly affect the received signal at the destination. When there is also a direct link between the source and the destination, as in partial-decode-forward, the codebooks at the source are generated conditioned on the relay’s transmission, resulting in coherent cooperation. The scheme is formally described next.

**V. PROOF OF THEOREM**

We demonstrate in this section the achievable capacity for the semideterministic relay channel using the new scheme cooperative-bin-forward. As described in the introduction, this scheme does not require the relay to decode part of the message. Instead, the relay simply bins its received signal and maps the bin-index to a codeword to be transmitted in the next block. As in partial-decode-forward, the codebooks at the source are generated conditioned on the relay’s transmission, resulting in coherent cooperation. The scheme is formally described next.

**Proof:**

Fix a pmf \( p_{X,X_r}(x,x_r) \) and \( \epsilon > 0 \). Split \( R \) as \( R' + R'' \), with the message \( M \) denoted accordingly as \( (M',M'') \). Divide the total communication time into \( B \) blocks, each of length \( n \).

**Codebook Generation:**

For each block \( b \in [1 : B] \), a codebook is generated independently of the other blocks as follows.

- **Cooperation codewords**
  Generate \( 2^{nR} \) codewords \( x^n_{rb}(l_{b-1}) \) i.i.d. according to \( p_{X_r} \), where \( l_{b-1} \in [1 : 2^{nR}] \).

- **Cribbed codewords**
  For each \( l_{b-1} \), generate \( 2^{nR} \) codewords \( z^n_{rb}(m'_b|l_{b-1}) \) according to \( \prod_{i=1}^n p_{Z|X_r} \left( z_{rb}|x_{rb}(l_{b-1}) \right) \), where \( m'_b \in [1 : 2^{nR}] \).

- **Transmission codewords**
  For each \( l_{b-1} \) and each \( m'_b \), generate \( 2^{nR''} \) codewords \( x^{'n}_{rb}(m''_b|l_{b-1},m'_b) \), where \( m''_b \in [1 : 2^{nR''}] \) according to \( \prod_{i=1}^n p_{X|X_r,z} \left( x_{rb}|z_{rb}(l_{b-1}),z_{rb}(m'_b|l_{b-1}) \right) \).

- **Binning**
  Partition the set of all \( Z^n \) into \( 2^{nR} \) bins, by choosing a bin for each \( z^n \) independently and uniformly at random. Denote the index of the chosen bin for \( z^n \) by \( \text{bin}_b(z^n) \).

\( ^3 \)Given the analogy of the source-to-relay link in the relay channel with the cribbing link in the multiple-access channel with cribbing encoders, we call the \( z^n_{rb} \) codewords as cribbed codewords.
Encoding:
Fix \( l_0 = 1 \) and \((m'_B, m''_B) = (1, 1)\). Since the message in the last block is fixed, the effective rate of communication is \( \frac{B-1}{B} R \), which can be made as close as desired to \( R \) by choosing a sufficiently large \( B \).

In block \( l_b \), \( l_{b-1} \) is known to the source encoder. To communicate message \( m_b = (m'_b, m''_b) \), it transmits \( x_b^l(m''_b|l_{b-1}, m'_b) \). The relay transmits \( x_b^n(l_{b-1}) \). Due to the deterministic link from source to relay and the codebook construction, the received signal at the relay in block \( b \) is the codeword \( z^n_b(m'_b|l_{b-1}) \). The source and the relay set \( b_b \) to be the index of the bin containing \( z^n_b(m'_b|l_{b-1}) \).

Decoding:
The decoder performs backward decoding, starting from block \( B \) and moving towards block 1, performing the following two steps for each block \( b \):

1. Assuming that \( b_b \) is known from previous operations, the decoder, for each \( l_{b-1} \in [1 : 2^nR] \), finds the unique \( m'_b \) such that
   \[
   \text{bin}_b(z^n(m'_b|l_{b-1})) = l_b.
   \]
   Whenever a unique \( m'_b \) cannot be found for some \( l_{b-1} \), the decoder chooses any \( m'_b \) arbitrarily. So after this operation, the decoder has chosen one \( m'_b \) for each \( l_{b-1} \), given its knowledge of \( l_b \). We will signify this explicitly by denoting the chosen message as \( \hat{m}'_b(l_{b-1}, b_b) \).

2. Now the decoder looks for the unique \((\hat{l}_{b-1}, \hat{m}''_b)\) such that
   \[
   \left( x^n_{\hat{l}_{b-1}}, z^n_{\hat{m}'_b(l_{b-1}, l_b)}|l_{b-1} \right), x^n_{\hat{m}''_b(l_{b-1}, \hat{l}_{b-1})}, y^n_b \right) \in \mathcal{T}_c^{(n)}.
   \]
   Note that the first step does not depend on the received signal in block \( b \) at the destination. However, it depends on the received signal in block \( b+1 \), due to the involvement of \( b_b \).

Probability of Error:
The following error analysis reveals that in order to achieve the highest rate, the scheme will set \( R' \approx \hat{R} \approx H(Z|X_r) \). It is easy to see that when \( R' \approx \hat{R} \approx H(Z|X_r) \), given its knowledge of \( l_{b-1} \), the relay can indeed recover \( m_b \), even though it is not required to do so in this new scheme. In other words, for a given \( l_{b-1} \), since \( R' \approx \hat{R} \), each message \( m'_b \) is mapped to a different bin, and therefore cooperatively communicating the bin index is indeed equivalent to cooperatively communicating the partial message \( m'_b \). Thus, cooperative-bin-forward for this basic setup is indeed equivalent to partial-decode-forward. We will see however in the next section that when we have states, even though we still set \( R' \approx \hat{R} \), the relay will not be able to decode any part of the message so the binning aspect of the scheme will be instrumental.

By symmetry, we can assume without loss of generality that the true messages and bin-indices corresponding to the current block are all 1, i.e.

\[
(L_{b-1}, M'_b, M''_b) = (1, 1, 1).
\]

We bound the probability of decoding error in block \( b \) conditioned on successful decoding for blocks \( \{B, B-1, \ldots, b+1\} \), averaged over the randomness in the messages and codebook generation. In particular, successful decoding in block \( b+1 \) means that \( L_b \) has been decoded successfully, where we remind ourselves that

\[
L_b = \text{Bin}_b(Z^n_1(1|1)).
\]

An error occurs in block \( b \) only if any of the following events occur:

(a) \( \hat{M}'_b(1, L_b) \neq 1 \).
(b) \( (\hat{L}_{b-1}, \hat{M}''_b) \neq (1, 1) \) given \( \hat{M}'_b(1, L_b) = 1 \).

We analyze the two events in the following. One can notice from the above partitioning of the error events that we are ensuring that \( \hat{M}'_b(l_{b-1}, L_b) \) is equal to 1 only for \( l_{b-1} = L_{b-1} = 1 \), and not worrying about what \( \hat{M}'_b(l_{b-1}, L_b) \) is for any other value of \( l_{b-1} \). However, it is still important to fix at most one \( m'_b \) pair for each \( l_{b-1} \), even if it is arbitrary for all \( l_{b-1} \neq 1 \), which is what the first step of the decoding does\(^4\). This allows us to restrict our attention to at most \( 2^{n(R+R')} \) options while analyzing the probability of decoding \( L_{b-1} \) incorrectly during the second decoding step, instead of \( 2^{n(R+R+R')} \).

\(^4\)Of course, one could also discard any \( l_{b-1} \) for which a unique \( \hat{M}'_b(l_{b-1}, L_b) \) cannot be identified; we stick to making one arbitrary choice in such cases only because it makes the exposition simpler.
Event (a): \( \bar{M}'_b(1, L_b) \neq 1 \):

We have

\[
\Pr\left( \bar{M}'_b(1, L_b) \neq 1 \right) = \Pr(\text{Bin}_b(Z_b^0(m'_b|1)) = L_b \text{ for some } m'_b > 1)
\]

\[
\overset{(i)}{=} \Pr(\text{Bin}_b(Z_b^0(m'_b|1)) = \text{Bin}_b(Z_b^0(1|1)) \text{ for some } m'_b > 1)
\]

\[
\leq \sum_{m'_b>1} \Pr(\text{Bin}_b(Z_b^0(m'_b|1)) = \text{Bin}_b(Z_b^0(1|1)))
\]

\[
= \sum_{m'_b>1} \Pr(\text{Bin}_b(Z_b^0(m'_b|1)) = \text{Bin}_b(Z_b^0(1|1)), Z_b^0(m'_b|1) = Z_b^0(1|1))
\]

\[
+ \sum_{m'_b>1} \Pr(\text{Bin}_b(Z_b^0(m'_b|1)) = \text{Bin}_b(Z_b^0(1|1)), Z_b^0(m'_b|1) \neq Z_b^0(1|1))
\]

\[
\overset{(ii)}{\leq} 2^n R', 2^{-(H(Z|X_r)-\delta(\epsilon))} + 2^n R', 2^{-n\bar{R}},
\]

where

- (i) follows since \( L_b = \text{Bin}_b(Z_b^0(1|1)) \),
- the first term in (ii) is obtained because for \( m'_b \neq 1 \), the codewords \( Z_b^0(m'_b|1) \) and \( Z_b^0(1|1) \) are generated independently according to \( \prod_{i=1}^n p_{Z|X}(\cdot|X_{rb}(l_b-1)) \), and the second term in (ii) arises because the binning is performed uniformly at random and independently for each sequence,

and we use \( \delta(\epsilon) \) to denote any function of \( \epsilon \) for which \( \delta(\epsilon) \to 0 \) as \( \epsilon \to 0 \). Hence, we get that

\[
\Pr\left( \bar{M}'_b(1, L_b) \neq 1 \right) \to 0 \text{ as } n \to \infty,
\]

if the following two constraints are satisfied:

\[
R' < \bar{R},
\]

\[
R' < H(Z|X_r) - \delta(\epsilon).
\]

Event (b): \( (\hat{L}_{b-1}, \hat{M}'_b) \neq (1, 1) \) given \( \hat{M}'_b(1, L_b) = 1 \):

The probability of this event is upper bounded by

\[
\Pr\left( \text{Condition (6) is not satisfied by } (l_{b-1}, m''_b) = (1, 1) \mid \hat{M}'_b(1, L_b) = 1 \right)
\]

\[
+ \Pr\left( \text{Condition (6) is satisfied for some } (l_{b-1}, m''_b) \neq (1, 1) \mid \hat{M}'_b(1, L_b) = 1 \right).
\]

The first term goes to zero as \( n \to \infty \) by the law of large numbers. The second term can be analyzed by standard applications of the packing lemma [28] as follows:

\[
\Pr\left( \text{Condition (6) is satisfied for some } (l_{b-1}, m''_b) \neq (1, 1) \mid \hat{M}'_b(1, L_b) = 1 \right)
\]

\[
\leq \sum_{l_{b-1}=1, m''_b \geq 1} \Pr\left( (X_{rb}^n(1), Z_b^0(1|1), X_b^0(m''_b|1, 1), Y_b^n) \in T^{(n)}_c \right)
\]

\[
+ \sum_{l_{b-1}=1, m''_b \geq 1} \Pr\left( (X_{rb}^n(l_{b-1}), Z_b^0(l_{b-1}, L_b), Y_b^n) \in T^{(n)}_c \right)
\]

\[
\leq 2^n R'' 2^{-n(I(X;Y|X_r,Z)-\delta(\epsilon))} + 2^n (R + R'') 2^{-n(I(X;X_r;Y)-\delta(\epsilon))},
\]

which follows by applying the packing lemma. Thus, we get that

\[
\Pr\left( (\hat{L}_{b-1}, \hat{M}'_b) \neq (1, 1) \mid \hat{M}'_b(1, L_b) = 1 \right) \to 0, \text{ as } n \to \infty,
\]

if

\[
R'' < I(X;Y|X_r,Z) - \delta(\epsilon),
\]

\[
\tilde{R} + R'' < I(X;X_r;Y) - \delta(\epsilon).
\]
Performing Fourier-Motzkin elimination of \( \tilde{R} \), \( R' \) and \( R'' \) using the rate constraints and also \( R = R' + R'' \), letting \( n \to \infty \), \( B \to \infty \) and \( \epsilon \to 0 \), we get that the rates specified in Theorem 1 are indeed achievable by the scheme presented in this section. One can also simplify the Fourier-Motzkin elimination step by setting \( \tilde{R} \) to be \( R' + \delta(\epsilon) \). Since the converse part of this theorem is already known, we do not repeat the arguments here.

VI. PROOF OF THEOREM 2

The achievability part of the theorem is similar to the scheme presented in the previous section. Due to the availability of causal state information at the source encoder and the decoder, the source encoder constructs codebooks for each state symbol and treats the state sequence as a time-sharing sequence (i.e. it performs multiplexing). Note that since the relay does not have state information, it might not be able to decode part of the message. However, it can still perform the bin-forward operation, allowing us to establish coherence between the source and the relay transmissions without sacrificing unnecessarily on the rate. The decoding is also similar to the previous section, except for the extra demultiplexing component. The converse part of the theorem is presented towards the end of this section. Thus, we see that while partial-decode-forward cannot be applied, cooperative-bin-forward allows us to achieve the capacity region of the state-dependent semideterministic relay channel.

Proof:

Fix a pmf \( p_X, (x_r) p_X|X,S(x|x_r,s) \) and \( \epsilon > 0 \). Split \( R \) as \( R' + R'' \), with the message \( M \) denoted accordingly as \( (M', M'') \). Divide the total communication time into \( B \) blocks, each of length \( n \).

\[
\begin{array}{cccccc}
S = 1 & & & & & \\
\cdot & \cdot & \cdot & \cdot & \cdot & m' = 4 \\
\cdot & \cdot & \cdot & \cdot & \cdot & m' = 3 \\
\cdot & \cdot & \cdot & \cdot & \cdot & m' = 2 \\
\cdot & \cdot & \cdot & \cdot & \cdot & m' = 1 \\
S = 0 & & & & & \\
\cdot & \cdot & \cdot & \cdot & \cdot & m' = 4 \\
\cdot & \cdot & \cdot & \cdot & \cdot & m' = 3 \\
\cdot & \cdot & \cdot & \cdot & \cdot & m' = 2 \\
\cdot & \cdot & \cdot & \cdot & \cdot & m' = 1 \\
\end{array}
\]

Fig. 7: The figure depicts the cribbed codewords generated for encoding \( m' \) for a given \( z^n \). Each node corresponds to a \( z \) symbol that is generated independently according to \( p_Z|X,S(x|x_r,l),s \). The red circles show how encoder 1 chooses the codeword if it wants to transmit \( m' = 2 \) and observes \( s^n = (0, 0, 1, 0, 1, 0) \). This construction is not identical but equivalent to that described in [28, Section 7.4.1].

Codebook Generation:

For each block \( b \in [1 : B] \), a codebook is generated independently of the other blocks as follows.

- Cooperation codewords
  Generate \( 2^nR \) codewords \( x^n_{b,1}(l_b-1) \), i.i.d. according to \( p_X, b \), where \( l_b-1 \in [1 : 2^nR] \).

- Cribbed codewords
  For each \( l_b-1 \) and each \( s \in S \), generate a codebook of \( 2^nR \) codewords. The \( i \)th symbol of such a codeword is chosen independently according to \( p_Z|X_S, s(x_r,l),s \). The result of this is that for each \( l_b-1 \), each \( m'_b \in [1 : 2^nR] \) and each \( s^n_b = (s_{b1}, s_{b2}, \ldots, s_{bn}) \), the source encoder can form an effective codeword \( z^n_b(m'_b|l_b-1, s^n_b) \), whose \( i \)th symbol can be causally chosen as the \( i \)th symbol of the \( m'_b \)th codeword from the codebook corresponding to \( l_b-1 \) and \( s_{bi} \). See Figure 7.

- Transmission codewords
  For each \( l_b-1 \), each \( m''_b \in [1 : 2^nR] \) and each \( s \in S \), generate a codebook of \( 2^nR \) codewords. The \( i \)th symbol of such a codeword is generated independently according to \( p_X|X_S, z, s(x_r,l), z_b(m''_b|l_b-1, s, s) \). The result of this construction is that for each \( l_b-1 \), each \( m''_b \in [1 : 2^nR] \), each \( m'_b \in [1 : 2^nR'] \) and each \( s^n_b \), the source encoder can form an effective codeword \( x^n_{b}(m''_b|l_b-1, m'_b, s^n_b) \), whose \( i \)th symbol can be causally chosen as the \( i \)th symbol of the \( m''_b \)th codeword from the codebook corresponding to \( l_b-1 \), \( m'_b \) and \( s_{bi} \).

- Binning
  Partition the set of all \( Z^n \) into \( 2^nR \) bins, by choosing a bin for each \( z^n \) independently and uniformly at random. Denote the index of the chosen bin for \( z^n \) by \( b^n(z^n) \).

Encoding:

Fix \( l_0 = 1 \) and \( (m'_b, m''_b) = (1, 1) \). Since the message in the last block is fixed, the effective rate of communication is \( B^{-1}R \), which can be made as close as desired to \( R \) by choosing a sufficiently large \( B \).

In block \( b \), \( l_{b-1} \) is known to the source encoder. To communicate message \( m_b = (m'_b, m''_b) \), it transmits \( x^n_{b}(m''_b|l_b-1, m'_b, s^n_b) \). The relay transmits \( x^n_{b+1}(l_{b+1}) \). Due to the deterministic link from source to relay and the codebook construction, the received
signal at the relay in block $b$ is the codeword $z^n_b(m'_b | l_{b-1}, s^n_0)$. The source and the relay set $l_b$ to be the index of the bin containing $z^n_b(m'_b | l_{b-1}, s^n_0)$.

From the encoding operation described above, we can see that the label $l_b$ depends on $(l_{b-1}, m'_b, s^n_0)$. We do not require the relay to decode $m'_b$, but the source and the relay can still establish cooperation by directly performing a binning on the $z^n_b$ codeword to agree on the $u^n_{b+1}$ codeword to be used in the next block, thus providing the scheme with the title “cooperative-bin-forward”. The term cooperative is added to emphasize that the source and the relay agree on the binning and transmit coherently. Thus, the scheme achieves cooperation by communicating $l_b$ to the relay, instead of $m'_b$. While the relay is not required to decode the partial message, we still need the destination to be able to decode all parts of the transmitted message successfully. In the following, appropriate conditions are imposed so that the destination can utilize the state information at its disposal to achieve successful decoding.

**Decoding:**

The decoder performs backward decoding, starting from block $B$ and moving towards block 1, performing the following two steps for each block $b$:

1. Assuming that $l_b$ is known from previous operations, the decoder, for each $l_{b-1} \in [1 : 2nR]$, finds the unique $m'_b$ such that

$$\text{bin}_b(z^n_b(m'_b | l_{b-1}, s^n_0)) = l_b.$$  

Whenever a unique $m'_b$ cannot be found for some $l_{b-1}$, the decoder chooses any $m'_b$ arbitrarily. So after this operation, the decoder has chosen one $m'_b$ for each $l_{b-1}$, given its knowledge of $l_b$ and $s^n_0$. We will signify this explicitly by denoting the chosen message as $\tilde{m}'_b(l_{b-1}, s^n_0, l_b)$. We will signify this explicitly by denoting the chosen message as $\tilde{m}'_b(l_{b-1}, s^n_0, l_b)$. We will signify this explicitly by denoting the chosen message as $\tilde{m}'_b(l_{b-1}, s^n_0, l_b)$.

2. Now the decoder looks for the unique $(\hat{l}_{b-1}, \tilde{m}_n')$ such that

$$\left(x^n_{\hat{l}_{b-1}}(\hat{l}_{b-1}), z^n_b(\tilde{m}'_b(\hat{l}_{b-1}, s^n_0, l_b) | \hat{l}_{b-1}, s^n_0), x^n_{\tilde{m}'_b}(\tilde{m}'_b(\hat{l}_{b-1}, s^n_0, l_b), s^n_0), s^n_0, \tilde{y}_b^n \right) \in \mathcal{T}^{(n)}.$$  

**Probability of Error:**

In the following error analysis, we will observe that in order to achieve the largest rate the scheme will set $R' \approx H(Z|X_r, S)$. The causal multiplexing-demultiplexing strategy proposed above effectively creates a different codebook for $m'_b$ for each typical $s^n_0$ sequence. The total number of $z^n_b$ codewords constructed by the source encoder is therefore $\approx 2^{nH(S)} \cdot R' \approx 2^{nH(Z|X_r)}$. However, these codewords cannot be distinct since there are only $\approx 2^{nH(Z|X_r)}$ distinct sequences $z^n_b$ (conditioned on $l_{b-1}$). This implies that multiple $(s^n_0, m'_b)$ pairs will be mapped to the same codeword $z^n_b$ and therefore, the relay will not be able to decode $m'_b$ due to the lack of state information.

By symmetry, we can assume without loss of generality that the true messages and bin-indices corresponding to the current block are all 1, i.e.

$$(L_{b-1}, M'_b, M''_b) = (1, 1, 1).$$

We bound the probability of decoding error in block $b$ conditioned on successful decoding for blocks $\{B, B-1, \ldots, b + 1\}$, averaged over the randomness in the messages and codebook generation. In particular, successful decoding in block $b + 1$ means that $L_b$ has been decoded successfully, where we remind ourselves that

$$L_b = \text{Bin}_b(Z^n_b(1|1, S^n_0)).$$

An error occurs in block $b$ only if any of the following events occur:

(a) $\tilde{M}_b'(1, S^n_0, L_b) \neq 1$, 
(b) $(L_{b-1}, M''_b) \neq (1, 1)$ given $\tilde{M}_b'(1, S^n_0, L_b) = 1$.

We analyze the two terms events in the following two subsections.

**Event (a): $\tilde{M}_b'(1, S^n_0, L_{b-1}) \neq 1$:** We have

$$\Pr \left( \tilde{M}_b'(1, S^n_0, L_{b-1}) \neq 1 \right) = \Pr \left( \text{Bin}_b(Z^n_b(m'_b | 1, S^n_0)) = L_b \text{ for some } m'_b > 1 \right)$$

$$= \Pr \left( \text{Bin}_b(Z^n_b(m'_b | 1, S^n_0)) = \text{Bin}_b(Z^n_b(1|1, S^n_0)) \text{ for some } m'_b > 1 \right)$$

$$\leq \sum_{m'_b > 1} \Pr \left( \text{Bin}_b(Z^n_b(m'_b | 1, S^n_0)) = \text{Bin}_b(Z^n_b(1|1, S^n_0)) \right)$$

$$= \sum_{m'_b > 1} \Pr \left( \text{Bin}_b(Z^n_b(m'_b | 1, S^n_0)) = \text{Bin}_b(Z^n_b(1|1, S^n_0)), Z^n_b(m'_b | 1, S^n_0) = Z^n_b(1|1, S^n_0) \right)$$

$$\quad + \sum_{m'_b > 1} \Pr \left( \text{Bin}_b(Z^n_b(m'_b | 1, S^n_0)) = \text{Bin}_b(Z^n_b(1|1, S^n_0)), Z^n_b(m'_b | 1, S^n_0) \neq Z^n_b(1|1, S^n_0) \right)$$

$$= \sum_{m'_b > 1} \Pr \left( \text{Bin}_b(Z^n_b(m'_b | 1, S^n_0)) = \text{Bin}_b(Z^n_b(1|1, S^n_0)) \right)$$

$$\quad + \sum_{m'_b > 1} \Pr \left( \text{Bin}_b(Z^n_b(m'_b | 1, S^n_0)) = \text{Bin}_b(Z^n_b(1|1, S^n_0)), Z^n_b(m'_b | 1, S^n_0) \neq Z^n_b(1|1, S^n_0) \right)$$
where we use $\delta(\epsilon)$ to denote any function of $\epsilon$ for which $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. Hence, we get that
\[
\Pr \left( \hat{M}_b'(1, S_b^n, L_b) \neq 1 \right) \to 0 \text{ as } n \to \infty,
\]
if the following two constraints are satisfied:
\[
\begin{align*}
R' &< \tilde{R}, \\
R' &< H(Z | X_r, S) - \delta(\epsilon).
\end{align*}
\] (12) \hspace{2cm} (13)

**Event (b):** $(\hat{L}_{b-1}, \hat{M}_b') \neq (1, 1)$ given $\hat{M}_b'(1, S_b^n, L_b) = 1$: The probability of this event is upper bounded by
\[
\Pr \left( \text{Condition (I)} \text{ is not satisfied by } (l_{b-1}, m_b') = (1, 1) \mid \hat{M}_b'(1, S_b^n, L_b) = 1 \right) + \Pr \left( \text{Condition (I)} \text{ is satisfied for some } (l_{b-1}, m_b') \neq (1, 1) \mid \hat{M}_b'(1, S_b^n, L_b) = 1 \right).
\]

The first term goes to zero as $n \to \infty$ by the law of large numbers. The second term can be analyzed by standard applications of the packing lemma [28] as follows:
\[
\Pr \left( \text{Condition (I)} \text{ is satisfied for some } (l_{b-1}, m_b') \neq (1, 1) \mid \hat{M}_b'(1, S_b^n, L_b) = 1 \right) \leq \sum_{l_{b-1}=1, m_b' > 1} \Pr \left( (X_{rb}^b(1), Z_b^n(1 | l_{b-1}, S_b^n), X_b^n(m_b'| l_{b-1}, S_b^n), S_b^n, Y_b^n) \in T_{\epsilon}^{(n)} \right)
\]
\[
+ \sum_{l_{b-1}>1, m_b' \geq 1} \Pr \left( (X_{rb}^b(\hat{l}_{b-1}), Z_b^n(m_b'| \hat{l}_{b-1}, S_b^n, L_b), \hat{l}_{b-1}, S_b^n), X_b^n(m_b'| \hat{l}_{b-1}, S_b^n, L_b), S_b^n, Y_b^n) \in T_{\epsilon}^{(n)} \right)
\]
\[
\leq 2^{nR'} \cdot 2^{-n(H(Z \mid X_r, S) - \delta(\epsilon))} + 2^{n(R' + R'') \cdot 2^{-n(I(X; Y \mid X_r, Z, S) - \delta(\epsilon))}}.
\]

which follows by applying the packing lemma. Note that when $l_{b-1} > 1$, it so happens due to the codebook construction that $y_b^n$ is independent of all the other sequences for any value of $(m_b', m_b'')$. So the joint distribution of the sequences has the same factorization no matter what $m_b'$ is chosen for $l_{b-1} > 1$. The only fact that matters for our analysis is that at most one $m_b'$ has been chosen somehow for each $l_{b-1} > 1$. This allows us to write the fourth event as the union of at most $2^{n(R' + R'')} \cdot 2^{-n(I(X, X_r, Z, S) - \delta(\epsilon))}$ events, where each corresponds to a different value of $(l_{b-1}, m_b')$.

Thus, we get that
\[
\Pr \left( (\hat{L}_{b-1}, \hat{M}_b') \neq (1, 1) \mid \hat{M}_b'(1, L_b) = 1 \right) \to 0, \text{ as } n \to \infty,
\]
if
\[
\begin{align*}
R' &< I(X; Y \mid X_r, Z, S) - \delta(\epsilon), \\
\tilde{R} + R'' &< I(X, X_r; Y | S) - \delta(\epsilon).
\end{align*}
\] (14) \hspace{2cm} (15)

Performing Fourier-Motzkin elimination, letting $n \to \infty$, $B \to \infty$ and $\epsilon \to 0$, we get that the rates specified in Theorem [2] are indeed achieved by the achievability scheme presented in this section.

**Converse:**
Given a reliable code, we have by Fano’s inequality
\[
H(M | Y^n, S^n) \leq n\epsilon_n,
\]
where $\epsilon_n \to 0$ as $n \to \infty$. Then, we prove the first bound on $R$ as follows:
\[
nR = H(M)
\]
\[
\overset{(a)}{=} H(M | S^n)
\]
\[
\overset{(b)}{=} H(M, Z^n | S^n)
\]
\[
= H(Z^n | S^n) + H(M | Z^n, S^n)
\]
\[
\overset{(c)}{\leq} H(Z^n | S^n) + I(M; Y^n | Z^n, S^n) + n\epsilon_n
\]
\[\begin{align*}
&\leq \sum_{i=1}^{n} H(Z_i|S^n, Z^{i-1}) + \sum_{i=1}^{n} I(M, X_i; Y_i|Y^{i-1}, Z^n, S^n, X^n) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} H(Z_i|S^n, Z^{i-1}, X_{r,i}) + \sum_{i=1}^{n} I(M, X_i; Y_i|Y^{i-1}, Z^n, S^n, X^n, X^n_r) + n\epsilon_n \\
&\leq nH(Z_Q|S_Q, X_{r,Q}, Q) + nI(X_Q; Y_Q|Z_Q, S_Q, X_{r,Q}, Q) + n\epsilon_n \\
&\leq nI(X_Q; Y_Q|Z_Q, S_Q, X_{r,Q}) + n\epsilon_n,
\end{align*}\]

where

- \( Q \) is a random variable uniformly distributed over \([1 : n]\), independent of \((X^n, X^n_r, S^n, Y^n)\),
- (a) follows because \( M \) is independent of \( S^n \),
- (b) follows because \( Z^n \) is a function of \( M \) and \( S^n \),
- (c) follows by Fano’s inequality,
- (d) follows by the chain rule of mutual information and because \( X_i \) is a function of \((M, S^n)\),
- (e) follows because \( X_{r,i} \) is a function of \( Z^{i-1} \),
- (f) follows because conditioning reduces entropy and \( Y_i \) is independent of other random variables given \((X_i, X_{r,i}, S_i)\), and
- the final step follows because conditioning reduces entropy and \( Q - (X_Q, X_{r,Q}, S_Q) - Y_Q \).

The second bound on \( R \) is proved as follows:

\[nR = H(M) = H(M|S^n) \leq I(M; Y^n|S^n) + n\epsilon_n = I(M, X^n, X^n_r; Y^n|S^n) + n\epsilon_n \leq \sum_{i=1}^{n} I(X_i, X_{r,i}; Y_i|S_i) + n\epsilon_n = nI(X_Q, X_{r,Q}; Y_Q|S_Q, Q) + n\epsilon_n \leq nI(X_Q, X_{r,Q}; Y_Q|S_Q) + n\epsilon_n.\]

Thus, we have

\[R \leq \min(I(X_Q, X_{r,Q}; Y_Q|S_Q), H(Z_Q|S_Q, X_{r,Q}) + I(X_Q; Y_Q|Z_Q, S_Q, X_{r,Q})) + \epsilon_n.\]

Note that
- \( S_Q \) is independent of \( Q \) and has marginal pmf \( p_S \) due to the i.i.d. assumption on the state;
- since \( S_i \) is independent of \( X_{r,i} = x_{r,i}(Z^{i-1}) \) for all \( 1 \leq i \leq n \), we have that \( S_Q \) is independent of \( X_{r,Q} \);
- we have that \( p_{Y_Q|X_Q, X_{r,Q}, S_Q}(y|x, x_r, s) \) is equal to \( p_{Y|X, X_r, s}(y|x, x_r, s) \), since \( Y_Q \) is the output of the channel when the inputs are \((X_Q, X_{r,Q}, S_Q)\), and
- similarly, we also have \( Z_Q = z(X_Q, X_{r,Q}, S_Q) \).

Hence the joint pmf of the random variables \((X_Q, X_{r,Q}, S_Q, Y_Q)\) factorizes as

\[p_{S_Q, X_Q, X_{r,Q}, Y_Q}(s, x, x_r, y) = p_{S_Q}(s)p_{X_Q}(x_r)p_{X_Q|X_{r,Q}, S_Q}(x|x_r, s)p_{Y_Q|X_Q, X_{r,Q}, S_Q}(y|x, x_r, s) = p_{S}(s)p_{X_Q}(x_r)p_{Y_Q|X_Q, S_Q}(x|x_r, s)p_{Y|X_r, S}(y|x, x_r, s).\]

So, we can define the random variables \( X \triangleq X_Q, X_r \triangleq X_{r,Q}, S \triangleq S_Q, Z \triangleq Z_Q \) and \( Y \triangleq Y_Q \) to get

\[R \leq \min\{I(X, X_r; Y|S), H(Z|S, X_r) + I(X; Y|Z, S, X_r)\} + \epsilon_n,\]

where the pmf of the random variables has the form \( p_S(s)p_X(x_r)p_X|X_S,S(x|x_r, s)p_Y|X_r, S(y|x, x_r, s) \) and \( Z = z(X, X_r, S) \). Since \( \epsilon_n \rightarrow 0 \) as \( n \rightarrow \infty \), the converse is proved.

This concludes the proof of Theorem 2.

VII. PROOF OF THEOREM 3

The achievability part of this theorem is obtained by combining the cooperative-bin-forward scheme from the previous section with instantaneous relaying. This requires an auxiliary random variable, as described next.
Fix \( p_U(u)p_{X|U,S}(x|u,s) \), \( X_r = x_r(u,z) \) and \( \epsilon > 0 \). Split \( R \) as \( R' + R'' \), with the message \( m \) denoted accordingly as \( M = (M', M'') \). Divide the total communication time into \( B \) blocks, each of length \( n \).

**Codebook Generation:**

For each block \( b \in [1 : B] \), a codebook is generated independently of the other blocks as follows.

- **Cooperation codewords**
  
  Generate \( 2^{nR} \) codewords \( u^n_b(l_{b-1}) \), i.i.d. according to \( p_U \), where \( l_{b-1} \in [1 : 2^{n\tilde{R}}] \).

- **Cribbed codewords**
  
  For each \( l_{b-1} \) and each \( s \in S \), generate a codebook of \( 2^{nR'} \) codewords. The \( i \)th symbol of such a codeword is chosen independently according to \( p_{Z|U,S}(\cdot|u_{bi}(l_{b-1}), s) \). The result of this is that for each \( l_{b-1} \), each \( m'^b_b \in [1 : 2^{nR'}] \) and each \( s^n_b = (s_{b1}, s_{b2}, \ldots, s_{bn}) \), the source encoder can form an effective codeword \( z^n_b(m'^b_b|l_{b-1}, s^n_b) \), whose \( i \)th symbol can be causally chosen as the \( i \)th symbol of the \( m'^b_b \)-th codeword from the codebook corresponding to \( l_{b-1} \) and \( s_{bi} \).

- **Transmission codewords**
  
  For each \( l_{b-1} \), each \( m'^b_b \in [1 : 2^{nR}] \) and each \( s \in S \), generate a codebook of \( 2^{nR'} \) codewords. The \( i \)th symbol of such a codeword is generated independently according to \( p_{X|U,Z,S}(|u_{bi}(l_{b-1}), z_{bi}(m'^b_b|l_{b-1}, s), s) \). The result of this construction is that for each \( l_{b-1} \), each \( m'^b_b \in [1 : 2^{nR}] \), each \( m'^b_b \in [1 : 2^{nR''}] \) and each \( s^n_b \), the source encoder can form an effective codeword \( x^n_b(m'^b_b|l_{b-1}, m'^b_b, s^n_b) \), whose \( i \)th symbol can be causally chosen as the \( i \)th symbol of the \( m'^b_b \)-th codeword from the codebook corresponding to \( l_{b-1} \), \( m'^b_b \) and \( s_{bi} \).

- **Binning**
  
  Partition the set of all \( Z^n \) into \( 2^{n\tilde{R}} \) bins, by choosing a bin for each \( z^n \) independently and uniformly at random. Denote the index of the chosen bin for \( z^n \) by \( b_iu(z^n) \).

**Encoding:**

Fix \( l_0 = 1 \) and \((m'^b_0, m'^b_B) = (1, 1) \). Since the message in the last block is fixed, the effective rate of communication is \( \frac{R'-\tilde{R}}{B-1} R \), which can be made as close as desired to \( R \) by choosing a sufficiently large \( B \).

In block \( b \), assuming \( l_{b-1} \) is known to the source encoder, it transmits \( x^n_b(m'^b_b|l_{b-1}, m'^b_b, s^n_b) \). The relay transmits \( x^n_{lb} \), the \( i \)th symbol of which is obtained as \( x_r(u_{bi}(l_{b-1}), z_{bi}(m'^b_b|l_{b-1}, s^n_b)) \). At the end of block \( b \), the source and the relay set \( l_b \) to be the index of the bin containing \( z^n_b(m'^b_b|l_{b-1}, s^n_b) \).

**Decoding:**

The decoding operation is nearly the same as that in the previous section. The decoder performs the following two steps for each block \( b \), where \( b \in \{B, B-1, \cdots, 1\} \):

1. Assuming that \( l_b \) is known from previous operations, the decoder, for each \( l_{b-1} \in [1 : 2^{n\tilde{R}}] \), finds the unique \( m'^b_b \) such that
   
   \[
   b_iu(z^n_b(m'^b_b|l_{b-1}, s^n_b)) = l_b.
   \]

   Whenever a unique \( m'^b_b \) cannot be found for some \( l_{b-1} \), the decoder chooses any \( m'^b_b \) arbitrarily. So after this operation, the decoder has chosen one \( m'^b_b \) for each \( l_{b-1} \), given its knowledge of \( l_b \) and \( s^n_b \). We will signify this explicitly by denoting the chosen message as \( \hat{m}'_{lb}(l_{b-1}, s^n_b, l_b) \).

2. Now the decoder looks for the unique \((\hat{l}_{b-1}, \hat{m}'_{lb})\) such that
   
   \[
   \left( u^n_b(\hat{l}_{b-1}), z^n_b(\hat{m}'_{lb}(l_{b-1}, s^n_b, l_b)|l_{b-1}, s^n_b), x^n_b(\hat{m}'_{lb}(l_{b-1}, s^n_b, l_b), s^n_b), \hat{y}_b^n \right) \in \mathcal{T}_c^n. \tag{16}
   \]

**Probability of Error:**

By following a similar path as the previous section, we get the following conditions for vanishing probability of error:

\[
\begin{align*}
R' &< \tilde{R}, \\
R' &< H(Z|U,S) - \delta(\epsilon), \\
R'' &< I(X;Y|U,Z,S) - \delta(\epsilon), \\
\tilde{R} + R'' &< I(U;X:Y|S) - \delta(\epsilon).
\end{align*}
\]

Performing Fourier-Motzkin elimination completes the proof of achievability.
Converse:
Given a reliable code, define for each $1 \leq i \leq n$, the random variable $U_i \triangleq (Z_i^{i-1}, S_i^{i-1})$. Note that with this definition, $X_{r,i}$ becomes a function of $(U_i, Z_i)$. We have for any reliable code, by Fano’s inequality,
\[
H(M|Y^n, S^n) \leq n\epsilon_n.
\]
Then,
\[
nR = H(M|S^n)
\]
\[
= H(M, Z^n|S^n)
\]
\[
\leq H(Z^n|S^n) + I(M; Y^n|S^n, Z^n) + n\epsilon_n
\]
\[
\leq \sum_{i=1}^n H(Z_i|Z_i^{i-1}, S_i^{i-1}, S_i) + \sum_{i=1}^n I(X_i; Y_i|Z_i^{i-1}, S_i^{i-1}, Z_i, S_i) + n\epsilon_n,
\]
where the final step uses the fact that $Y_i$ is independent of other random variables given $X_i, Z_i^{i-1}, Z_i, S_i$, since $X_{r,i}$ is a function of $Z^i$. Using the definition of $U_i$ in the above, we get that
\[
nR \leq \sum_{i=1}^n H(Z_i|U_i, S_i) + \sum_{i=1}^n I(X_i; Y_i|U_i, Z_i, S_i) + n\epsilon_n
\]
\[
= nH(Z|U_Q, S_Q, Q) + nI(X_Q; Y_Q|U_Q, Z_Q, S_Q, Q) + n\epsilon_n,
\]
where $Q$ is uniformly distributed over $[1 : n]$ and independent of $U^n, X^n, X_r^n, S^n, Y^n$.

The remaining bound on $R$ is proved below:
\[
nR = H(M|S^n)
\]
\[
\leq I(M; Y^n|S^n) + n\epsilon_n
\]
\[
= \sum_{i=1}^n I(M; Y_i^{i-1}, S^n) + n\epsilon_n
\]
\[
(=) \sum_{i=1}^n I(M, X_i, X_{r,i}, Z_i^{i-1}, S_i^{i-1}; Y_i^{i-1}, S^n) + n\epsilon_n
\]
\[
\leq \sum_{i=1}^n I(X_i, X_{r,i}, Z_i^{i-1}, S_i^{i-1}; Y_i|S_i) + n\epsilon_n
\]
\[
= \sum_{i=1}^n I(U_i, X_i, X_{r,i}; Y_i|S_i) + n\epsilon_n
\]
\[
= \sum_{i=1}^n I(U_i, X_i; Y_i|S_i) + n\epsilon_n
\]
\[
= nI(U_Q, X_Q; Y_Q|S_Q, Q) + n\epsilon_n
\]
\[
\leq nI(Q, U_Q, X_Q; Y_Q|S_Q) + n\epsilon_n,
\]
where step $(a)$ is true since $(X_i, X_{r,i}, Z_i^{i-1}, S_i^{i-1})$ is a function of $(M, S^n)$, and step $(b)$ follows because $X_{r,i}$ is a function of $(U_i, Z_i)$, hence a function of $(U_i, X_i, S_i)$.

Following similar arguments as the previous section, we can define $U \triangleq (Q, U_Q)$, $X \triangleq X_Q$, $X_r \triangleq X_{r,Q}$, $S \triangleq S_Q$, $Z \triangleq Z_Q$ and $Y \triangleq Y_Q$ to get
\[
R \leq \min\{I(U, X; Y|S), H(Z|U, S) + I(X; Y|U, Z, S)\} + \epsilon_n,
\]
where the pmf of the random variables has the form $p_2(s)p_v(u)p_{X|U,S}(x|u,s)p_{Y|X,S}(y|x,r,s)$, $Z = z(X, S)$ and $X_r = x_r(U, Z)$. Since $\epsilon_n \to 0$ as $n \to 0$, the converse is completed. The bound on cardinality of the auxiliary random variable can be obtained using arguments based on Caratheodory’s theorem as described in [28] Appendix C.

This concludes the proof of Theorem 3.

VIII. PROOF OF THEOREM 4

The achievability scheme is more intricate than the previous sections due to the additional complications in the model, but builds on the same idea. Each encoder in the multiple access channel has an operation similar to the source encoder of the relay channels considered in the previous sections. The source encoder of the relay channels controlled the signal received at the relay by employing rate-splitting and superposition coding. This signal was used to choose a cooperation codeword for the next block. For the multiple access channel, each encoder controls the received signal at the other encoder in the same
manner, so that at the end of a block, these two cribbed signals are known to both encoders, which are used to agree on a cooperation codeword for the next block.

We point out the fact that it is crucial for both encoders to know both the cribbed signals at the end of a block, so that they can agree on a cooperation codeword for the next block. Encoder 1 knows \(z^n_1\), because it receives this signal. Since the model assumes that the cribbing link is of the form \(z_1(X_1, S_1)\), encoder 1 is able to control the \(z^n_1\) signal received by encoder 2, and thus encoder 1 also knows \(z^n_1\). Similarly, encoder 2 also knows \(z^n_2\) and \(z^n_2\). If \(Z_1\) were assumed to be \(z_1(X_1, X_2, S_1)\) or \(z_1(X_1, X_2, S_1, S_2)\), then encoder 1 would not have knowledge of the received signal at encoder 2 due to the involvement of \(X_2\) and \(S_2\), and it would not be possible to employ the scheme. The reason we are able to assume that the received signal at the relay in the previous sections is \(z(X, X_r, S)\) and not just \(z(X, S)\) is that the relay has no message of its own, so \(X_r\) in fact depends only on past signals transmitted by the source encoder, so the source encoder can still control the \(z^n\) signal.

Proof:

Fix a pmf \(p_U(u)p_{X_1}(x_1|u, s_1)p_{X_2}(x_2|u, s_2)\) and \(\epsilon > 0\). Split \(R_1\) as \(R'_1 + R''_1\), with the message \(M_1\) denoted accordingly as \((M'_1, M''_1)\), and similarly split \(R_2\) as \(R'_2 + R''_2\), with the message \(M_2\) denoted accordingly as \((M'_2, M''_2)\). Divide the total communication time into \(B\) blocks, each of length \(n\). In the achievability scheme proposed in [3] for the case of no state, \(M'_1\) corresponds to the part of \(M_1\) that is decoded by encoder 2. As can be guessed on the previous sections, this is not the case in the cooperative-bin-forward scheme presented below.

Codebook Generation:

For each block \(b \in [1 : B]\), a codebook is generated independently of the other blocks as follows:

- **Cooperation codewords**

  Generate \(2^{n(K_1 + R_2)}\) codewords \(u^n_b(l_{1,b-1}, l_{2,b-1})\), i.i.d. according to \(p_U\), where \(l_{1,b-1} \in [1 : 2^nR_1]\) and \(l_{2,b-1} \in [1 : 2^nR_2]\).

  In the following, we will sometimes abbreviate \((l_{1,b-1}, l_{2,b-1})\) by \(l_{b-1}\).

- **Cribbed codewords - I**

  For each \(l_{b-1}\) and each \(s_1 \in S_1\), generate a codebook of \(2^{nR'_1}\) codewords. The \(i\)th symbol of such a codeword is chosen independently according to \(p_{Z_1|U,S}(\cdot|u_{b}(l_{b-1}), s_1)\). The result of this is that for each \(l_{b-1}\), each \(m'_{1,b} \in [1 : 2^nR'_1]\) and each \(s_{1b} = (s_{1b_1}, s_{1b_2}, \ldots, s_{1b_n})\), encoder 1 can form an effective codeword \(z^n_{1b}(m'_{1,b}|l_{b-1}, s_{1b})\), whose \(i\)th symbol can be causally chosen as the \(i\)th symbol of the \(m'_{1,b}\)-th codeword from the codebook corresponding to \(l_{b-1}\) and \(s_{1b}i\).

- **Cribbed codewords - II**

  Similarly, for each \(l_{b-1}\) and each \(s_2 \in S_2\), generate a codebook of \(2^{nR'_2}\) codewords. The \(i\)th symbol of such a codeword is chosen independently according to \(p_{Z_2|U,S}(\cdot|u_{b}(l_{b-1}), s_2)\). The result of this is that for each \(l_{b-1}\), each \(m'_{2,b} \in [1 : 2^nR'_2]\) and each \(s_{2b} = (s_{2b_1}, s_{2b_2}, \ldots, s_{2b_n})\), encoder 2 can form an effective codeword \(z^n_{2b}(m'_{2,b}|l_{b-1}, s_{2b})\), whose \(i\)th symbol can be causally chosen as the \(i\)th symbol of the \(m'_{2,b}\)-th codeword from the codebook corresponding to \(l_{b-1}\) and \(s_{2bi}\).

- **Transmission codewords - I**

  For each \(l_{b-1}\), each \(m''_{1,b} \in [1 : 2^nR''_1]\) and each \(s_1 \in S_1\), generate a codebook of \(2^{nR''_1}\) codewords. The \(i\)th symbol of such a codeword is generated independently according to \(p_{X_1|U,Z_1,S}(\cdot|u_{b}(l_{b-1}), z_{1b}(m''_{1,b}|l_{b-1}, s_1), s_1)\). The result of this construction is that for each \(l_{b-1}\), each \(m''_{1,b} \in [1 : 2^nR''_1]\), each \(s_{1b} \in [1 : 2^nR''_1]\) and each \(s_{1b}i\), encoder 1 can form an effective codeword \(x_{1b}(m''_{1,b}|l_{b-1}, m''_{1,b}, s_{1b}i)\), whose \(i\)th symbol can be causally chosen as the \(i\)th symbol of the \(m''_{1,b}\)-th codeword from the codebook corresponding to \(l_{b-1}\), \(m''_{1,b}\) and \(s_{1bi}\).

- **Transmission codewords - II**

  Similarly, for each \(l_{b-1}\), each \(m''_{2,b} \in [1 : 2^nR''_2]\) and each \(s_2 \in S_2\), generate a codebook of \(2^{nR''_2}\) codewords. The \(i\)th symbol of such a codeword is generated independently according to \(p_{X_2|U,Z_2,S}(\cdot|u_{b}(l_{b-1}), z_{2b}(m''_{2,b}|l_{b-1}, s_2), s_2)\). The result of this construction is that for each \(l_{b-1}\), each \(m''_{2,b} \in [1 : 2^nR''_2]\), each \(s_{2b} \in [1 : 2^nR''_2]\) and each \(s_{2bi}\), encoder 2 can form an effective codeword \(x_{2b}(m''_{2,b}|l_{b-1}, m''_{2,b}, s_{2bi})\), whose \(i\)th symbol can be causally chosen as the \(i\)th symbol of the \(m''_{2,b}\)-th codeword from the codebook corresponding to \(l_{b-1}\), \(m''_{2,b}\) and \(s_{2bi}\).

- **Binning**

  Finally, partition the set \(Z^n_1\) into \(2^nR_1\) bins, by choosing a bin for each \(z^n_1i\) independently and uniformly at random from \([1 : 2^nR_1]\). Denote the chosen bin for \(z^n_1i\) by \(b_{1b}(z^n_1i)\). Similarly, partition the set \(Z^n_2\) into \(2^nR_2\) bins, by choosing a bin for each \(z^n_2\) independently and uniformly at random from \([1 : 2^nR_2]\). Denote the chosen bin for \(z^n_2\) by \(b_{2b}(z^n_2)\).

Encoding:

Henceforth, whenever convenient, we will abbreviate \((l_{1,b-1}, l_{2,b-1})\) by \(l_{b-1}\). Set \((l_{1,1}, l_{2,1}) = (1, 1)\) and

\[
(m'_{1,B}, m''_{1,B}, m'_{2,B}, m''_{2,B}) = (1, 1, 1, 1).
\]

Since the message in the last block is fixed, the effective rate of communication will be \((\frac{B-1}{B} R_1, \frac{B-1}{B} R_2)\), which can be made as close as desired to \((R_1, R_2)\) by choosing a sufficiently large \(B\). We now describe the encoding for block \(b\). Assume
both encoders have agreed upon some \( l_{b-1} = (l_{b-1,1}, l_{b-1,2}) \) based on operations in previous blocks. Then, encoders encode messages \( m_{1,b} \) and \( m_{2,b} \) by \( x_{1b}^n(m_{1,b}''|l_{b-1}, m_{1,b}', s_{1b}^n) \) and \( x_{2b}^n(m_{2,b}''|l_{b-1}, m_{2,b}', s_{2b}^n) \) respectively. This operation is valid because it does not require noncausal knowledge of the state sequences, it can be done “on the fly”. At the end of block \( b \), both encoders have knowledge of the cribbed codewords \( z_{1b}^n(m_{1,b}''|l_{b-1}, s_{1b}^n) \) and \( z_{2b}^n(m_{2,b}''|l_{b-1}, s_{2b}^n) \). They set

\[
l_{1,b} = \text{bin}_0(z_{1b}^n), \quad \text{and} \quad l_{2,b} = \text{bin}_0(z_{2b}^n).
\]

Decoding:

The decoder performs backward decoding. For each block \( b \in \{B, B-1, B-2, \ldots, 2\} \), assuming that \( l_b = (l_{b,1}, l_{b,2}) \) is known from previous operations:

1. The decoder first takes a pass through all \( l_{b-1} = (l_{b-1,1}, l_{b,2}) \) and for each \( l_{b-1} \), finds the unique \((m_{1,b}', m_{2,b}')\) such that

\[
\text{bin}_0(z_{1b}^n(m_{1,b}'|l_{b-1}, s_{1b}^n)) = l_{b,1} \quad \text{and} \quad \text{bin}_0(z_{2b}^n(m_{2,b}'|l_{b-1}, s_{2b}^n)) = l_{b,2}.
\]

Whenever a unique \((m_{1,b}', m_{2,b}')\) cannot be found for some \( l_{b-1} \), the decoder chooses any \((m_{1,b}', m_{2,b}')\) arbitrarily. So after this operation, the decoder has chosen one \((m_{1,b}', m_{2,b}')\) for each \( l_{b-1} \), given its knowledge of \((l_{b,1}, s_{1b}^n, s_{2b}^n)\). We will signify this explicitly by denoting the chosen messages as \( \hat{m}_{1,b}'(l_{b-1}, s_{1b}^n, l_{1,b}) \) and \( \hat{m}_{2,b}'(l_{b-1}, s_{2b}^n, l_{2,b}) \) respectively.

2. Now the decoder looks for the unique \((\hat{b}_{b-1}, \hat{m}_{1,b}'', \hat{m}_{2,b}'')\) such that \((17)\) appearing at the top of this page is satisfied.

Analysis of the Error Probability:

By symmetry, we can assume without loss of generality that the true messages and bin-indices corresponding to the current block are all 1, i.e.

\[
(L_{b-1}, M_{1,b}', M_{2,b}', M_{1,b}''', M_{2,b}''') = (1, 1, 1, 1).
\]

We bound the probability of decoding error in block \( b \) conditioned on successful decoding for blocks \( B, B-1, \ldots, b+1 \), averaged over the randomness in the messages and codebook generation. In particular, successful decoding in block \( b+1 \) means that \((L_{1,b}, L_{2,b})\) has been decoded successfully, where we remind ourselves that

\[
L_{1,b} = \text{Bin}_0(Z_{1b}^n(1|1, S_{1b}^n)) \quad \text{and} \quad L_{2,b} = \text{Bin}_0(Z_{2b}^n(1|1, S_{2b}^n)).
\]

An error occurs in block \( b \) only if any of the following events occur:

(a) \( \hat{M}_{1,b}'(1, S_{1b}^n, L_{1,b}) \neq 1 \)

(b) \( \hat{M}_{2,b}'(1, S_{2b}^n, L_{2,b}) \neq 1 \)

(c) \( (L_{b-1}, M_{1,b}'', M_{2,b}'') \neq (1, 1, 1) \) given \((M_{1,b}''(1, S_{1b}^n, L_{1,b}), M_{2,b}''(1, S_{2b}^n, L_{2,b})) = (1, 1)\)

We analyze each of the above three events in the following.

Event (a) \( \hat{M}_{1,b}'(1, S_{1b}^n, L_{1,b}) \neq 1 \): We have

\[
\Pr(\hat{M}_{1,b}'(1, S_{1b}^n, L_{1,b}) \neq 1) = \Pr(\text{Bin}_0(Z_{1b}^n(m_{1,b}'|1, S_{1b}^n)) = L_{1,b} \text{ for some } m_{1,b}' > 1) = \Pr(\text{Bin}_0(Z_{1b}^n(m_{1,b}'|1, S_{1b}^n)) = \text{Bin}_0(Z_{1b}^n(1|1, S_{1b}^n)) \text{ for some } m_{1,b}' > 1) = \sum_{m_{1,b}' > 1} \Pr(\text{Bin}_0(Z_{1b}^n(m_{1,b}'|1, S_{1b}^n)) = \text{Bin}_0(Z_{1b}^n(1|1, S_{1b}^n))) = \sum_{m_{1,b}' > 1} \Pr(\text{Bin}_0(Z_{1b}^n(m_{1,b}'|1, S_{1b}^n)) = \text{Bin}_0(Z_{1b}^n(1|1, S_{1b}^n), Z_{1b}^n(m_{1,b}'|1, S_{1b}^n) = Z_{1b}^n(1|1, S_{1b}^n)))
\]

\[
\leq \sum_{m_{1,b}' > 1} \Pr(\text{Bin}_0(Z_{1b}^n(m_{1,b}'|1, S_{1b}^n)) = \text{Bin}_0(Z_{1b}^n(1|1, S_{1b}^n), Z_{1b}^n(m_{1,b}'|1, S_{1b}^n) \neq Z_{1b}^n(1|1, S_{1b}^n)))
\]

\[
\leq \sum_{m_{1,b}' > 1} \Pr(Z_{1b}^n(m_{1,b}'|1, S_{1b}^n) = Z_{1b}^n(1|1, S_{1b}^n))
\]

The same analysis applies to the second event.
where we use $\delta(\epsilon)$ to denote any function of $\epsilon$ for which $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. Hence, we get that
\[
\Pr \left( \hat{M}_{1,b}(1, S_{1b}^{n}, L_{1,b}) \neq 1 \right) \to 0 \text{ as } n \to \infty,
\]
if the following two constraints are satisfied:
\[
R'_1 < \bar{R}_1, \\
R'_1 < H(Z_1|U, S_1) - \delta(\epsilon).
\]

Event (b): $\tilde{M}_{2,b}(1, S_{2b}^{n}, L_{2,b}) \neq 1$: Similar to the previous subsection, we can conclude that
\[
\Pr \left( \hat{M}_{2,b}(1, S_{2b}^{n}, L_{2,b}) \neq 1 \right) \to 0 \text{ as } n \to \infty,
\]
if the following two constraints are satisfied:
\[
R'_2 < \bar{R}_2, \\
R'_2 < H(Z_2|U, S_2) - \delta(\epsilon).
\]

Event (c): $(\hat{L}_{b-1}, \hat{M}_{1,b}', \hat{M}_{2,b}') \neq (1, 1, 1)$ given $(\hat{M}_{1,b}(1, S_{1b}^{n}, L_{1,b}), \hat{M}_{2,b}(1, S_{2b}^{n}, L_{2,b})) = (1, 1)$: The probability of this event is upper bounded by
\[
\Pr \left( \text{Condition } (17) \text{ is not satisfied by } (b_{-1}, m_{1,b}', m_{2,b}') = (1, 1, 1) \mid (\hat{M}_{1,b}(1, S_{1b}^{n}, L_{1,b}), \hat{M}_{2,b}(1, S_{2b}^{n}, L_{2,b})) = (1, 1) \right)
\]
\[
+ \Pr \left( \text{Condition } (17) \text{ is satisfied for some } (b_{-1}, m_{1,b}', m_{2,b}') \neq (1, 1, 1) \mid (\hat{M}_{1,b}(1, S_{1b}^{n}, L_{1,b}), \hat{M}_{2,b}(1, S_{2b}^{n}, L_{2,b})) = (1, 1) \right).
\]
The first term goes to zero as $n \to \infty$ by the law of large numbers.
The second term can be handled by considering the following four different cases separately and applying the packing lemma [28] appropriately in each case.

- $(\hat{L}_{b-1}, \hat{M}_{1,b}', \hat{M}_{2,b}') = (1, 1, >1)$ given $(\hat{M}_{1,b}(1, S_{1b}^{n}, L_{1,b}), \hat{M}_{2,b}(1, S_{2b}^{n}, L_{2,b})) = (1, 1)$
- $(\hat{L}_{b-1}, \hat{M}_{1,b}', \hat{M}_{2,b}') = (1, >1, 1)$ given $(\hat{M}_{1,b}(1, S_{1b}^{n}, L_{1,b}), \hat{M}_{2,b}(1, S_{2b}^{n}, L_{2,b})) = (1, 1)$
- $(\hat{L}_{b-1}, \hat{M}_{1,b}', \hat{M}_{2,b}') = (1, >1, >1)$ given $(\hat{M}_{1,b}(1, S_{1b}^{n}, L_{1,b}), \hat{M}_{2,b}(1, S_{2b}^{n}, L_{2,b})) = (1, 1)$
- $(\hat{L}_{b-1}, \hat{M}_{1,b}', \hat{M}_{2,b}') = (>1, *, *)$

A standard application of the packing lemma gives us that the probability of each of the first three events goes to zero as $n \to \infty$ if the following constraints are respectively satisfied:
\[
R''_1 < I(X_2; Y|U, Z_2, X_1, S_1, S_2) - \delta(\epsilon), \\
R''_1 < I(X_1; Y|U, Z_1, X_2, S_1, S_2) - \delta(\epsilon), \\
R'_1 + R'_2 < I(X_1, X_2; Y|S_1, S_2) - \delta(\epsilon).
\]
Applying the packing lemma gives us the following condition for vanishing probability of the fourth event,
\[
\bar{R}_1 + \bar{R}_2 + R''_1 + R''_2 < I(U, Z_1, Z_2, X_1, X_2; Y|S_1, S_2) - \delta(\epsilon)
\]
\[
= I(X_1, X_2; Y|S_1, S_2) - \delta(\epsilon).
\]
Collecting all the constraints established so far, we have
\[
R'_1 < \bar{R}_1, \tag{18}
\]
\[
R'_1 < H(Z_1|U, S_1) - \delta(\epsilon), \tag{19}
\]
\[
R'_2 < H(Z_2|U, S_2) - \delta(\epsilon), \tag{20}
\]
\[
R''_2 < I(X_2; Y|U, Z_2, X_1, S_1, S_2) - \delta(\epsilon), \tag{21}
\]
\[
R''_1 < I(X_1; Y|U, Z_1, X_2, S_1, S_2) - \delta(\epsilon), \tag{22}
\]
\[
R'_1 + R''_1 < I(X_1, X_2; Y|U, Z_1, Z_2, S_1, S_2) - \delta(\epsilon), \tag{23}
\]
\[
\bar{R}_1 + \bar{R}_2 + R''_1 + R''_2 < I(X_1, X_2; Y|S_1, S_2) - \delta(\epsilon). \tag{24}
\]
Performing Fourier-Motzkin elimination of $\bar{R}_1, \bar{R}_2, R'_1, R'_2, R''_1$ and $R''_2$, and letting $n \to \infty$, $B \to \infty$ and $\epsilon \to 0$, we get that communication at arbitrarily small error probability is possible for the rates specified in Theorem 4.
Converse:
The proof of the converse can be constructed by using similar arguments as in [4]. Note that we have by Fano’s inequality the following condition for any reliable code:
\[ H(M_1, M_2 | Y^n, S^n_1, S^n_2) \leq n\epsilon_n, \]
where \( \epsilon_n \to 0 \) as \( n \to \infty \). Define \( U_i \) as
\[ U_i \triangleq (Z_i^{i-1}, Z_i^{i-2}, S_i^{i-1}, S_i^{i-1}). \]
An upper bound on \( R_1 \) is established by the following:
\[
\begin{align*}
nR_1 &= H(M_1) \\
&\overset{(a)}{=} H(M_1|M_2, S^n_1, S^n_2) \\
&\overset{(b)}{=} H(M_1, Z_1^n|M_2, S^n_1, S^n_2) \\
&= H(Z_1^n|M_2, S^n_1, S^n_2) + H(M_1|Z_1^n, M_2, S^n_1, S^n_2) \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} H(Z_{1i}|Z_1^{i-1}, M_2, S^n_1, S^n_2) + \sum_{i=1}^{n} I(M_1; Y_i|Z_1^{i-1}, Z_1^n, M_2, S^n_1, S^n_2) + n\epsilon_n \\
&\overset{(d)}{=} \sum_{i=1}^{n} H(Z_{1i}|Z_1^{i-1}, M_2, S^n_1, S^n_2, Z_2^{i-1}) + \sum_{i=1}^{n} I(M_1; X_{1i}; Y_i|Z_1^{i-1}, Z_1^n, M_2, S^n_1, S^n_2, X_2^n) + n\epsilon_n \\
&\overset{(e)}{=} \sum_{i=1}^{n} H(Z_{1i}|Z_1^{i-1}, S^n_1, S^n_2, Z_2^{i-1}) + \sum_{i=1}^{n} I(X_{1i}; Y_i|Z_1^{i-1}, S^n_1, S^n_2, X_2^n) + n\epsilon_n \\
&\overset{(f)}{\leq} \sum_{i=1}^{n} H(Z_{1i}|U_i, S_1) + \sum_{i=1}^{n} I(X_{1i}; Y_i|U_i, S_1, S_2, X_2) + n\epsilon_n \\
&= nH(Z_{1Q}|U_Q, S_1Q, Q) + nI(X_{1Q}; Q|U_Q, Z_{1Q}, S_1Q, S_2Q, X_2Q, Q) + n\epsilon_n
\end{align*}
\]
where
- \( Q \) is a random variable uniformly distributed on \([1 : n]\), independent of other random variables,
- (a) follows because \( M_1 \) is independent of \((M_2, S^n_1, S^n_2)\),
- (b) follows since \( Z_1^n \) is a function of \((M_1, M_2, S^n_1, S^n_2)\),
- (c) follows by Fano’s inequality,
- (d) follows since \((X_2, Z_2)\) is a function of \((M_2, S^n_1, Z_1^{i-1})\),
- (e) follows since \( X_{1i} \) is a function of \((M_1, S^n_1, X_2^{i-1})\),
- (f) follows since (i) conditioning reduces entropy and (ii) conditioned on \((X_{1i}, X_2, S_1, S_2)\), \( Y_i \) is independent of \((M_1, M_2, X_1^n, X_2^n, S_1^n, S_2^n)\).
Similarly, we get an upper bound on \( R_2 \):
\[
nR_2 \leq nH(Z_{2Q}|U_Q, S_2Q, Q) + nI(X_{2Q}; Y_Q|U_Q, Z_{2Q}, S_1Q, S_2Q, X_1Q, Q) + n\epsilon_n.
\]
Applying similar arguments to the sum rate, we get:
\[
n(R_1 + R_2) = H(M_1, M_2) \\
= H(M_1, M_2|S^n_1, S^n_2) \\
= H(M_1, M_2, Z^n_1, Z^n_2|S^n_1, S^n_2) \\
= H(Z^n_1, Z^n_2|S^n_1, S^n_2) + H(M_1, M_2|Z^n_1, Z^n_2, S^n_1, S^n_2) \\
\leq \sum_{i=1}^{n} H(Z_{1i}, Z_{2i}|U_i, S_1, S_2) + H(M_1, M_2|Z^n_1, Z^n_2, S^n_1, S^n_2) + n\epsilon_n \\
\leq \sum_{i=1}^{n} H(Z_{1i}, Z_{2i}|U_i, S_1, S_2) + \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i|U_i, S_1, S_2) + n\epsilon_n \\
= nH(Z_{1Q}, Z_{2Q}|U_Q, S_1Q, S_2Q, Q) + nI(X_{1Q}, X_{2Q}; Y_Q|U_Q, Z_{1Q}, Z_{2Q}, S_1Q, S_2Q, Q) + n\epsilon_n.
\]
Another upper bound on the sum rate can be easily established as follows:

\[ n(R_1 + R_2) = H(M_1, M_2) \]
\[ = H(M_1, M_2 | S_1^n, S_2^n) \]
\[ \leq I(M_1, M_2; Y^n | S_1^n, S_2^n) + n\epsilon_n \]
\[ \leq \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i | S_{1i}, S_{2i}) + n\epsilon_n \]
\[ = nI(X_{1Q}, X_{2Q}; Y_Q | S_{1Q}, S_{2Q}, Q) + n\epsilon_n \]
\[ \leq nI(X_{1Q}, X_{2Q}; Y_Q | S_{1Q}, S_{2Q}) + n\epsilon_n. \]

We note the following four conditions:

- \((S_{1Q}, S_{2Q})\) is independent of \((Q, U_Q)\), and has the pmf \(p_{S_1, S_2}(s_1, s_2)\).
- \(X_{1Q} - (Q, U_Q, S_{1Q}) - (Q, U_Q, S_{2Q}) - X_{2Q}\),
- \(p_{Y_Q|U_Q, X_{1Q}, X_{2Q}, S_{1Q}, S_{2Q}}(y|u, x_1, x_2, s_1, s_2)\) is equal to \(p_{Y_1|X_1, x_2, s_1, s_2}(y|x_1, x_2, s_1, s_2)\).
- \(Z_{1Q} = z_1(X_{1Q}, S_{1Q})\) and \(Z_{2Q} = z_2(X_{2Q}, S_{2Q})\).

The proof of the first condition follows because \((S_{1i}, S_{2i})\) are generated independently of \(U_i = (Z_{1i}^{-1}, Z_{2i}^{-1}, S_{1i}^{-1}, S_{2i}^{-1})\) for all \(1 \leq i \leq n\), and due to the i.i.d. assumption on \((S_{1i}, S_{2i})\). To prove that the second condition is satisfied, consider the following. For any \(1 \leq i \leq n\):

\[
p(s_1^i, s_2^i, x_{1i}, x_{2i}, z_1^{-i}, z_2^{-i}) = \sum_{m_1, m_2} p(m_1, m_2, s_1^i, s_2^i, x_{1i}, x_{2i}, z_1^{-i}, z_2^{-i}) \]
\[
= \sum_{m_1, m_2} \left[ p(m_1)p(m_2)p(s_1^i, s_2^i)p(x_{1i}|m_1, s_1^i, z_2^{-i}) \left( \prod_{j=1}^{i-1} p(z_{1j}|m_1, s_1^j, z_2^{-1}) \right) \right.
\[
\times \left. p(x_{2i}|m_2, s_2^i, z_1^{-i}) \left( \prod_{j=1}^{i-1} p(z_{2j}|m_2, s_2^j, z_1^{-1}) \right) \right]
\]
\[
= p(s_1^i, s_2^i) \sum_{m_1} p(m_1)p(x_{1i}, z_1^{-i}|m_1, s_1^i, z_2^{-i}) \sum_{m_2} p(m_2)p(x_{2i}, z_2^{-i}|m_2, s_2^i, z_1^{-i})
\]
\[
= p(s_1^i, s_2^i)p(x_{1i}|z_1^{-i}, s_1^i, z_2^{-i})p(x_{2i}|z_2^{-i}, s_2^i, z_1^{-i}).
\]

The above factorization implies that

\[
X_{1i} - (S_{1i}^i, Z_{1i}^{-i}, Z_{2i}^{-i}) - (X_{2i}, S_{2i}), \quad \text{for all } 1 \leq i \leq n.
\]
\[
\Rightarrow X_{1i} - (U_i, S_{1i}) - (X_{2i}, S_{2i}), \quad \text{for all } 1 \leq i \leq n.
\]
\[
\Rightarrow X_{1i} - (U_i, S_{1i}) - (U_i, S_{2i}, X_{2i}), \quad \text{for all } 1 \leq i \leq n.
\]
\[
\Rightarrow X_{1Q} - (U_Q, S_{1Q}, Q) - (U_Q, S_{2Q}, X_{2Q}, Q).
\]

Similarly, we also have

\[
X_{2Q} - (U_Q, S_{2Q}, Q) - (U_Q, S_{1Q}, X_{1Q}, Q).
\]

These two Markov chains together imply the desired Markov chain \(X_{1Q} - (U_Q, S_{1Q}, Q) - (U_Q, S_{2Q}, Q) - X_{2Q}\). The third condition follows by the definition of the auxiliary random variable \(U_Q\) and because \(Y_Q\) is the channel output when the inputs are \(X_{1Q}, X_{2Q}, S_{1Q}, S_{2Q}\). Similarly, the fourth condition is true because \(Z_{1Q}\) and \(Z_{2Q}\) are the cribbed signals due to \((X_{1Q}, S_{1Q})\) and \((X_{2Q}, S_{2Q})\) respectively.

So we can define random variables \(U \triangleq (Q, U_Q)\), \(X_1 \triangleq X_{1Q}\), \(X_2 \triangleq X_{2Q}\), \(S_1 \triangleq S_{1Q}\), \(S_2 \triangleq S_{2Q}\), and \(Y \triangleq Y_Q\), such that the joint pmf of these random variables has the factorization

\[
p_{S_1, S_2}(s_1, s_2)p_U(u)p_{X_1|U, S_1}(x_1|u, s_1)p_{X_2|U, S_2}(x_2|u, s_2)p_{Y|X_1, X_2, S_1, S_2}(y|x_1, x_2, s_1, s_2),
\]

\(^5\)We drop subscripts denoting the random variables when analyzing the factorization to reduce the length of the expressions.

\(^6\)The expressions from this line onwards, read in a literal manner, might seem strange for representing factorizations due to the presence of terms of the form \(p(z_1^{-i}|z_2^{-i})p(z_1^{-i}|z_2^{-i})\), however these are valid factorizations, possibly containing redundant conditioning in some terms. The Markovity conclusions drawn from such a form are valid, because the factorization might at most contain redundant conditionings. An example might make the point clearer. Assume that a joint factorization is of the form \(p(x, y)p(x'|y)p(z|x', y)\). If we marginalize by summing over \((x, x')\) first and then over \((z, z')\), we get an expression of the form \(p(y|y')p(y'|y)\), while reversing the order of summation gives us \(p(y)p(y')\). Thus, \(Y\) and \(Y'\) are independent, so the former expression is equal to the latter, though the former contains redundant conditioning.
and \( Z_1 = z_1(X_1, S_1) \) and \( Z_2 = z_2(X_2, S_2) \). Noting that \( \epsilon_n \to 0 \) as \( n \to \infty \), the four constraints on the rates that we have established become

\[
R_1 \leq I(X_1; Y|U, X_2, Z_1, S_1, S_2) + H(Z_1|U, S_1),
\]

\[
R_2 \leq I(X_2; Y|U, X_1, Z_2, S_1, S_2) + H(Z_2|U, S_2),
\]

\[
R_1 + R_2 \leq I(X_1, X_2; Y|U, Z_1, Z_2, S_1, S_2) + H(Z_1, Z_2|U, S_1, S_2),
\]

\[
R_1 + R_2 \leq I(X_1, X_2; Y|S_1, S_2).
\]

Thus, we get that the region stated in Theorem 4 is an outer bound to the achievable rate region. The bound on cardinality of the auxiliary random variable can be obtained using arguments based on Caratheodory’s theorem as described in [28, Appendix C].

This concludes the proof of Theorem 4.

\[ \square \]

IX. PROOF OF THEOREM 5

The achievability part of this theorem builds on the cooperative-bin-forward scheme from the previous section by combining it with instantaneous relaying (a.k.a. codetrees or Shannon strategies). To avoid unnecessary repetition, we only provide the differences in the achievability part relative to that in the previous section.

**Proof:**

Fix a pmf \( p_U(u)p_{X_1|U, S_1}(x_1|u, s_1)p_{X_2|U, S_2, Z_1}(x_2|u, s_2, z_1) \) and \( \epsilon > 0 \). Rate-splitting is performed as in the previous section.

**Codebook Generation:**

The cooperation codewords and the codebooks used by Encoder 1 are generated in the same manner as the previous section. Encoder 2 generates codewords by treating the causally observed \( z_1 \) symbol in the same manner as the causally observed \( s_2 \) symbol. More precisely, the codebooks constructed by Encoder 2 are described in the following two paragraphs.

For each \( l_{b-1}, \) each \( s_2 \in S_2 \) and each \( z_1 \in Z_1 \), generate a codebook of \( 2^nR'_2 \) codewords. The \( i \)th symbol of such a codeword is chosen independently according to \( p_{Z_2|U, S_2, Z_1}(\cdot|u_{l_1}(l_{b-1}), s_2, z_1) \). The result of this is that for each \( l_{b-1}, \) each \( m'_{2,b} \in [1:2^nR'_2], \) each \( s_{2b} \) and each \( z_1^{b_1}, \) encoder 2 can form an effective codeword \( z_{2b}^n(m'_{2,b}|l_{b-1}, s_{2b}, z_1^{b_1}), \) whose \( i \)th symbol can be causally chosen as the \( i \)th symbol of the \( m'_{2,b} \)th codeword from the codebook corresponding to \( l_{b-1}, s_{2b}, \) and \( z_1^{b_1}. \)

For each \( l_{b-1}, \) each \( m'_{2,b} \in [1:2^nR'_2], \) each \( s_2 \in S_2 \) and each \( z_1 \in Z_1 \), generate a codebook of \( 2^nR'_2 \) codewords. The \( i \)th symbol of such a codeword is generated independently according to \( p_{X_2|U, Z_2, S_2, Z_1}(\cdot|u_{l_1}(l_{b-1}), z_{2b}(m'_{2,b}|l_{b-1}, s_2, z_1), s_2, z_1) \). The result of this construction is that for each \( l_{b-1}, \) each \( m'_{2,b} \in [1:2^nR'_2], \) each \( s_{2b} \) and each \( z_1^{b_1}, \) encoder 2 can form an effective codeword \( x_{2b}^n(m'_{2,b}|l_{b-1}, m'_{2,b}, s_{2b}, z_1^{b_1}), \) whose \( i \)th symbol can be causally chosen as the \( i \)th symbol of the \( m'_{2,b} \)th codeword from the codebook corresponding to \( l_{b-1}, m'_{2,b}, s_{2b} \) and \( z_1^{b_1}. \)

The binning is performed as in the previous section.

**Encoding:**

The encoding at Encoder 1 is identical to that in the previous section. Encoder 2 transmits \( x_{2b}^n(m''_{2,b}|l_{b-1}, m'_{2,b}, s_{2b}, z_1^{b_1}) \) which can be chosen as described above.

**Decoding:**

The decoder performs backward decoding over the blocks, where it performs two steps as in the previous section, with the first step changed to the following.

The decoder first takes a pass through all \( l_{b-1} = (l_{1,b-1}, l_{2,b-1}) \) and for each \( l_{b-1}, \) finds the unique \( (m'_{1,b}, m'_{2,b}) \) such that

\[
\bin_{l_1}(z_{1b}^n(m'_{1,b}|l_{b-1}, s_{1b}^n)) = l_{1,b} \quad \text{and} \quad \bin_{l_2}(z_{2b}^n(m'_{2,b}|l_{b-1}, s_{2b}^n, z_1^{b_1}(m'_{1,b}|l_{b-1}, s_{1b}^n))) = l_{2,b}.
\]

**Probability of Error:**

In the previous section, we obtained the conditions [18] for vanishing probability of error. The only difference now is that the fourth condition is replaced by

\[
R'_2 < H(Z_2|U, S_2, Z_1) - \delta(\epsilon).
\]

This is obtained by analyzing the probability of event \( (b) \) conditioned on the complement of event \( (a) \). The other conditions remain the same. Performing Fourier-Motzkin elimination of \( R_1, R'_2, R''_1, R''_2, R'''_1 \) and \( R''''_2 \), and letting \( n \to \infty, B \to \infty \) and \( \epsilon \to 0 \), we get that communication at arbitrarily small error probability is possible for the rates specified in Theorem 5.
Converse:
The only difference in the converse compares to that of the previous section is that we need to show a different bound on $R_2$ and we need to prove the factorization of the pmf. The bound on $R_1$ and the two bounds on the sum rate $R_1 + R_2$ are the same and require no changes in the arguments.

The new bound on $R_2$ can be shown by following the same line of arguments with minor changes. We provide the chain of inequalities below for completeness. The auxiliary random variable $U_i$ appearing below is defined to be $(Z_{1i}^{-1}, Z_{2i}^{-1}, s_{1i}^{-1}, s_{2i}^{-1})$.

\[
\begin{align*}
nR_2 &= H(M_2) \\
&= (a) H(M_2|X_1^n, X_2^n) \\
&= (b) H(M_2, Z_2^n|X_1^n, X_2^n) \\
&= H(Z_2^n|X_1^n, X_2^n) + I(M_2; X_1^n, X_2^n) + n\epsilon_n \\
&= \sum_{i=1}^{n} H(Z_{2i}|Z_{2i}^{-1}, s_{1i}^{-1}, s_{2i}^{-1}) + \sum_{i=1}^{n} I(M_2; Y_i|Y_i^{-1}, Z_{2i}^{-1}, s_{1i}^{-1}, s_{2i}^{-1}) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} H(Z_{2i}|U_i, X_{2i}, Z_{1i}) + \sum_{i=1}^{n} I(X_{2i}; Y_i|U_i, X_{2i}, Z_{1i}) + n\epsilon_n \\
&= nH(Z_{2Q}|U_Q, S_{2Q}, Z_{1Q}, Q) + nI(X_{2Q}; Y_Q|U_Q, Z_{2Q}, S_{1Q}, S_{2Q}, X_{1Q}, Q) + n\epsilon_n
\end{align*}
\]

where

- $Q$ is a random variable uniformly distributed on $[1 : n]$, independent of other random variables,
- (a) follows because $M_2$ is independent of $(X_1^n, X_2^n)$,
- (b) follows since $Z_2^n$ is a function of $(X_1^n, X_2^n, s_{1i}^{-1}, s_{2i}^{-1})$,
- (c) follows by Fano's inequality,
- (d) follows since $(X_{1i}, Z_{1i})$ is a function of $(M_1, s_{1i}^{-1}, Z_{2i}^{-1})$,
- (e) follows since $X_{2i}$ is a function of $(M_2, S_{2i}^{-1}, X_{1i})$,
- (f) follows since (i) conditioning reduces entropy and (ii) conditioned on $(X_{1i}, X_{2i}, S_{1i}, S_{2i}), Y_i$ is independent of $(M_1, M_2, X_{1i}^n, X_{2i}^n, S_{1i}^n, S_{2i}^n)$.

Regarding the joint pmf, we note the following conditions

- $(S_{1Q}, S_{2Q})$ is independent of $(Q, U_Q)$, and has the pmf $p_{S_1, S_2}(s_1, s_2)$,
- $X_{1Q} = (Q, U_Q, S_{1Q}) - S_{2Q}$,
- $X_{2Q} = (Q, U_Q, S_{2Q}, Z_{1Q}) - (X_{1Q}, S_{1Q})$,
- $p_{Y_{1Q}|U_{1Q}, X_{1Q}, X_{2Q}, S_{1Q}, S_{2Q}}(y|x_1, x_2, s_1, s_2)$ is equal to $p_{Y|X_1, X_2, S_1, S_2}(y|x_1, x_2, s_1, s_2)$,
- $Z_{1Q} = z_1(X_{1Q}, S_{1Q})$ and $Z_{2Q} = z_2(X_{2Q}, S_{2Q})$.

The first, fourth and fifth conditions do not need new arguments. The second and third condition can be proved as follows. For any $1 \leq i \leq n$,

\[
\begin{align*}
p(s_1^i, s_2^i, x_1^i, z_1^{i-1}, z_2^{i-1}) &= \sum_{m_1, m_2} p(m_1, m_2, s_1^i, s_2^i, x_1^i, z_1^{i-1}, z_2^{i-1}) \\
&= \sum_{m_1, m_2} p(m_1)p(m_2)p(s_1^i, s_2^i)p(x_1^i|m_1, s_1^i, z_2^{i-1}) \prod_{j=1}^{i-1} p(z_{1j}|m_1, s_1^j, z_2^{j-1}) \prod_{j=1}^{i-1} p(z_{2j}|m_2, s_2^j, z_1^j) \\
&= p(s_1^i, s_2^i)p(x_1^i, z_1^{i-1}|s_1^i, z_2^{i-1})p(z_2^{i-1}|s_2^i, z_1^i).
\end{align*}
\]

This implies $X_{1Q} = (U_Q, S_{1Q}, Q) - S_{2Q}$. For the third condition, we have for any $1 \leq i \leq n$,

\[
p(s_1^i, s_2^i, x_1^i, x_2^i, z_1^{i-1}, z_1^{i-1})
\]
Thus, we get that the region stated in Theorem 5 is an outer bound to the achievable rate region. The bound on cardinality of the relay channel with two state components, one known to the source and the other to the relay, with an uninformed destination, some cooperation between the source and the relay.

involve a compression operation at the relay, where part of the compression can be reconstructed at the source, thus enabling suggest an approach. Note that binning is a form of compression, so a natural extension of cooperative-bin-forward might establish by exploiting the deterministic components in the models. To ensure some kind of agreement between nodes in via a decoding operation at the relay. In cooperative-bin-forward, the decoding operation was removed and the agreement was used by both schemes is that different nodes agree on some information. In partial-decode-forward, the agreement is established partial-decode-forward, the agreement was established become

**ACKNOWLEDGMENTS**

The authors gratefully acknowledge helpful discussions with Xiugang Wu.
REFERENCES

[1] A. Gamal and M. Aref, “The capacity of the semideterministic relay channel (corresp.),” IEEE Transactions on Information Theory, vol. 28, no. 3, pp. 536–536, May 1982.

[2] Y.-H. Kim, “Capacity of a class of deterministic relay channels,” IEEE Transactions on Information Theory, vol. 54, no. 3, pp. 1328–1329, March 2008.

[3] F. Willems and E. van der Meulen, “The discrete memoryless multiple-access channel with cribbing encoders,” IEEE Transactions on Information Theory, vol. 31, no. 3, pp. 313–327, May 1985.

[4] H. Asnani and H. Permuter, “Multiple-access channel with partial and controlled cribbing encoders,” IEEE Transactions on Information Theory, vol. 59, no. 4, pp. 2252–2266, April 2013.

[5] A. El Gamal, N. Hassanzadeh, and J. Mammen, “Relay networks with delays,” IEEE Transactions on Information Theory, vol. 53, no. 10, pp. 3413–3431, Oct 2007.

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

[7] B. Akhbari, M. Mirmohseni, and M. Aref, “Compress-and-forward strategy for relay channel with causal and non-causal channel state information,” IET Commun., vol. 4, no. 10, pp. 1174–1186, Jul 2010.

[8] M. Khormuji and M. Skoglund, “The relay channel with partial causal state information,” in International Symposium on Information Theory and Its Applications, Dec 2008, pp. 1–6.

[9] M. Khormuji, A. El Gamal, and M. Skoglund, “State-dependent relay channel: Achievable rate and capacity of a semideterministic class,” IEEE Transactions on Information Theory, vol. 59, no. 5, pp. 2629–2638, May 2013.

[10] Z. Deng, F. Lang, B.-Y. Wang, and S. mei Zhao, “Capacity of a class of relay channel with orthogonal components and non-causal channel state,” in IEEE International Symposium on Information Theory, July 2013, pp. 2696–2700.

[11] A. Zaidi, S. Kotagiri, J. Laneman, and L. Vandendorpe, “Cooperative relaying with state available noncausally at the relay,” IEEE Transactions on Information Theory, vol. 56, no. 5, pp. 2272–2298, May 2010.

[12] A. Zaidi, S. Shamai, P. Piantanida, and L. Vandendorpe, “Bounds on the capacity of the relay channel with noncausal state at the source,” IEEE Transactions on Information Theory, vol. 59, no. 5, pp. 2639–2672, May 2013.

[13] I. Aguerri and D. Gunduz, “Capacity of a class of relay channels with state,” in IEEE Information Theory Workshop, Sep 2012, pp. 277–281.

[14] A. Lapidoth and Y. Steinberg, “The multiple-access channel with causal side information: Common state,” IEEE Transactions on Information Theory, vol. 59, no. 1, pp. 32–50, Jan 2013.

[15] ——, “The multiple-access channel with causal side information: Double state,” IEEE Transactions on Information Theory, vol. 59, no. 3, pp. 1379–1393, March 2013.

[16] M. Li, O. Simeone, and A. Yener, “Multiple access channels with states causally known at transmitters,” IEEE Transactions on Information Theory, vol. 59, no. 3, pp. 1394–1404, March 2013.

[17] T. Philosof, R. Zamir, U. Erez, and A. Khisti, “Lattice strategies for the dirty multiple access channel,” IEEE Transactions on Information Theory, vol. 57, no. 8, pp. 5006–5035, Aug 2011.

[18] S. Sigurjonsson and Y.-H. Kim, “On multiple user channels with state information at the transmitters,” in Proceedings of International Symposium on Information Theory, Sep 2005, pp. 72–76.

[19] S. Jafar, “Capacity with causal and noncausal side information: A unified view,” IEEE Transactions on Information Theory, vol. 52, no. 12, pp. 5468–5474, Dec 2006.

[20] G. Como and S. Yuksel, “On the capacity of memoryless finite-state multiple-access channels with asymmetric state information at the encoders,” IEEE Transactions on Information Theory, vol. 57, no. 3, pp. 1267–1273, Mar 2011.

[21] N. Sen, F. Alajaji, S. Yuksel, and G. Como, “Memoryless multiple access channel with asymmetric noisy state information at the encoders,” IEEE Transactions on Information Theory, vol. 59, no. 11, pp. 7052–7070, Nov 2013.

[22] F. Willems, “The discrete memoryless multiple access channel with partially cooperating encoders (corresp.),” IEEE Transactions on Information Theory, vol. 29, no. 3, pp. 441–445, May 1983.

[23] A. Somekh-Baruch, S. Shamai, and S. Verdu, “Cooperative multiple-access encoding with states available at one transmitter,” IEEE Transactions on Information Theory, vol. 54, no. 10, pp. 4448–4469, Oct 2008.

[24] S. Bross and A. Lapidoth, “The state-dependent multiple-access channel with states available at a cribbing encoder,” in IEEE 26th Convention of Electrical and Electronics Engineers in Israel, Nov 2010, pp. 665–669.

[25] H. Permuter, S. Shamai, and A. Somekh-Baruch, “Message and state cooperation in multiple access channels,” IEEE Transactions on Information Theory, vol. 57, no. 10, pp. 6379–6396, Oct 2011.

[26] M. Zamanighomi, M. Emadi, F. Chaharsoghi, and M.-R. Aref, “Multiple access channel with correlated channel states and cooperating encoders,” in IEEE Information Theory Workshop, Oct 2011, pp. 628–632.

[27] M. Emadi, M. Khormuji, M. Skoglund, and M.-R. Aref, “On the achievable rate region of a state-dependent mac with cooperating encoders,” in Swedish Communication Technologies Workshop, Oct 2012, pp. 48–52.

[28] A. E. Gamal and Y.-H. Kim, Network Information Theory. New York, NY, USA: Cambridge University Press, 2012.