Cooperative Binning for Semideterministic Channels

Ritesh Kolte, *Student Member, IEEE*, Ayfer Özgür, *Member, IEEE*, and Haim Permuter, *Senior Member, IEEE*

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 suboptimal 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
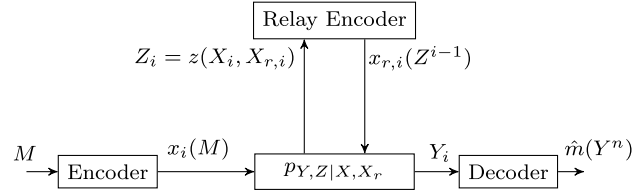
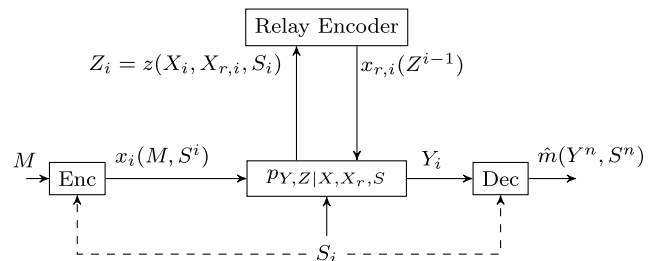Fig. 1. Semideterministic Relay Channel.

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

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), where the relay also does not decode but bins its received signal, however this bin-index is communicated to the destination without any cooperation with the source node. Indeed, hash-forward has been considered in the context of the primitive relay channel [2], where the multiple-access channel (MAC) from the source and the relay to the destination has orthogonal components and therefore cooperative transmission between the source and the relay is not needed. For the vanilla semideterministic relay

Fig. 3.   Multiple Access Channel with Strictly-Causal Partial-Cribbing Encoders.

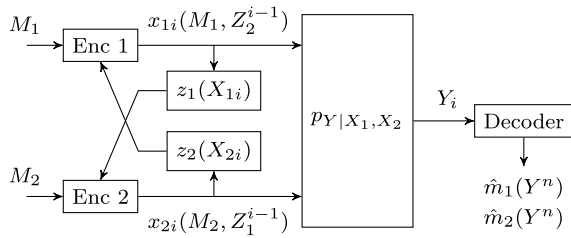Fig. 4.   State-dependent Multiple-Access Channel with Strictly-Causal Partial-Cribbing Encoders.

channel in Figure 1, cooperative-bin-forward recovers the capacity achieved by partial-decode-forward. However, partial-decode-forward is suboptimal for the state-dependent semideterministic relay channel in Figure 2, while 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.[1] 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 [4]. Here, we consider a natural extension of this setup with states, depicted in Figure 4.[2] 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.

Fig. 5.   State-dependent Semideterministic Relay-Without-Delay Channel with Causal State Information at Source and Destination. For this to make sense, we need to define the received signal at the relay so that it does not depend on the current transmission of the relay, in contrast to Figure 2.

Fig. 6.   State-dependent Multiple-Access Channel with Partially-Cribbing Encoders, one strictly causal and the other causal. Note that both links cannot be changed to causal.

Hence, the model in Figure 4 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

---

[1]Earlier work [4] has observed that the best rates that can be achieved by assuming the overheard signal to be a very coarse quantization of the channel input, are close to the best rates that can be achieved under the unrealistic assumption of perfect cribbing (overhearing via a noiseless link). Thus, to work with the noisy overheard signals available in practice, we can in fact manually perform quantization on the overheard signals without any appreciable loss in performance. Furthermore, modeling the output of this quantization as a deterministic function of the channel input is reasonable since the quantization can be made coarse.

[2]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 $z_2(X_{1i}, X_{2i}, S_{2i})$.
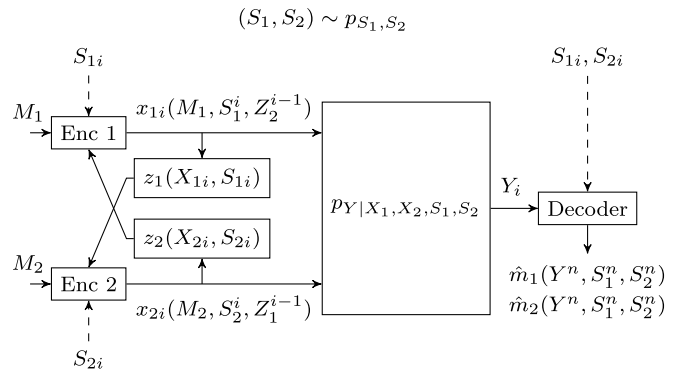
relay-without-delay channel, the transmission of the relay is allowed to depend on its past and current received signal. The capacity region for this setup without state is characterized in [5], using partial-decode-forward combined with instantaneous relaying (a.k.a. codetrees or Shannon strategies). 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 [4], 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.

To summarize, the main contribution of this paper is to introduce a new multi-user coding strategy which we call cooperative-bin-forward. This new strategy can be regarded as a generalization of partial-decode-forward and hash-forward from previous literature. As in partial-decode-forward, nodes cooperate to send codewords to the destination, however this is done without requiring any partial decoding; instead received signals are simply binned as in hash-forward. In this respect, the main contribution of our paper is to show that cooperation can be induced without any decoding (full or partial) and to show that this is sufficient to achieve the capacity in a number of practically well-motivated semideterministic settings with states.

A. 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 at decoder was considered in [13], and optimality of a partial-decode-compress-forward scheme was proved. To the best of our knowledge, the state-dependent relay channels considered in this paper have not been previously studied, and as mentioned in the introduction, standard combinations of available ideas are not sufficient to obtain good achievability schemes.

There has been interesting recent work on state-dependent multiple access channels where the state is only known to the encoders. The encoders are not allowed to cooperate in these setups, so the main challenge is to handle the lack of state information at the destination. When the state is known in a strictly causal manner, it was shown in [14]–[16] that in contrast to point-to-point channels, ignoring the state information is suboptimal. An improvement in achievable rates can be obtained by explicitly communicating the stale state information to the destination. The aforementioned papers accomplished this using block-Markov schemes that encode

messages of the current block as well as some information about the state and messages from previous block. When the state information is known noncausally, the work [17] considered the *dirty-paper* special case (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 [3] and [22], 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 [24]. 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. The works [28] and [29] consider state-dependent multiple access channels in which one of the encoders knows both messages, and the state information is available strictly causally, causally or non-causally to one or both of the encoders.

B. 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 ϵ-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 $p_S(s)p_{Y|X,X_r,S}(y|x,x_r,s)$ 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:

$$x_i : [1 : 2^{nR}] \times \mathcal{S}^i \to \mathcal{X}, \quad 1 \le i \le n,$$
$$x_{r,i} : \mathcal{Z}^{i-1} \to \mathcal{X}_r, \quad 1 \le i \le n,$$
$$\hat{m} : \mathcal{Y}^n \times \mathcal{S}^n \to [1 : 2^{nR}],$$

such that

$$\Pr\{\hat{m}(Y^n, S^n) \ne M\} \le \epsilon,$$

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 $p_S(s)p_{Y|X,X_r,S}(y|x, x_r, s)$ 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_i:

$$x_{r,i} : \mathcal{Z}^i \to \mathcal{X}_r, \quad 1 \le i \le n.$$

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 $p_{S_1,S_2}(s_1, s_2)$ $p_{Y|X_1,X_2,S_1,S_2}(y|x_1, x_2, s_1, s_2)$ 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:

$$x_{1,i} : [1 : 2^{nR_1}] \times \mathcal{S}_1^i \times \mathcal{Z}_2^{i-1} \to \mathcal{X}_1, \quad 1 \le i \le n,$$
$$x_{2,i} : [1 : 2^{nR_2}] \times \mathcal{S}_2^i \times \mathcal{Z}_1^{i-1} \to \mathcal{X}_2, \quad 1 \le i \le n,$$
$$\hat{m}_1 : \mathcal{Y}^n \times \mathcal{S}_1^n \times \mathcal{S}_2^n \to [1 : 2^{nR_1}],$$
$$\hat{m}_2 : \mathcal{Y}^n \times \mathcal{S}_1^n \times \mathcal{S}_2^n \to [1 : 2^{nR_2}],$$

such that

$$\Pr\left\{\left(\hat{m}_1(Y^n, S_1^n, S_2^n), \hat{m}_2(Y^n, S_1^n, S_2^n)\right) \ne (M_1, M_2)\right\} \le \epsilon,$$

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, which begin with an informal description of how the idea of cooperative-bin-forward is used to develop the achievability scheme, before presenting the formal proof.

A. State-Dependent Semideterministic Relay Channels

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 proofs for the more complicated setups.

Theorem 1: The capacity of the semideterministic relay channel, shown in Figure 1, is given by

$$\max_{p_{X,X_r}(x,x_r)} \min\{I(X, X_r; Y),\ H(Z|X_r) + I(X; Y|X_r, Z)\}. \tag{1}$$

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

$$\max_{p_{X_r} p_{X|X_r,S}} \min \left\{ \begin{array}{l} I(X, X_r; Y|S), \\ H(Z|S, X_r) + I(X; Y|S, X_r, Z) \end{array} \right\}. \tag{2}$$

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 $p_{X_r}(x_r)p_{X|X_r,S}(x|x_r, s)$, 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 the pmf.

The cooperative aspect of the scheme is captured by the fact that the pmf has the form $p_{X|X_r,S}(x|x_r, s)$. Combining bin-forward with state-multiplexing-demultiplexing without cooperation would correspond to choosing the pmf in the form $p_{X|S}(x|s)$ and therefore would lead to smaller rate.

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 achieves 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

$$\max_{p_U\, p_{X|U,S}\, X_r = x_r(U,Z)} \min \left\{ \begin{array}{l} I(U, X; Y|S), \\ H(Z|U, S) + I(X; Y|U, Z, S) \end{array} \right\}, \tag{3}$$

where $|\mathcal{U}| \leq |\mathcal{S}|\,(|\mathcal{X}||\mathcal{X}_r| - 1) + 2$.

The capacity region for the setup of Theorem 3 in the absence of states is characterized in [5, Proposition 7]. Setting S to be deterministic 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.

B. State-Dependent Multiple-Access Channels

The next two theorems describe the capacity regions for the two state-dependent 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

$$\begin{aligned} R_1 &\leq I(X_1; Y|U, X_2, Z_1, S_1, S_2) + H(Z_1|U, S_1), \\ R_2 &\leq I(X_2; Y|U, X_1, Z_2, S_1, S_2) + H(Z_2|U, S_2), \\ R_1 + R_2 &\leq I(X_1, X_2; Y|U, Z_1, Z_2, S_1, S_2) \\ &\quad + H(Z_1, Z_2|U, S_1, S_2), \\ R_1 + R_2 &\leq I(X_1, X_2; Y|S_1, S_2), \end{aligned} \tag{4}$$

for pmf of the form

$$p_U(u) p_{X_1|U,S_1}(x_1|u, s_1) p_{X_2|U,S_2}(x_2|u, s_2),$$

with $Z_1 = z_1(X_1, S_1)$ and $Z_2 = z_2(X_2, S_2)$, with $|\mathcal{U}| \leq \min\{|\mathcal{S}_1||\mathcal{S}_2|(|\mathcal{Y}| - 1) + 4,\ |\mathcal{S}_1||\mathcal{S}_2|(|\mathcal{X}_1||\mathcal{X}_2| - 1) + 3\}$.

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. At the other extreme, the result for strictly causal partial cribbing from [4] is recovered by setting S_1 and S_2 to be deterministic.

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

$$\begin{aligned} R_1 &\leq I(X_1; Y|U, X_2, Z_1, S_1, S_2) + H(Z_1|U, S_1), \\ R_2 &\leq I(X_2; Y|U, X_1, Z_2, S_1, S_2) + H(Z_2|U, S_2, Z_1), \\ R_1 + R_2 &\leq I(X_1, X_2; Y|U, Z_1, Z_2, S_1, S_2) \\ &\quad + H(Z_1, Z_2|U, S_1, S_2), \\ R_1 + R_2 &\leq I(X_1, X_2; Y|S_1, S_2), \end{aligned} \tag{5}$$

for pmf of the form

$$p_U(u) p_{X_1|U,S_1}(x_1|u, s_1) p_{X_2|U,S_2,Z_1}(x_2|u, s_2, z_1),$$

with $Z_1 = z_1(X_1, S_1)$ and $Z_2 = z_2(X_2, S_2)$, with $|\mathcal{U}| \leq \min\{|\mathcal{S}_1||\mathcal{S}_2|(|\mathcal{Y}| - 1) + 4,\ |\mathcal{S}_1||\mathcal{S}_2|(|\mathcal{X}_1||\mathcal{X}_2| - 1) + 3\}$.

Note that if $p_{X_2|U,S_2,Z_1}(x_2|u, s_2, z_1)$ in Theorem 5 is replaced by $p_{X_2|U,S_2}(x_2|u, s_2)$, then the region becomes identical to that in Theorem 4. Setting S_1 and S_2 to be deterministic 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

$$p_S(s) = \begin{cases} p/2, & \text{if } s = 0, \\ p/2, & \text{if } s = 1, \\ 1 - p, & \text{if } s = 2, \end{cases}$$

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 [30, Fig. 7.7], while the channel $p_{Y|X,X_r,S}$ 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 $1 - H_2\left(\frac{p}{2}\right)$. 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 $1 - H_2\left(\frac{p}{2}\right)$ 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 state-multiplexed codebook at the source. Hence, it can recover the message by finding the unique source codeword, if any, 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 allowed in the general models that we consider in this paper, the source and relay need to perform the bin-forward operation in a cooperative fashion, as can be realized by considering the special case of the semideterministic relay channel without states (Theorem 1), which already requires dependence between X and X_r.

V. PROOF OF THEOREM 1

We demonstrate in this section the achievability of 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 $2^{n\widetilde{R}}$ codewords $x_{rb}^n(l_{b-1})$ i.i.d. according to p_{X_r}, where $l_{b-1} \in [1 : 2^{n\widetilde{R}}]$.
- *Cribbed Codewords[3]:*
  For each l_{b-1}, generate $2^{nR'}$ codewords $z_b^n(m_b'|l_{b-1})$ according to $\prod_{i=1}^{n} p_{Z|X_r}(\cdot|x_{rbi}(l_{b-1}))$, where $m_b' \in [1 : 2^{nR'}]$.
- *Transmission Codewords:*
  For each l_{b-1} and each m_b', generate $2^{nR''}$ codewords $x_b^n(m_b''|l_{b-1}, m_b')$, where $m_b'' \in [1 : 2^{nR''}]$ according to $\prod_{i=1}^{n} p_{X|X_r,Z}(\cdot|x_{rbi}(l_{b-1}), z_{bi}(m_1'|l_{b-1}))$.
- *Binning:*
  Partition the set of all $\mathcal{Z}^n$ into $2^{n\widetilde{R}}$ 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 $\text{bin}_b(z^n)$.

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 $\frac{B-1}{B}R$, 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 $x_b^n(m_b''|l_{b-1}, m_b')$. The relay transmits $x_{rb}^n(l_{b-1})$. 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 $z_b^n(m_b'|l_{b-1})$. The source and the relay set l_b to be the index of the bin containing $z_b^n(m_b'|l_{b-1})$.

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 l_b is known from previous operations, the decoder, for each $l_{b-1} \in [1 : 2^{n\widetilde{R}}]$, finds the unique m_b' such that

$$\text{bin}_b(z_b^n(m_b'|l_{b-1})) = l_b.$$

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 $\hat{m}_b'(l_{b-1}, l_b)$.

---

[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_b^n codewords as cribbed codewords.

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

$$\left(x_{rb}^n \le (\hat{l}_{b-1}), z_b^n(\hat{m}_b'(\hat{l}_{b-1}, l_b)|\hat{l}_{b-1}),\right.$$
$$\left. x_b^n(\hat{m}_b''|\hat{l}_{b-1}, \hat{m}_b'(\hat{l}_{b-1}, l_b)), y_b^n\right) \in \mathcal{T}_\epsilon^{(n)}. \quad (6)$$

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 l_b.

Probability of Error: The following error analysis reveals that in order to achieve the highest rate, the scheme will set $R' \approx \widetilde{R} \approx H(Z|X_r)$. It is easy to see that when $R' \approx H(Z|X_r)$, 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 $R' \approx \widetilde{R}$, 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 $R' \approx \widetilde{R}$, 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

$$L_b = \mathsf{Bin}_b(Z_b^n(1|1)).$$

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

(a) $\hat{M}_b'(1, L_b) \ne 1$,
(b) $(\hat{L}_{b-1}, \hat{M}_b'') \ne (1, 1)$ 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 $l_{b-1} \ne 1$, which is what the first step of the decoding does.[4] This allows us to restrict our attention to at most $2^{n(\widetilde{R}+R'')}$ options while analyzing the probability of decoding L_{b-1} incorrectly during the second decoding step, instead of $2^{n(\widetilde{R}+R'+R'')}$.

---

[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): $\hat{M}_b'(1, L_b) \ne 1$: We have

$$\Pr\left(\hat{M}_b'(1, L_b) \ne 1\right)$$
$$= \Pr\left(\mathsf{Bin}_b(Z_b^n(m_b'|1)) = L_b \text{ for some } m_b' > 1\right)$$
$$\overset{(i)}{=} \Pr\left(\mathsf{Bin}_b(Z_b^n(m_b'|1)) = \mathsf{Bin}_b(Z_b^n(1|1)) \text{ for some } m_b' > 1\right)$$
$$\le \sum_{m_b' > 1} \Pr\left(\mathsf{Bin}_b(Z_b^n(m_b'|1)) = \mathsf{Bin}_b(Z_b^n(1|1))\right)$$
$$= \sum_{m_b' > 1} \Pr\left(\begin{array}{c} \mathsf{Bin}_b(Z_b^n(m_b'|1)) = \mathsf{Bin}_b(Z_b^n(1|1)), \\ Z_b^n(m_b'|1) = Z_b^n(1|1) \end{array}\right)$$
$$+ \sum_{m_b' > 1} \Pr\left(\begin{array}{c} \mathsf{Bin}_b(Z_b^n(m_b'|1)) = \mathsf{Bin}_b(Z_b^n(1|1)), \\ Z_b^n(m_b'|1) \ne Z_b^n(1|1) \end{array}\right)$$
$$\le \sum_{m_b' > 1} \Pr\left(Z_b^n(m_b'|1) = Z_b^n(1|1)\right)$$
$$+ \sum_{m_b' > 1} \Pr\left(\begin{array}{c} \mathsf{Bin}_b(Z_b^n(m_b'|1)) \\ = \mathsf{Bin}_b(Z_b^n(1|1)) \end{array}\Big| Z_b^n(m_b'|1) \ne Z_b^n(1|1)\right)$$
$$\overset{(ii)}{\le} 2^{nR'} \cdot 2^{-n(H(Z|X_r)-\delta(\epsilon))} + 2^{nR'} \cdot 2^{-n\widetilde{R}},$$

where

- (i) follows since $L_b = \mathsf{Bin}_b(Z_b^n(1|1))$,
- the first term in (ii) is obtained because for $m_b' \ne 1$, the codewords $Z_b^n(m_b'|1)$ and $Z_b^n(1|1)$ are generated independently according to $\prod_{i=1}^n p_{Z|X_r}(\cdot|x_{rbi}(l_{b-1}))$, 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 ϵ for which $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. Hence, we get that

$$\Pr\left(\hat{M}_b'(1, L_b) \ne 1\right) \to 0 \text{ as } n \to \infty,$$

if the following two constraints are satisfied:

$$R' < \widetilde{R}, \quad (7)$$
$$R' < H(Z|X_r) - \delta(\epsilon). \quad (8)$$

Event (b): $(\hat{L}_{b-1}, \hat{M}_b'') \ne (1, 1)$ given $\hat{M}_b'(1, L_b) = 1$: The probability of this event is upper bounded by

$$\Pr\left(\begin{array}{c} (6) \text{ not true for} \\ (l_{b-1}, m_b'') = (1, 1) \end{array}\Big| \hat{M}_b'(1, L_b) = 1\right)$$
$$+ \Pr\left(\begin{array}{c} (6) \text{ true for some} \\ (l_{b-1}, m_b'') \ne (1, 1) \end{array}\Big| \hat{M}_b'(1, L_b) = 1\right).$$

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 [30] as follows:

$$\Pr\left(\begin{array}{c} (6) \text{ true for some} \\ (l_{b-1}, m_b'') \ne (1, 1) \end{array}\Big| \hat{M}_b'(1, L_b) = 1\right)$$
$$\le \sum_{\substack{l_{b-1}=1, \\ m_b''>1}} \Pr\left(\left(X_{rb}^n(1), Z_b^n(1|1), X_b^n(\hat{m}_b''|1, 1), Y_b^n\right) \in \mathcal{T}_\epsilon^{(n)}\right)$$

$$+ \sum_{\substack{l_{b-1} > 1, \\ m_b'' \geq 1}} \Pr \Bigg(\Big(X_{rb}^n(\hat{l}_{b-1}), Z_b^n(\hat{m}_b'(\hat{l}_{b-1}, L_b) | \hat{l}_{b-1}),$$

$$X_b^n(\hat{m}_b'' | \hat{l}_{b-1}, \hat{m}_b'(\hat{l}_{b-1}, L_b)), Y_b^n \Big) \in \mathcal{T}_\epsilon^{(n)} \Bigg)$$

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

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

$$\Pr \Big((\hat{L}_{b-1}, \hat{M}_b'') \neq (1, 1) \mid \hat{M}_b'(1, L_b) = 1 \Big) \to 0, \quad \text{as } n \to \infty,$$

if

$$R'' < I(X; Y | X_r, Z) - \delta(\epsilon), \tag{9}$$
$$\widetilde{R} + R'' < I(X, X_r; Y) - \delta(\epsilon). \tag{10}$$

Performing Fourier-Motzkin elimination of $\widetilde{R}$, 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 $\widetilde{R}$ to be $R' + \delta(\epsilon)$. Since this theorem is already known, we do not repeat the arguments of the converse 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 $p_{X_r}(x_r) p_{X|X_r,S}(x|x_r, s)$ 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.

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

- *Cooperation Codewords:*
  Generate $2^{n\widetilde{R}}$ codewords $x_{rb}^n(l_{b-1})$, i.i.d. according to p_{X_r}, where $l_{b-1} \in [1 : 2^{n\widetilde{R}}]$.
- *Cribbed Codewords:*
  For each l_{b-1} and each $s \in \mathcal{S}$, generate a codebook of $2^{nR'}$ codewords. The ith symbol of such a codeword is chosen independently according to
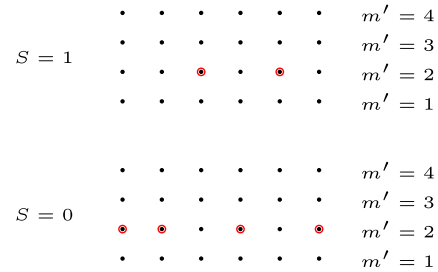

Fig. 7. The figure depicts the cribbed codewords generated for encoding m' for a given $x_r^n(l)$. Each node corresponds to a z symbol that is generated independently according to $p_{Z|X_r,S}(\cdot|x_{r,i}(l), s)$. The red circles show how encoder 1 chooses the codeword if it wants to transmit $m' = 2$ and observes $s^n = (0, 0, 1, 0, 1, 0)$. This construction is not identical but equivalent to that described in [30, Sec. 7.4.1].

$p_{Z|X_r,S}(\cdot|x_{rbi}(l_{b-1}), s)$. The result of this is that for each l_{b-1}, each $m_b' \in [1 : 2^{nR'}]$ and each $s_b^n = (s_{b1}, s_{b2}, \ldots, s_{bn})$, the source encoder can form an effective codeword $z_b^n(m_b'|l_{b-1}, s_b^n)$, whose ith symbol can be causally chosen as the ith 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 $s \in \mathcal{S}$, generate a codebook of $2^{nR''}$ codewords. The ith symbol of such a codeword is generated independently according to $p_{X|X_r,Z,S}(\cdot|x_{rbi}(l_{b-1}), z_{bi}(m_b'|l_{b-1}, s), s)$. 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_b^n, the source encoder can form an effective codeword $x_b^n(m_b''|l_{b-1}, m_b', s_b^n)$, whose ith symbol can be causally chosen as the ith 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 $\mathcal{Z}^n$ into $2^{n\widetilde{R}}$ 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 $\text{bin}_b(z^n)$.

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 $\frac{B-1}{B} R$, 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 $x_b^n(m_b''|l_{b-1}, m_b', s_b^n)$. This operation is valid because it does not require noncausal knowledge of the state sequences, it can be done "on the fly". The relay transmits $x_{rb}^n(l_{b-1})$. 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 $z_b^n(m_b'|l_{b-1}, s_b^n)$. The source and the relay set l_b to be the index of the bin containing $z_b^n(m_b'|l_{b-1}, s_b^n)$.

From the encoding operation described above, we can see that the label l_b depends on (l_{b-1}, m_b', s_b^n). 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_b^n codeword to agree on the u_{b+1}^n 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 $l_{b-1} \in [1 : 2^{n\tilde{R}}]$, finds the unique m'_b such that

$$\text{bin}_b(z_b^n(m'_b|l_{b-1}, s_b^n)) = l_b.$$

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_b^n. We will signify this explicitly by denoting the chosen message as $\hat{m}'_b(l_{b-1}, s_b^n, l_b)$.

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

$$\left(x_{rb}^n(\hat{l}_{b-1}),\ z_b^n(\hat{m}'_b(\hat{l}_{b-1}, s_b^n, l_b)|\hat{l}_{b-1}, s_b^n), \right.$$
$$\left. x_b^n(\hat{m}''_b|\hat{l}_{b-1}, \hat{m}'_b(\hat{l}_{b-1}, s_b^n, l_b), s_b^n),\ s_b^n, y_b^n \right) \in \mathcal{T}_\epsilon^{(n)} \quad (11)$$

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 s_b^n sequence. The total number of z_b^n codewords constructed by the source encoder considering only the typical s_b^n sequences is therefore $\approx 2^{nH(S)} \cdot 2^{nR'} \approx 2^{nH(S,Z|X_r)}$. However, these codewords cannot be distinct since there are only $\approx 2^{nH(Z|X_r)}$ distinct typical sequences z_b^n (conditioned on l_{b-1}). This implies that multiple (s_b^n, m'_b) pairs will be mapped to the same codeword z_b^n and therefore, the relay will be 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

$$L_b = \text{Bin}_b(Z_b^n(1|1, S_b^n)).$$

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

(a) $\hat{M}'_b(1, S_b^n, L_b) \neq 1$,

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

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

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

$$\Pr\left(\hat{M}'_b(1, S_b^n, L_{1,b}) \neq 1 \right)$$
$$= \Pr\left(\text{Bin}_b(Z_b^n(m'_b|1, S_b^n)) = L_b \text{ for some } m'_b > 1 \right)$$
$$= \Pr\left(\begin{array}{c} \text{Bin}_b(Z_b^n(m'_b|1, S_b^n)) \\ = \text{Bin}_b(Z_b^n(1|1, S_b^n)) \end{array} \text{ for some } m'_b > 1 \right)$$
$$\leq \sum_{m'_b > 1} \Pr\left(\text{Bin}_b(Z_b^n(m'_b|1, S_b^n)) = \text{Bin}_b(Z_b^n(1|1, S_b^n)) \right)$$
$$= \sum_{m'_b > 1} \Pr\left(\begin{array}{c} \text{Bin}_b(Z_b^n(m'_b|1, S_b^n)) = \text{Bin}_b(Z_b^n(1|1, S_b^n)), \\ Z_b^n(m'_b|1, S_b^n) = Z_b^n(1|1, S_b^n) \end{array} \right)$$
$$+ \sum_{m'_b > 1} \Pr\left(\begin{array}{c} \text{Bin}_b(Z_b^n(m'_b|1, S_b^n)) = \text{Bin}_b(Z_b^n(1|1, S_b^n)), \\ Z_b^n(m'_b|1, S_b^n) \neq Z_b^n(1|1, S_b^n) \end{array} \right)$$
$$\leq \sum_{m'_b > 1} \Pr\left(Z_b^n(m'_b|1, S_b^n) = Z_b^n(1|1, S_b^n) \right)$$
$$+ \sum_{m'_b > 1} \Pr\left(\begin{array}{c} \text{Bin}_b(Z_b^n(m'_b|1, S_b^n)) \\ = \text{Bin}_b(Z_b^n(1|1, S_b^n)) \end{array} \left| \begin{array}{c} Z_b^n(m'_b|1, S_b^n) \\ \neq Z_b^n(1|1, S_b^n) \end{array} \right. \right)$$
$$\leq 2^{nR'} \cdot 2^{-n(H(Z|X_r, S) - \delta(\epsilon))} + 2^{nR'} \cdot 2^{-n\tilde{R}},$$

where we use $\delta(\epsilon)$ to denote any function of ϵ 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:

$$R' < \tilde{R}, \quad (12)$$
$$R' < H(Z|X_r, S) - \delta(\epsilon). \quad (13)$$

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

$$\Pr\left(\begin{array}{c} (11) \text{ not true for} \\ (l_{b-1}, m''_b) = (1, 1) \end{array} \left| \hat{M}'_b(1, S_b^n, L_b) = 1 \right. \right)$$
$$+ \Pr\left(\begin{array}{c} (11) \text{ true for some} \\ (l_{b-1}, m''_b) \neq (1, 1) \end{array} \left| \hat{M}'_b(1, S_b^n, L_b) = 1 \right. \right).$$

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 [30] as follows:

$$\Pr\left(\begin{array}{c} (11) \text{ true for some} \\ (l_{b-1}, m''_b) \neq (1, 1) \end{array} \left| \hat{M}'_b(1, S_b^n, L_b) = 1 \right. \right)$$
$$\leq \sum_{\substack{l_{b-1}=1, \\ m''_b > 1}} \Pr\left(\left(\begin{array}{c} X_{rb}^n(1), Z_b^n(1|1, S_b^n), \\ X_b^n(m''_b|1, 1, S_b^n), S_b^n, Y_b^n \end{array} \right) \in \mathcal{T}_\epsilon^{(n)} \right)$$
$$+ \sum_{\substack{l_{b-1}>1, \\ m''_b \geq 1}} \Pr\left(\left(X_{rb}^n(\hat{l}_{b-1}),\ Z_b^n(\hat{m}'_b(\hat{l}_{b-1}, S_b^n, L_b)|\hat{l}_{b-1}, S_b^n), \right.\right.$$
$$\left.\left. X_b^n(\hat{m}''_b|\hat{l}_{b-1}, \hat{m}'_b(\hat{l}_{b-1}, S_b^n, l_b), S_b^n),\ S_b^n, Y_b^n \right) \in \mathcal{T}_\epsilon^{(n)} \right)$$
$$\leq 2^{nR''} 2^{-n(I(X;Y|X_r, Z, S) - \delta(\epsilon))}$$
$$+ 2^{n(\tilde{R}+R'')} 2^{-n(I(X, X_r, Z; Y|S) - \delta(\epsilon))},$$

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 $2^{n(\widetilde{R}+R'')}$ events, where each corresponds to a different value of (l_{b-1}, m_b'').

Thus, we get that

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

if

$$R'' < I(X; Y|X_r, Z, S) - \delta(\epsilon), \tag{14}$$
$$\widetilde{R} + R'' < I(X, X_r; Y|S) - \delta(\epsilon). \tag{15}$$

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:

$$
\begin{aligned}
nR &= H(M) \\
&\overset{(a)}{=} H(M|S^n) \\
&\overset{(b)}{=} H(M, Z^n|S^n) \\
&= H(Z^n|S^n) + H(M|Z^n, S^n) \\
&\overset{(c)}{\leq} H(Z^n|S^n) + I(M; Y^n|Z^n, S^n) + n\epsilon_n \\
&\overset{(d)}{=} \sum_{i=1}^n H(Z_i|S^n, Z^{i-1}) + \sum_{i=1}^n I(M, X_i; Y_i|Y^{i-1}, Z^n, S^n) \\
&\qquad + n\epsilon_n \\
&\overset{(e)}{=} \sum_{i=1}^n H(Z_i|S^n, Z^{i-1}, X_{r,i}) \\
&\qquad + \sum_{i=1}^n I(M, X_i; Y_i|Y^{i-1}, Z^n, S^n, X_r^n) + n\epsilon_n \\
&\overset{(f)}{\leq} \sum_{i=1}^n H(Z_i|S_i, X_{r,i}) + \sum_{i=1}^n I(X_i; Y_i|Z_i, S_i, X_{r,i}) + n\epsilon_n \\
&= nH(Z_Q|S_Q, X_{rQ}, Q) + nI(X_Q; Y_Q|Z_Q, S_Q, X_{rQ}, Q) \\
&\qquad + n\epsilon_n \\
&\leq nH(Z_Q|S_Q, X_{rQ}) + nI(X_Q; Y_Q|Z_Q, S_Q, X_{rQ}) + n\epsilon_n,
\end{aligned}
$$

where

- Q is a random variable uniformly distributed over $[1 : n]$, independent of (X^n, X_r^n, 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 $Q - (X_Q, X_{rQ}, S_Q) - Y_Q$.

The second bound on R is proved as follows:

$$
\begin{aligned}
nR &= H(M) \\
&= H(M|S^n) \\
&\leq I(M; Y^n|S^n) + n\epsilon_n \\
&= I(M, X^n, X_r^n; Y^n|S^n) + n\epsilon_n \\
&\leq \sum_{i=1}^n I(X_i, X_{r,i}; Y_i|S_i) + n\epsilon_n \\
&= nI(X_Q, X_{rQ}; Y_Q|S_Q, Q) + n\epsilon_n \\
&\leq nI(X_Q, X_{rQ}; Y_Q|S_Q) + n\epsilon_n.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
R \leq \min\big(& I(X_Q, X_{rQ}; Y_Q|S_Q), \\
& H(Z_Q|S_Q, X_{rQ}) + I(X_Q; Y_Q|Z_Q, S_Q, X_{rQ})\big) + \epsilon_n
\end{aligned}
$$

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_{rQ};
- we have that $p_{Y_Q|X_Q, X_{rQ}, S_Q}(y|x, x_r, s)$ 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_{rQ}, S_Q), and
- similarly, we also have $Z_Q = z(X_Q, X_{rQ}, S_Q)$.

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

$$
\begin{aligned}
& p_{S_Q, X_Q, X_{rQ}, Y_Q}(s, x, x_r, y) \\
&= p_{S_Q}(s) p_{X_{rQ}}(x_r) p_{X_Q|X_{rQ}, S_Q}(x|x_r, s) \\
&\quad \times p_{Y_Q|X_Q, X_{rQ}, S_Q}(y|x, x_r, s) \\
&= p_S(s) p_{X_{rQ}}(x_r) p_{X_Q|X_{rQ}, S_Q}(x|x_r, s) p_{Y|X, X_r, S}(y|x, x_r, s).
\end{aligned}
$$

So, we can define the random variables $X \triangleq X_Q$, $X_r \triangleq X_{rQ}$, $S \triangleq S_Q$, $Z \triangleq Z_Q$ 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 $p_S(s) p_{X_r}(x_r) p_{X|X_r, S}(x|x_r, s) p_{Y|X, X_r, S}(y|x, x_r, s)$ and $Z = z(X, X_r, S)$. Since $\epsilon_n \to 0$ as $n \to \infty$, 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.

Proof: 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 $2^{n\tilde{R}}$ codewords $u_b^n(l_{b-1})$, 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 $s \in \mathcal{S}$, generate a codebook of $2^{nR'}$ codewords. The ith 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 $m_b' \in [1 : 2^{nR'}]$ and each $s_b^n = (s_{b1}, s_{b2}, \ldots, s_{bn})$, the source encoder can form an effective codeword $z_b^n(m_b'|l_{b-1}, s_b^n)$, whose ith symbol can be causally chosen as the ith symbol of the m_b'-th codeword from the codebook corresponding to l_{b-1} and s_{bi}.

- *Transmission Codewords:*
  For each l_{b-1}, each $m_b' \in [1 : 2^{nR'}]$ and each $s \in \mathcal{S}$, generate a codebook of $2^{nR''}$ codewords. The ith symbol of such a codeword is generated independently according to $p_{X|U,Z,S}(\cdot|u_{bi}(l_{b-1}), z_{bi}(m_b'|l_{b-1}, s), s)$. 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_b^n, the source encoder can form an effective codeword $x_b^n(m_b''|l_{b-1}, m_b', s_b^n)$, whose ith symbol can be causally chosen as the ith 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 $\mathcal{Z}^n$ 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 $\mathsf{bin}_b(z^n)$.

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 $\frac{B-1}{B}R$, 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 $x_b^n(m_b''|l_{b-1}, m_b', s_b^n)$. The relay transmits x_{rb}^n, the ith symbol of which is obtained as $x_r(u_{bi}(l_{b-1}), z_{bi}(m_b'|l_{b-1}, s_b^n))$. At the end of block b, the source and the relay set l_b to be the index of the bin containing $z_b^n(m_b'|l_{b-1}, s_b^n)$.

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' such that

$$\mathsf{bin}_b(z_b^n(m_b'|l_{b-1}, s_b^n)) = l_b.$$

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_b^n. We will signify this explicitly by denoting the chosen message as $\hat{m}_b'(l_{b-1}, s_b^n, l_b)$.

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

$$\left(u_b^n(\hat{l}_{b-1}), z_b^n(\hat{m}_b'(\hat{l}_{b-1}, s_b^n, l_b)|\hat{l}_{b-1}, s_b^n), \right.$$
$$\left. x_b^n(\hat{m}_b''|\hat{l}_{b-1}, \hat{m}_b'(\hat{l}_{b-1}, s_b^n, l_b), s_b^n), s_b^n, y_b^n\right) \in \mathcal{T}_\epsilon^{(n)}.$$

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

$$\begin{aligned} R' &< \tilde{R}, \\ R' &< H(Z|U, S) - \delta(\epsilon), \\ R'' &< I(X; Y|U, Z, S) - \delta(\epsilon), \\ \tilde{R} + R'' &< I(U, X; Y|S) - \delta(\epsilon). \end{aligned}$$

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 $U_i \triangleq (Z^{i-1}, S^{i-1})$. 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,

$$\begin{aligned} nR &= H(M|S^n) \\ &= H(M, Z^n|S^n) \\ &\leq H(Z^n|S^n) + I(M; Y^n|S^n, Z^n) + n\epsilon_n \\ &\leq \sum_{i=1}^n H(Z_i|Z^{i-1}, S^{i-1}, S_i) \\ &\quad + \sum_{i=1}^n I(X_i; Y_i|Z^{i-1}, S^{i-1}, Z_i, S_i) + n\epsilon_n, \end{aligned}$$

where the final step uses the fact that Y_i is independent of other random variables given X_i, Z^{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

$$\begin{aligned} nR &\leq \sum_{i=1}^n H(Z_i|U_i, S_i) + \sum_{i=1}^n I(X_i; Y_i|U_i, Z_i, S_i) + n\epsilon_n \\ &= nH(Z|U_Q, S_Q, Q) + nI(X_Q; Y_Q|U_Q, Z_Q, S_Q, Q) + n\epsilon_n, \end{aligned}$$

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:

$$\begin{aligned} nR &= H(M|S^n) \\ &\leq I(M; Y^n|S^n) + n\epsilon_n \\ &= \sum_{i=1}^n I(M; Y_i|Y^{i-1}, S^n) + n\epsilon_n \\ &\overset{(a)}{=} \sum_{i=1}^n I(M, X_i, X_{r,i}, Z^{i-1}, S^{i-1}; Y_i|Y^{i-1}, S^n) + n\epsilon_n \\ &\leq \sum_{i=1}^n I(X_i, X_{r,i}, Z^{i-1}, S^{i-1}; Y_i|S_i) + n\epsilon_n \end{aligned}$$

$$= \sum_{i=1}^{n} I(U_i, X_i, X_{r,i}; Y_i|S_i) + n\epsilon_n$$

$$= \sum_{i=1}^{n} I(U_i, X_i; Y_i|S_i) + n\epsilon_n$$

$$= nI(U_Q, X_Q; Y_Q|S_Q, Q) + n\epsilon_n$$

$$\leq nI(Q, U_Q, X_Q; Y_Q|S_Q) + n\epsilon_n,$$

where step (a) is true since $(X_i, X_{r,i}, Z^{i-1}, S^{i-1})$ 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 $U \triangleq (Q, U_Q)$, $X \triangleq X_Q$, $X_r \triangleq X_{rQ}$, $S \triangleq S_Q$, $Z \triangleq Z_Q$ 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 $p_S(s)p_U(u)p_{X|U,S}(x|u, s)p_{Y|X,X_r,S}(y|x, x_r, s)$, $Z = z(X, S)$ 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 [30, 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 is 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_2^n, 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_1^n signal received by encoder 2, and thus encoder 1 also knows z_1^n. Similarly, encoder 2 also knows z_1^n and z_2^n. 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 $p_U(u)p_{X_1|U,S_1}(x_1|u, s_1)p_{X_2|U,S_2}(x_2|u, s_2)$ 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 [4] 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 based 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 $2^{n(\tilde{R}_1+\tilde{R}_2)}$ codewords $u_b^n(l_{1,b-1}, l_{2,b-1})$, i.i.d. according to p_U, where $l_{1,b-1} \in [1 : 2^{n\tilde{R}_1}]$ and $l_{2,b-1} \in [1 : 2^{n\tilde{R}_2}]$. 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 $s_1 \in \mathcal{S}_1$, generate a codebook of $2^{nR_1'}$ codewords. The ith symbol of such a codeword is chosen independently according to $p_{Z_1|U,S_1}(\cdot|u_{bi}(l_{b-1}), s_1)$. The result of this is that for each l_{b-1}, each $m_{1,b}' \in [1 : 2^{nR_1'}]$ and each $s_{1b}^n = (s_{1b1}, s_{1b2}, \ldots, s_{1bn})$, encoder 1 can form an effective codeword $z_{1b}^n(m_{1,b}'|l_{b-1}, s_{1b}^n)$, whose ith symbol can be causally chosen as the ith symbol of the $m_{1,b}'$-th codeword from the codebook corresponding to l_{b-1} and s_{1bi}.

- *Cribbed Codewords - II:*
  Similarly, for each l_{b-1} and each $s_2 \in \mathcal{S}_2$, generate a codebook of $2^{nR_2'}$ codewords. The ith symbol of such a codeword is chosen independently according to $p_{Z_2|U,S_2}(\cdot|u_{bi}(l_{b-1}), s_2)$. The result of this is that for each l_{b-1}, each $m_{2,b}' \in [1 : 2^{nR_2'}]$ and each $s_{2b}^n = (s_{2b1}, s_{2b2}, \ldots, s_{2bn})$, encoder 2 can form an effective codeword $z_{2b}^n(m_{2,b}'|l_{b-1}, s_{2b}^n)$, whose ith symbol can be causally chosen as the ith 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 $s_1 \in \mathcal{S}_1$, generate a codebook of $2^{nR_1''}$ codewords. The ith symbol of such a codeword is generated independently according to $p_{X_1|U,Z_1,S_1}(\cdot|u_{bi}(l_{b-1}), z_{1bi}(m_{1,b}'|l_{b-1}, s_1), s_1)$. The result of this construction is that for each l_{b-1}, each $m_{1,b}' \in [1 : 2^{nR_1'}]$, each $m_{1,b}'' \in [1 : 2^{nR_1''}]$ and each s_{1b}^n, encoder 1 can form an effective codeword $x_{1b}^n(m_{1,b}''|l_{b-1}, m_{1,b}', s_{1b}^n)$, whose ith symbol can be causally chosen as the ith 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 $s_2 \in \mathcal{S}_2$, generate a codebook of $2^{nR_2''}$ codewords. The ith symbol of such a codeword is generated independently according to $p_{X_2|U,Z_2,S_2}(\cdot|u_{bi}(l_{b-1}), z_{2bi}(m_{2,b}'|l_{b-1}, s_2), s_2)$. The result of this construction is that for each l_{b-1}, each $m_{2,b}' \in [1 : 2^{nR_2'}]$, each $m_{2,b}'' \in [1 : 2^{nR_2''}]$ and each s_{2b}^n, encoder 2 can form an effective codeword $x_{2b}^n(m_{2,b}''|l_{b-1}, m_{2,b}', s_{2b}^n)$, whose ith symbol can be causally chosen as the ith 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 $\mathcal{Z}_1^n$ into $2^{n\widetilde{R}_1}$ bins, by choosing a bin for each z_1^n independently and uniformly at random from $[1 : 2^{n\widetilde{R}_1}]$. Denote the chosen bin for z_1^n by $\text{bin}_b(z_1^n)$. Similarly, partition the set $\mathcal{Z}_2^n$ into $2^{n\widetilde{R}_2}$ bins, by choosing a bin for each z_2^n independently and uniformly at random from $[1 : 2^{n\widetilde{R}_2}]$. Denote the chosen bin for z_2^n by $\text{bin}_b(z_2^n)$.

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 $\left(\frac{B-1}{B} R_1, \frac{B-1}{B} R_2\right)$, 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 $l_{b-1} = (l_{1,b-1}, l_{2,b-1})$ 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

$$l_{1,b} = \text{bin}_b(z_{1b}^n), \quad \text{and} \quad l_{2,b} = \text{bin}_b(z_{2b}^n).$$

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

(1) 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

$$\text{bin}_b(z_{1b}^n(m'_{1,b}|l_{b-1}, s_{1b}^n)) = l_{1,b},$$

and

$$\text{bin}_b(z_{2b}^n(m'_{2,b}|l_{b-1}, s_{2b}^n)) = l_{2,b}.$$

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 $(l_b, s_{1b}^n, s_{2b}^n)$. 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 $(\hat{l}_{b-1}, \hat{m}''_{1,b}, \hat{m}''_{2,b})$ such that (16) (appearing at the bottom 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.

$$(L_{b-1}, M'_{1,b}, M'_{2,b}, M''_{1,b}, M''_{2,b}) = (1, 1, 1, 1, 1).$$

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

$$L_{1,b} = \text{Bin}_b(Z_{1b}^n(1|1, S_{1b}^n))$$

and

$$L_{2,b} = \text{Bin}_b(Z_{2b}^n(1|1, S_{2b}^n)).$$

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

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

$$\Pr\left(\hat{M}'_{1,b}(1, S_{1b}^n, L_{1,b}) \neq 1\right)$$
$$= \Pr\left(\text{Bin}_b(Z_{1b}^n(m'_{1,b}|1, S_{1b}^n)) = L_{1,b} \text{ for some } m'_{1,b} > 1\right)$$
$$= \Pr\begin{pmatrix}\text{Bin}_b(Z_{1b}^n(m'_{1,b}|1, S_{1b}^n)) \\ = \text{Bin}_b(Z_{1b}^n(1|1, S_{1b}^n)) \text{ for some } m'_{1,b} > 1\end{pmatrix}$$
$$\leq \sum_{m'_{1,b} > 1} \Pr\left(\text{Bin}_b(Z_{1b}^n(m'_{1,b}|1, S_{1b}^n)) = \text{Bin}_b(Z_{1b}^n(1|1, S_{1b}^n))\right)$$
$$= \sum_{m'_{1,b} > 1} \Pr\begin{pmatrix}\text{Bin}_b(Z_{1b}^n(m'_{1,b}|1, S_{1b}^n)) \\ = \text{Bin}_b(Z_{1b}^n(1|1, S_{1b}^n)), \\ Z_{1b}^n(m'_{1,b}|1, S_{1b}^n) = Z_{1b}^n(1|1, S_{1b}^n)\end{pmatrix}$$
$$+ \sum_{m'_{1,b} > 1} \Pr\begin{pmatrix}\text{Bin}_b(Z_{1b}^n(m'_{1,b}|1, S_{1b}^n)) \\ = \text{Bin}_b(Z_{1b}^n(1|1, S_{1b}^n)), \\ Z_{1b}^n(m'_{1,b}|1, S_{1b}^n) \neq Z_{1b}^n(1|1, S_{1b}^n)\end{pmatrix}$$
$$\leq \sum_{m'_{1,b} > 1} \Pr\left(Z_{1b}^n(m'_{1,b}|1, S_{1b}^n) = Z_{1b}^n(1|1, S_{1b}^n)\right)$$
$$+ \sum_{m'_{1,b} > 1} \Pr\begin{pmatrix}\text{Bin}_b(Z_{1b}^n(m'_{1,b}|1, S_{1b}^n)) \,\Big|\, Z_{1b}^n(m'_{1,b}|1, S_{1b}^n) \\ = \text{Bin}_b(Z_{1b}^n(1|1, S_{1b}^n)) \,\Big|\, \neq Z_{1b}^n(1|1, S_{1b}^n)\end{pmatrix}$$
$$\leq 2^{nR'_1} \cdot 2^{-n(H(Z_1|U,S_1)-\delta(\epsilon))} + 2^{nR'_1} \cdot 2^{-n\widetilde{R}_1},$$

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

$$\Pr\left(\hat{M}'_{1,b}(1, S_{1b}^n, L_{1,b}) \neq 1\right) \to 0 \text{ as } n \to \infty,$$

$$\begin{aligned}\Big(u_b^n(\hat{l}_{b-1}), \ & z_{1b}^n(\hat{m}'_{1,b}(\hat{l}_{b-1}, s_{1b}^n, l_{1,b}) \,|\, \hat{l}_{b-1}, s_{1b}^n), \ z_{2b}^n(\hat{m}'_{2,b}(\hat{l}_{b-1}, s_{2b}^n, l_{2,b}) \,|\, \hat{l}_{b-1}, s_{2b}^n), \\ & x_{1b}^n(\hat{m}''_{1,b} \,|\, \hat{l}_{b-1}, \hat{m}'_{1,b}(\hat{l}_{b-1}, s_{1b}^n, l_{1,b}), s_{1b}^n), \ x_{2b}^n(\hat{m}''_{2,b} \,|\, \hat{l}_{b-1}, \hat{m}'_{2,b}(\hat{l}_{b-1}, s_{2b}^n, l_{2,b}), s_{2b}^n), \ s_{1b}^n, s_{2b}^n, y_b^n\Big) \in \mathcal{T}_\epsilon^{(n)}\end{aligned} \quad (16)$$

if the following two constraints are satisfied:

$$R_1' < \widetilde{R}_1,$$
$$R_1' < H(Z_1|U, S_1) - \delta(\epsilon).$$

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

$$R_2' < \widetilde{R}_2,$$
$$R_2' < H(Z_2|U, S_2) - \delta(\epsilon).$$

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 the sum of

$$\Pr\left(\begin{array}{c|c} (16) \text{ not true for} & \hat{M}_{1,b}'(1, S_{1b}^n, L_{1,b}) = 1, \\ (l_{b-1}, m_{1,b}'', m_{2,b}'') = (1, 1, 1) & \hat{M}_{2,b}'(1, S_{2b}^n, L_{2,b}) = 1 \end{array}\right)$$

and

$$\Pr\left(\begin{array}{c|c} (16) \text{ true for some} & \hat{M}_{1,b}'(1, S_{1b}^n, L_{1,b}) = 1, \\ (l_{b-1}, m_{1,b}'', m_{2,b}'') \neq (1, 1, 1) & \hat{M}_{2,b}'(1, S_{2b}^n, L_{2,b}) = 1 \end{array}\right).$$

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 [30] 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:

$$R_2'' < I(X_2; Y|U, Z_2, X_1, S_1, S_2) - \delta(\epsilon),$$
$$R_1'' < I(X_1; Y|U, Z_1, X_2, S_1, S_2) - \delta(\epsilon),$$
$$R_1'' + R_2'' < I(X_1, X_2; Y|U, Z_1, Z_2, S_1, S_2) - \delta(\epsilon).$$

Applying the packing lemma gives us the following condition for vanishing probability of the fourth event,

$$\begin{aligned} \widetilde{R}_1 + \widetilde{R}_2 & \\ + R_1'' + R_2'' & < I(U, Z_1, Z_2, X_1, X_2; Y|S_1, S_2) - \delta(\epsilon) \\ & = I(X_1, X_2; Y|S_1, S_2) - \delta(\epsilon). \end{aligned}$$

Collecting all the constraints established so far, we have

$$R_1' < \widetilde{R}_1, \tag{17}$$
$$R_1' < H(Z_1|U, S_1) - \delta(\epsilon), \tag{18}$$
$$R_2' < \widetilde{R}_2, \tag{19}$$
$$R_2' < H(Z_2|U, S_2) - \delta(\epsilon), \tag{20}$$
$$R_2'' < I(X_2; Y|U, Z_2, X_1, S_1, S_2) - \delta(\epsilon), \tag{21}$$

$$R_1'' < I(X_1; Y|U, Z_1, X_2, S_1, S_2) - \delta(\epsilon), \tag{22}$$
$$R_1'' + R_2'' < I(X_1, X_2; Y|U, Z_1, Z_2, S_1, S_2) - \delta(\epsilon), \tag{23}$$
$$\begin{aligned} \widetilde{R}_1 + \widetilde{R}_2 & \\ + R_1'' + R_2'' & < I(X_1, X_2; Y|S_1, S_2) - \delta(\epsilon). \end{aligned} \tag{24}$$

Performing Fourier-Motzkin elimination of $\widetilde{R}_1$, $\widetilde{R}_2$, 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:

$$H(M_1, M_2|Y^n, S_1^n, S_2^n) \leq n\epsilon_n,$$

where $\epsilon_n \to 0$ as $n \to \infty$. Define U_i as

$$U_i \triangleq (Z_1^{i-1}, Z_2^{i-1}, S_1^{i-1}, S_2^{i-1}).$$

An upper bound on R_1 is established by the following:

$$\begin{aligned} nR_1 &= H(M_1) \\ &\stackrel{(a)}{=} H(M_1|M_2, S_1^n, S_2^n) \\ &\stackrel{(b)}{=} H(M_1, Z_1^n|M_2, S_1^n, S_2^n) \\ &= H(Z_1^n|M_2, S_1^n, S_2^n) + H(M_1|Z_1^n, M_2, S_1^n, S_2^n) \\ &\stackrel{(c)}{\leq} H(Z_1^n|M_2, S_1^n, S_2^n) + I(M_1; Y^n|Z_1^n, M_2, S_1^n, S_2^n) + n\epsilon_n \\ &= \sum_{i=1}^n H(Z_{1i}|Z_1^{i-1}, M_2, S_1^n, S_2^n) \\ &\quad + \sum_{i=1}^n I(M_1; Y_i|Y^{i-1}, Z_1^n, M_2, S_1^n, S_2^n) + n\epsilon_n \\ &\stackrel{(d)}{=} \sum_{i=1}^n H(Z_{1i}|Z_1^{i-1}, M_2, S_1^n, S_2^n, Z_2^{i-1}) \\ &\quad + \sum_{i=1}^n I(M_1; Y_i|Y^{i-1}, Z_1^n, M_2, S_1^n, S_2^n, X_2^n) + n\epsilon_n \\ &\stackrel{(e)}{=} \sum_{i=1}^n H(Z_{1i}|Z_1^{i-1}, M_2, S_1^n, S_2^n, Z_2^{i-1}) \\ &\quad + \sum_{i=1}^n I(M_1, X_{1i}; Y_i|Y^{i-1}, Z_1^n, M_2, S_1^n, S_2^n, X_2^n) + n\epsilon_n \\ &\stackrel{(f)}{\leq} \sum_{i=1}^n H(Z_{1i}|Z_1^{i-1}, S_1^i, S_2^{i-1}, Z_2^{i-1}) \\ &\quad + \sum_{i=1}^n I(X_{1i}; Y_i|Z_1^i, S_1^i, S_2^i, X_{2i}, Z_2^{i-1}) + n\epsilon_n \\ &= \sum_{i=1}^n H(Z_{1i}|U_i, S_{1i}) \\ &\quad + \sum_{i=1}^n I(X_{1i}; Y_i|U_i, Z_{1i}, S_{1i}, S_{2i}, X_{2i}) + n\epsilon_n \\ &= nH(Z_{1Q}|U_Q, S_{1Q}, Q) \\ &\quad + nI(X_{1Q}; Y_Q|U_Q, Z_{1Q}, S_{1Q}, S_{2Q}, X_{2Q}, Q) + n\epsilon_n \end{aligned}$$

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_1^n, S_2^n),
- (b) follows since Z_1^n is a function of (M_1, M_2, S_1^n, S_2^n),
- (c) follows by Fano's inequality,
- (d) follows since (X_{2i}, Z_{2i}) is a function of (M_2, S_2^n, Z_1^{i-1}),
- (e) follows since X_{1i} is a function of (M_1, S_1^n, X_2^{i-1}),
- (f) follows since (i) conditioning reduces entropy and (ii) conditioned on $(X_{1i}, X_{2i}, S_{1i}, S_{2i})$, Y_i is independent of $(M_1, M_2, X_1^n, X_2^n, S_1^n, S_2^n)$.

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:

$$n(R_1 + R_2)$$
$$= H(M_1, M_2)$$
$$= H(M_1, M_2|S_1^n, S_2^n)$$
$$= H(M_1, M_2, Z_1^n, Z_2^n|S_1^n, S_2^n)$$
$$= H(Z_1^n, Z_2^n|S_1^n, S_2^n) + H(M_1, M_2|Z_1^n, Z_2^n, S_1^n, S_2^n)$$
$$\leq \sum_{i=1}^{n} H(Z_{1i}, Z_{2i}|U_i, S_{1i}, S_{2i}) + H(M_1, M_2|Z_1^n, Z_2^n, S_1^n, S_2^n)$$
$$\leq \sum_{i=1}^{n} H(Z_{1i}, Z_{2i}|U_i, S_{1i}, S_{2i})$$
$$+ I(M_1, M_2; Y^n|Z_1^n, Z_2^n, S_1^n, S_2^n) + n\epsilon_n$$
$$\leq \sum_{i=1}^{n} H(Z_{1i}, Z_{2i}|U_i, S_{1i}, S_{2i})$$
$$+ \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_i|U_i, Z_{1i}, Z_{2i}, S_{1i}, S_{2i}) + n\epsilon_n$$
$$= nH(Z_{1Q}, Z_{2Q}|U_Q, S_{1Q}, S_{2Q}, Q)$$
$$+ nI(X_{1Q}, X_{2Q}; Y_Q|U_Q, Z_{1Q}, Z_{2Q}, S_{1Q}, S_{2Q}, Q) + n\epsilon_n.$$

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 $p_{Y|X_1, X_2, S_1, S_2}(y|x_1, x_2, s_1, s_2)$,
- $Z_{1Q} = z_1(X_{1Q}, S_{1Q})$ and $Z_{2Q} = z_2(X_{2Q}, S_{2Q})$.

The proof of the first condition follows because $(S_{1,i}, S_{2,i})$ are generated independently of $U_i = (Z_1^{i-1}, Z_2^{i-1}, S_1^{i-1}, S_2^{i-1})$ for all $1 \leq i \leq n$, and due to the i.i.d. assumption on (S_1, S_2). To prove that the second condition is satisfied, consider the following. For any $1 \leq i \leq n$,[5]

$$p(s_1^i, s_2^i, x_{1i}, x_{2i}, z_1^{i-1}, z_2^{i-1})$$
$$= \sum_{m_1, m_2} p(m_1, m_2, s_1^i, s_2^i, x_{1i}, x_{2i}, z_1^{i-1}, z_2^{i-1})$$
$$= \sum_{m_1, m_2} \Bigg[p(m_1)p(m_2)p(s_1^i, s_2^i)$$
$$\times p(x_{1i}|m_1, s_1^i, z_2^{i-1}) \left(\prod_{j=1}^{i-1} p(z_{1j}|m_1, s_1^j, z_2^{j-1}) \right)$$
$$\times p(x_{2i}|m_2, s_2^i, z_1^{i-1}) \left(\prod_{j=1}^{i-1} p(z_{2j}|m_2, s_2^j, z_1^{j-1}) \right) \Bigg]$$
$$\overset{6}{=} p(s_1^i, s_2^i) \sum_{m_1} p(m_1)p(x_{1i}, z_1^{i-1}|m_1, s_1^i, z_2^{i-1})$$
$$\times \sum_{m_2} p(m_2)p(x_{2i}, z_2^{i-1}|m_2, s_2^i, z_1^{i-1})$$
$$= p(s_1^i, s_2^i)p(x_{1i}, z_1^{i-1}|s_1^i, z_2^{i-1})p(x_{2i}, z_2^{i-1}|s_2^i, z_1^{i-1})$$
$$= p(s_1^i, s_2^i)p(x_{1i}|s_1^i, z_1^{i-1}, z_2^{i-1})p(z_1^{i-1}|s_1^i, z_2^{i-1})$$
$$\times p(x_{2i}, z_2^{i-1}|s_2^i, z_1^{i-1}).$$

The above factorization implies that

$$X_{1i} - (S_1^i, Z_1^{i-1}, Z_2^{i-1}) - (X_{2i}, S_{2i})$$
$$\Rightarrow X_{1i} - (U_i, S_{1i}) - (X_{2i}, S_{2i})$$
$$\Rightarrow X_{1i} - (U_i, S_{1i}) - (U_i, S_{2i}, X_{2i})$$
$$\Rightarrow X_{1Q} - (U_Q, S_{1Q}, Q) - (U_Q, S_{2Q}, X_{2Q}, Q).$$

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.

---

[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 $p(z_1^{i-1}|\cdot, z_2^{i-1})p(z_2^{i-1}|\cdot, z_1^{i-1})$, however these are valid factorizations, possibly containing redundant conditioning in some terms. The Markovity conclusions we draw 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 $p(x, y)p(x', y')p(z|x, y')p(z'|x', y)$. 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.

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

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

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 [30, Appendix C].

This concludes the proof of Theorem 4.   ∎

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 codebooks 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 $s_2 \in \mathcal{S}_2$ and each $z_1 \in \mathcal{Z}_1$, generate a codebook of $2^{nR_2'}$ codewords. The ith symbol of such a codeword is chosen independently according to $p_{Z_2|U,S_2,Z_1}(\cdot|u_{bi}(l_{b-1}), s_2, z_1)$. The result of this is that for each l_{b-1}, each $m_{2,b}' \in [1 : 2^{nR_2'}]$, each s_{2b}^n and each z_{1b}^n, encoder 2 can form an effective codeword $z_{2b}^n(m_{2,b}'|l_{b-1}, s_{2b}^n, z_{1b}^n)$, whose ith symbol can be causally chosen as the ith symbol of the $m_{2,b}'$-th codeword from the codebook corresponding to l_{b-1}, s_{2bi} and $z_{1,bi}$.

For each l_{b-1}, each $m_{2,b}' \in [1 : 2^{nR_2'}]$, each $s_2 \in \mathcal{S}_2$ and each $z_1 \in \mathcal{Z}_1$, generate a codebook of $2^{nR_2''}$ codewords. The ith symbol of such a code-word is generated independently according to $p_{X_2|U,Z_2,S_2,Z_1}(\cdot|u_{bi}(l_{b-1}), z_{2bi}(m_{2,b}'|l_{b-1}, s_2, z_1), s_2, z_1)$. The result of this construction is that for each l_{b-1}, each $m_{2,b}' \in [1 : 2^{nR_2'}]$, each $m_{2,b}'' \in [1 : 2^{nR_2''}]$, each s_{2b}^n and each z_{1b}^n, encoder 2

can form an effective codeword $x_{2b}^n(m_{2,b}''|l_{b-1}, m_{2,b}', s_{2b}^n, z_{1b}^n)$, whose ith symbol can be causally chosen as the ith symbol of the $m_{2,b}''$-th codeword from the codebook corresponding to l_{b-1}, $m_{2,b}'$, s_{2bi} and $z_{1,bi}$.

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 $x_{2b}^n(m_{2,b}''|l_{b-1}, m_{2,b}', s_{2b}^n, z_{1b}^n)$ 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

$$\mathsf{bin}_b(z_{1b}^n(m_{1,b}'|l_{b-1}, s_{1b}^n)) = l_{1,b}$$

and

$$\mathsf{bin}_b(z_{2b}^n(m_{2,b}'|l_{b-1}, s_{2b}^n, z_{1b}^n(m_{1,b}'|l_{b-1}, s_{1b}^n))) = l_{2,b}.$$

Probability of Error: In the previous section, we obtained the conditions (17)-(24) 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 $\widetilde{R}_1$, $\widetilde{R}_2$, 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 $(Z_1^{i-1}, Z_2^{i-1}, S_1^{i-1}, S_2^{i-1})$.

$$nR_2 = H(M_2)$$
$$\overset{(a)}{=} H(M_2|M_1, S_1^n, S_2^n)$$
$$\overset{(b)}{=} H(M_2, Z_2^n|M_1, S_1^n, S_2^n)$$
$$= H(Z_2^n|M_1, S_1^n, S_2^n) + H(M_2|Z_2^n, M_1, S_1^n, S_2^n)$$
$$\overset{(c)}{\leq} H(Z_2^n|M_1, S_1^n, S_2^n) + I(M_2; Y^n|Z_2^n, M_1, S_1^n, S_2^n) + n\epsilon_n$$
$$= \sum_{i=1}^{n} H(Z_{2i}|Z_2^{i-1}, M_1, S_1^n, S_2^n)$$
$$+ \sum_{i=1}^{n} I(M_2; Y_i|Y^{i-1}, Z_2^n, M_1, S_1^n, S_2^n) + n\epsilon_n$$

$$\overset{(d)}{=} \sum_{i=1}^{n} H(Z_{2i}|Z_2^{i-1}, M_1, S_1^n, S_2^n, Z_1^i)$$
$$+ \sum_{i=1}^{n} I(M_2; Y_i|Y^{i-1}, Z_2^n, M_1, S_1^n, S_2^n, X_1^n) + n\epsilon_n$$

$$\overset{(e)}{=} \sum_{i=1}^{n} H(Z_{2i}|Z_2^{i-1}, M_1, S_1^n, S_2^n, Z_1^i)$$
$$+ \sum_{i=1}^{n} I(M_2, X_{2i}; Y_i|Y^{i-1}, Z_2^n, M_1, S_1^n, S_2^n, X_1^n)$$
$$+ n\epsilon_n$$

$$\overset{(f)}{\leq} \sum_{i=1}^{n} H(Z_{2i}|Z_2^{i-1}, S_2^i, S_1^{i-1}, Z_1^i)$$
$$+ \sum_{i=1}^{n} I(X_{2i}; Y_i|Z_2^i, S_1^i, S_2^i, X_{1i}, Z_1^{i-1}) + n\epsilon_n$$

$$= \sum_{i=1}^{n} H(Z_{2i}|U_i, S_{2i}, Z_{1i})$$
$$+ \sum_{i=1}^{n} I(X_{2i}; Y_i|U_i, Z_{2i}, S_{1i}, S_{2i}, X_{1i}) + n\epsilon_n$$

$$= nH(Z_{2Q}|U_Q, S_{2Q}, Z_{1Q}, Q)$$
$$+ nI(X_{2Q}; Y_Q|U_Q, Z_{2Q}, S_{1Q}, S_{2Q}, X_{1Q}, Q) + n\epsilon_n$$

where

- Q is a random variable uniformly distributed on $[1:n]$, independent of other random variables,
- (a) follows because M_2 is independent of (M_1, S_1^n, S_2^n),
- (b) follows since Z_2^n is a function of (M_1, M_2, S_1^n, S_2^n),
- (c) follows by Fano's inequality,
- (d) follows since (X_{1i}, Z_{1i}) is a function of (M_1, S_1^n, Z_2^{i-1}),
- (e) follows since X_{2i} is a function of (M_2, S_2^n, X_1^i),
- (f) follows since (i) conditioning reduces entropy and (ii) conditioned on $(X_{1i}, X_{2i}, S_{1i}, S_{2i})$, Y_i is independent of $(M_1, M_2, X_1^n, X_2^n, S_1^n, S_2^n)$.

Regarding the joint pmf, we note the following conditions

- (S_{1Q}, S_{2Q}) is independent of (Q, U_Q), and has the pmf $p_{S_1, S_2}(s_1, s_1)$,
- $X_{1Q} - (Q, U_Q, S_{1Q}) - S_{2Q}$,
- $X_{2Q} - (Q, U_Q, S_{2Q}, Z_{1Q}) - (X_{1Q}, S_{1Q})$,
- $p_{Y_Q|U_Q, X_{1Q}, X_{2Q}, S_{1Q}, S_{2Q}}(y|u, x_1, x_2, s_1, s_2)$ is equal to $p_{Y|X_1, X_2, S_1, S_2}(y|x_1, x_2, s_1, s_2)$,
- $Z_{1Q} = z_1(X_{1Q}, S_{1Q})$ 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 $1 \leq i \leq n$,

$$p(s_1^i, s_2^i, x_{1i}, z_1^{i-1}, z_2^{i-1})$$
$$= \sum_{m_1, m_2} p(m_1, m_2, s_1^i, s_2^i, x_{1i}, z_1^{i-1}, z_2^{i-1})$$

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

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

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

The above factorization implies that

$$X_{2i} - (S_2^i, Z_1^i, Z_2^{i-1}) - (X_{1i}, S_{1i})$$
$$\Rightarrow X_{2i} - (U_i, S_{2i}, Z_{1i}) - (X_{1i}, S_{1i})$$
$$\Rightarrow X_{2Q} - (U_Q, S_{2Q}, Z_{1Q}, Q) - (X_{1Q}, S_{1Q}).$$

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

$$p_{S_1, S_2}(s_1, s_2)p_U(u)p_{X_1|U, S_1}(x_1|u, s_1)$$
$$\times p_{X_2|U, S_2, Z_1}(x_2|u, s_2, z_1)p_{Y|X_1, X_2, S_1, S_2}(y|x_1, x_2, s_1, s_2),$$

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

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

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 auxiliary random variable can be obtained using arguments based on Caratheodory's theorem as described in [30, Appendix C].

This concludes the proof of Theorem 5. ∎

X. Concluding Remarks and Some Open Problems

We presented the cooperative-bin-forward scheme and showed that it achieves the capacity region in a variety of semideterministic setups, where partial-decode-forward turns out to be suboptimal.

A number of interesting questions remain. Most importantly, how can the cooperative-bin-forward scheme be extended to, e.g. the model in Figure 2, when the source-relay link is not deterministic, but a general noisy link? Cooperative-bin-forward was developed in this paper as an alternative to partial-decode-forward. The latter naturally extends to the noisy case due to the decoding operation at the relay. The crucial high-level ingredient for establishing cooperation in both schemes is that different nodes agree on some common information. In partial-decode-forward, the agreement is established via a decoding operation at the relay. In cooperative-bin-forward, the decoding operation was removed and the nodes agreed on a bin-index, however the deterministic components in the models were crucial for this agreement. To ensure some kind of agreement between nodes in the general noisy case without using a decoding operation, one potential approach can be to explore the similarity between the operations of binning and compression. Note that binning is a form of compression, so a natural extension of cooperative-bin-forward might involve a compression operation at the relay, where part of the compression can be reconstructed at the source, thus enabling some cooperation between the source and the relay.

Another interesting question is that of designing optimal achievability schemes for all the state-dependent setups considered in this paper when the state is known only to the source encoders, causally or strictly causally. Finally, the semideterministic relay channel with two state components, one known to the source and the other to the relay, with an uninformed destination, is also an interesting open question.

Acknowledgments

References

[1] A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel (corresp.)," *IEEE Trans. Inf. Theory*, vol. 28, no. 3, p. 536, May 1982.

[2] Y.-H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1328–1329, Mar. 2008.

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

[4] H. Asnani and H. H. Permuter, "Multiple-access channel with partial and controlled cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2252–2266, Apr. 2013.

[5] A. El Gamal, N. Hassanpour, and J. Mammen, "Relay networks with delays," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3413–3431, Oct. 2007.

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

[7] B. Akhbari, M. Mirmohseni, and M. R. 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. N. Khormuji and M. Skoglund, "The relay channel with partial causal state information," in *Proc. Int. Symp. Inf. Theory Appl.*, Dec. 2008, pp. 1–6.

[9] M. N. Khormuji, A. El Gamal, and M. Skoglund, "State-dependent relay channel: Achievable rate and capacity of a semideterministic class," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2629–2638, May 2013.

[10] Z. Deng, F. Lang, B.-Y. Wang, and S.-M. Zhao, "Capacity of a class of relay channel with orthogonal components and non-causal channel state," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 2696–2700.

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

[12] A. Zaidi, S. Shamai (Shitz), P. Piantanida, and L. Vandendorpe, "Bounds on the capacity of the relay channel with noncausal state at the source," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2639–2672, May 2013.

[13] I. E. Aguerri and D. Gunduz, "Capacity of a class of relay channels with state," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2012, pp. 277–281.

[14] A. Lapidoth and Y. Steinberg, "The multiple-access channel with causal side information: Common state," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 32–50, Jan. 2013.

[15] A. Lapidoth and Y. Steinberg, "The multiple-access channel with causal side information: Double state," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1379–1393, Mar. 2013.

[16] M. Li, O. Simeone, and A. Yener, "Multiple access channels with states causally known at transmitters," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1394–1404, Mar. 2013.

[17] T. Philosof, R. Zamir, U. Erez, and A. J. Khisti, "Lattice strategies for the dirty multiple access channel," *IEEE Trans. Inf. 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 *Proc. Int. Symp. Inf. Theory*, Sep. 2005, pp. 72–76.

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

[20] G. Como and S. Yuksel, "On the capacity of memoryless finite-state multiple-access channels with asymmetric state information at the encoders," *IEEE Trans. Inf. 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 Trans. Inf. Theory*, vol. 59, no. 11, pp. 7052–7070, Nov. 2013.

[22] F. M. J. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders (corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 441–445, May 1983.

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

[24] S. I. Bross and A. Lapidoth, "The state-dependent multiple-access channel with states available at a cribbing encoder," in *Proc. IEEE 26th Conv. Elect. Electron. Eng. Israel*, Nov. 2010, pp. 665–669.

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

[26] M. Zamanighomi, M. J. Emadi, F. S. Chaharsooghi, and M. R. Aref, "Multiple access channel with correlated channel states and cooperating encoders," in *Proc. IEEE Inf. Theory Workshop*, Oct. 2011, pp. 628–632.

[27] M. J. Emadi, M. N. Khormuji, M. Skoglund, and M. R. Aref, "On the achievable rate region of a state-dependent MAC with cooperating encoders," in *Proc. Swedish Commun. Technol. Workshop*, Oct. 2012, pp. 48–52.

[28] A. Zaidi and S. Shamai (Shitz), "On cooperative multiple access channels with delayed CSI at transmitters," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6204–6230, Oct. 2014.

[29] A. Zaidi, P. Piantanida, and S. Shamai (Shitz), "Capacity region of cooperative multiple-access channel with states," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6153–6174, Oct. 2013.

[30] A. El Gamal and Y.-H. Kim, *Network Information Theory*. New York, NY, USA: Cambridge Univ. Press, 2012.

Ritesh Kolte (S'13) is a doctoral candidate in Electrical Engineering at Stanford University. His research interests are in the general area of information theory, optimization, statistical learning and inference, signal processing. He received the Bachelor of Technology and Master of Technology degrees from the Indian Institute of Technology Bombay in 2010 and the Master of Science degree from Stanford University in 2012, all in Electrical Engineering; and is a recipient of the Stanford Graduate Fellowship.

Ayfer Özgür (M'06) received her B.Sc. degrees in electrical engineering and physics from Middle East Technical University, Turkey, in 2001 and the M.Sc. degree in communications from the same university in 2004. From 2001 to 2004, she worked as hardware engineer for the Defense Industries Development Institute in Turkey. She received her Ph.D. degree in 2009 from the Information Processing Group at EPFL, Switzerland. In 2010 and 2011, she was a post-doctoral scholar with the Algorithmic Research in Network Information Group at EPFL. She is currently an Assistant Professor in the Electrical Engineering Department at Stanford University. Her research interests include network communications, wireless systems, and information and coding theory. Dr. Özgür received the EPFL Best Ph.D. Thesis Award in 2010 and a NSF CAREER award in 2013.

Haim Permuter (M'08–SM'13) received his B.Sc. (summa cum laude) and M.Sc.(summa cum laude) degrees in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 1997 and 2003 respectively, and the Ph.D. degree in Electrical Engineering from Stanford University, California in 2008.

Between 1997 and 2004, he was an officer at a research and development unit of the Israeli Defense Forces. Since 2009 he is with the department of Electrical and Computer Engineering at Ben-Gurion University where he is currently an associate professor.

Prof. Permuter is a recipient of several awards, among them the Fullbright Fellowship, the Stanford Graduate Fellowship (SGF), Allon Fellowship, and and the U.S.-Israel Binational Science Foundation Bergmann Memorial Award. He is currently serving on the editorial boards of the IEEE TRANSACTIONS ON INFORMATION THEORY.