Capacity Theorems for Discrete, Finite-State Broadcast Channels With Feedback and Unidirectional Receiver Cooperation

Ron Dabora and Andrea J. Goldsmith, *Fellow, IEEE*

***Abstract*—In this paper, we consider the discrete, time-varying broadcast channel (BC) with memory under the assumption that the channel states belong to a set of finite cardinality. We study the achievable rates in several scenarios of feedback and full unidirectional receiver cooperation. In particular, we focus on two scenarios: the first scenario is the general finite-state broadcast channel (FSBC) where both receivers send feedback to the transmitter while one receiver also sends its channel output to the second receiver. The second scenario is the degraded FSBC where only the strong receiver sends feedback to the transmitter. Using a superposition codebook construction, we derive the capacity regions for both scenarios. Combining elements from these two basic results, we obtain the capacity regions for a number of additional broadcast scenarios with feedback and unidirectional receiver cooperation.**

***Index Terms*—Broadcast channels, capacity, channels with memory, cooperation, feedback, finite-state channels (FSCs), network information theory, superposition codebook.**

I. Introduction

THE finite-state channel (FSC) was introduced as early as 1953 as a model for time-varying channels with memory [1]. In this model, the channel memory is captured by the state of the channel at the end of the previous symbol transmission. Put in mathematical terms, let X_i denote the channel input, Y_i the channel output, and S_i the channel state, all at time i; the transition function of the point-to-point (PtP) FSC at time i satisfies $p(y_i, s_i \,|\, x^i, y^{i-1}, s^{i-1}, s_0) = p(y_i, s_i \,|\, x_i, s_{i-1})$. Recently, capacity analysis for time-varying channels with memory and additive white Gaussian noise have been the focus of considerable interest. This has been motivated by the proliferation of mobile communications in which the channel is subject to multipath and correlated fading, both of which introduce memory into the time-varying channel. The capacity of PtP-FSCs without feedback was originally studied by Gallager [2] and has been applied to specific channels; see [3] and references therein. In recent papers, the capacity of FSCs with feedback has been studied [4], [5]. In particular, it was shown in [4] that feedback can increase the capacity of some FSCs. The capacity of finite-state multiple-access channels (FS-MACs) with and without feedback was also studied in a recent work [6].

The finite-state broadcast channel (FSBC) was originally introduced in [7] and [8]. In this scenario, a single transmitter communicates with two receivers over an FSC; see Fig. 1. In [7] and [8], we considered the degraded FSBC without feedback. Specifically, we first defined the notions of physical and stochastic degradedness for channels with memory and then showed that a superposition codebook *with memory* achieves capacity for degraded FSBCs. The capacity region was obtained as a limiting expression by taking the codeword length to infinity. It was also demonstrated in [8] that, for FSBCs, a codebook with memory achieves strictly higher rates than memoryless codebooks. In the current work, we extend [7] and [8] by considering the effects of feedback and full unidirectional receiver cooperation on the capacity region. It is well known that for physically degraded, memoryless broadcast channels (BCs), feedback does not increase the capacity region [9]. However, for *stochastically degraded* memoryless BCs, Ozarow showed that feedback does increase the capacity region [10]. Furthermore, in [11], it was shown that feedback from just one user can increase the capacity region of the stochastically degraded memoryless BC. Combined with the fact mentioned earlier, that feedback can increase the capacity of PtP-FSCs, we conclude that feedback can increase the capacity region of FSBCs, and specifically, the capacity region of physically degraded FSBCs can be increased by feedback from just one user. Broadcast scenarios which incorporate receiver cooperation as well as feedback were considered originally in [10], which referred to such scenarios as *augmented BCs*. In [12], the Gaussian augmented BC with feedback and Gaussian interference was analyzed for the case where the interference is noncausally available at the transmitter.

In the context of discrete BCs with random parameters, we note that the capacity of memoryless BCs with random parameters was investigated in [13] and [14]. In [13], degraded memoryless BCs with random parameters with causal and noncausal side information at the transmitter were considered, and in [14], an achievable rate region for general memoryless

Manuscript received January 13, 2009; revised December 28, 2009. Date of current version November 19, 2010. This work was supported by the DARPA ITMANET program under Grant 1105741-1-TFIND. The material in this paper was presented in part at the IEEE International Symposium on Information Theory (ISIT), Toronto, ON, Canada, July 2008 and at the Information Theory Workshop (ITW), Volos, Greece, June 2009.

R. Dabora is with the Department of Electrical and Computer Engineering, Ben-Gurion University, Be'er Sheva 84105, Israel (e-mail: ron@ee.bgu.ac.il).

A. J. Goldsmith is with the Department of Electrical Engineering, Stanford University, CA 94305 USA (e-mail: andrea@wsl.stanford.edu).

Communicated by M. C. Gastpar, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2010.2081010

0018-9448/$26.00 © 2010 IEEE

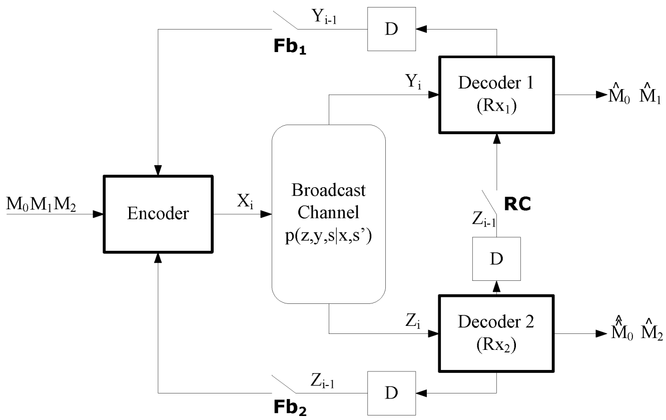

Fig. 1. The general FSBC with feedback and receiver cooperation. The "D"-block indicates a single symbol-time delay. Each of the three switches Fb_1, Fb_2, or RC can be either connected or disconnected. Switch RC enables cooperation from Rx_2 to Rx_1, switch Fb_2 enables feedback from Rx_2 to the transmitter, and switch Fb_1 enables feedback from Rx_1 to the transmitter.

BCs with random parameters was derived assuming that the states were noncausally known at the transmitter.

Receiver cooperation in *discrete, memoryless* broadcast channels (DMBCs) was first discussed in [10] as a tool for upper bounding the capacity region of BCs with feedback. In [15], receiver cooperation with both unidirectional and bidirectional orthogonal links was discussed for DMBCs with independent components, and in [16], both types of cooperation were studied for general DMBCs. A multistep conference for general DMBCs with orthogonal cooperation links was studied in [17]. Receiver cooperation for DMBCs with nonorthogonal links was considered in [18] and [19], where unidirectional and bidirectional cooperation was studied with and without feedback.

A. Main Contributions and Organization

We study FSBCs for several feedback and unidirectional receiver cooperation scenarios; see Fig. 1. As each of the feedback links as well as the cooperation link depicted in Fig. 1 may be connected or disconnected, there are eight possible configurations. The configuration without feedback and cooperation was analyzed in previous work [7], [8]. In this work, we address the remaining seven configurations. We refer to each configuration by indicating which switches are connected. For example, the configuration with all switches connected is referred to as Fb1-Fb2-RC and when only the two feedback links are connected it is referred to as Fb1-Fb2. We also indicate physically degraded situations by adding "PD" when referring to the configuration.

We derive the capacity regions for two scenarios. The first is Fb1-Fb2-RC for the general FSBC (Section IV-A). In the second scenario, the FSBC is physically degraded and only one receiver (the strong receiver) sends feedback but there is no receiver cooperation (i.e., only switch Fb_1 is connected; Section IV-B). We refer to this scenario as Fb1-PD. We develop a superposition *codetree* construction in which a codetree that represents the channel codeword is superimposed on a codetree that represents the cloud center. This extends the superposition construction to channels with memory. Considerable parts of the proofs are dedicated to showing that the properties of the superposition construction also hold for the case with feedback and memory. We also derive cardinality bounds for the auxiliary random vectors representing the cloud centers in the superposition construction for Fb1-PD. Here, the cardinality bounds reflect the effect of the feedback and the finite-state property. Finally, combining elements from the coding schemes for scenarios Fb1-Fb2-RC and Fb1-PD, we use superposition constructions to derive the capacity regions for several other FSBC scenarios with feedback and receiver cooperation.

The rest of this paper is organized as follows. In Section II, we introduce the notations, and in Section III, we present the channel model and definitions. In Section IV, we derive the capacity regions for scenarios Fb1-Fb2-RC and Fb1-PD, the proofs of which are provided in Appendixes A and B, and in Section V, we analyze all remaining configurations of feedback and receiver cooperation. Finally, Section VI concludes the work.

II. Notations

In the following, we denote random variables (RVs) with upper case letters, e.g., X, Y, and their realizations with lower case letters x, y. An RV X takes values in a set $\mathcal{X}$. We use $|\mathcal{X}|$ to denote the cardinality of a finite, discrete set $\mathcal{X}$, $\mathcal{X}^n$ to denote the n-fold Cartesian product of $\mathcal{X}$, $\times_{i=1}^n \mathcal{X}_i$ to denote $\mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n$ for (possibly different) sets $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$, and $p_X(x)$ to denote the probability mass function (p.m.f.) of a discrete RV X on $\mathcal{X}$. For brevity, we may omit the subscript X when it is the uppercase version of the realization x. We use $p_{X|Y}(x|y)$ to denote the conditional p.m.f. of X given $Y = y$. We denote vectors with boldface letters, e.g., $\mathbf{x}$, $\mathbf{y}$; the ith element of a vector $\mathbf{x}$ is denoted with x_i and we use x_i^j where $i < j$ to denote the vector $(x_i, x_{i+1}, \ldots, x_{j-1}, x_j)$; x^j is a short form notation for x_1^j, and $\mathbf{x} \equiv x^n$. A vector of RVs is denoted by $\mathbf{X} \equiv X^n$, and similarly, we define $X_i^j \triangleq (X_i, X_{i+1}, \ldots, X_{j-1}, X_j)$ for $i < j$. When $i > j$, $x_i^j = X_i^j = \emptyset$, the empty set. We use $H(\cdot)$ to denote the entropy of a discrete RV and $I(\cdot;\cdot)$ to denote the mutual information between two RVs, as defined in [2, Ch. 2]. $I(\cdot;\cdot)_q$ denotes the mutual information evaluated with the channel inputs distributed according to the p.m.f. q. Finally, $co\mathcal{R}$ denotes the convex hull of the set $\mathcal{R}$.

A. Directed Mutual Information and Causal Conditioning

Throughout this paper, we use the compact notations of directed mutual information and causal conditioning, originally introduced in [20, Sec. II.C], [21], [22]. Directed mutual information between two random vectors X^n and Y^n, given a third vector Z^n, is defined as

$$I(X^n \to Y^n | Z^n) \triangleq \sum_{i=1}^{n} I(X^i; Y_i | Y^{i-1}, Z^n). \qquad (1)$$

At every instant i, the mutual information

$$I(X^i; Y_i | Y^{i-1}, Z^n) = H(X^i | Y^{i-1}, Z^n) - H(X^i | Y^i, Z^n)$$

can be viewed as the incremental "decrease in uncertainty" about the sequence X^i at the receiver due to the reception of Y_i. Summing over all time indices, we obtain the total "decrease in uncertainty" for the n symbol transmissions. Directed mutual information arises naturally in the analysis of causal channels with memory.

The probability distribution of X^n causally conditioned on Y^n is defined as

$$Q(x^n \,|\, |\, y^{n-1}) \triangleq \prod_{i=1}^{n} p(x_i \,|\, x^{i-1}, y^{i-1}). \tag{2}$$

This is a compact way for describing a signal X^n in which every symbol X_i is generated using causal knowledge of the signal Y^n, i.e., X_i is a function of X^{i-1} and Y^{i-1}. Causal conditioning is widely used in describing results for feedback scenarios. A codeword x^n generated as a causal function of y^n is referred to as a *codetree*; see [20] for a detailed explanation.

We use three additional notations

$$I(X^n \to Y^n \| Z^n) \triangleq \sum_{i=1}^{n} I(X^i; Y_i \,|\, Y^{i-1}, Z^i)$$
$$Q(x^n \| y^n, z^{n-1}) \triangleq \prod_{i=1}^{n} p(x_i \,|\, x^{i-1}, y^i, z^{i-1})$$
$$Q(x^n \| y^{n-1}; z^n) \triangleq \prod_{i=1}^{n} p(x_i \,|\, x^{i-1}, y^{i-1}, z^n).$$

III. Channel Model

In the definition of the FSBC, we aim to capture the essential property of FSCs, namely that the channel state at the end of the previous symbol transmission represents all the channel past for the current transmission. This leads to the following formal definition of the FSBC.

Definition 1: The *FSBC* is defined by the triplet $\{\mathcal{X} \times \mathcal{S}, p(y, z, s \,|\, x, s'), \mathcal{Y} \times \mathcal{Z} \times \mathcal{S}\}$, where X is the input symbol, Y and Z are the output symbols, S' is the channel state at the end of the previous symbol transmission, and S is the channel state at the end of the current symbol transmission. $\mathcal{S}, \mathcal{X}, \mathcal{Y}$, and $\mathcal{Z}$ are discrete sets of finite cardinalities.

Throughout this work, we assume that the receivers and transmitter operate *without knowledge of the channel states*. Under this assumption, the p.m.f of a block of n transmissions factors as

$$\begin{aligned} &p(y^n, z^n, s^n, x^n \,|\, s_0) \\ &\quad = \prod_{i=1}^{n} p(y_i, z_i, s_i, x_i \,|\, y^{i-1}, z^{i-1}, s^{i-1}, x^{i-1}, s_0) \\ &\quad \overset{(a)}{=} \prod_{i=1}^{n} p(x_i \,|\, y^{i-1}, z^{i-1}, x^{i-1}) \\ &\qquad \times p(y_i, z_i, s_i \,|\, y^{i-1}, z^{i-1}, s^{i-1}, x^i, s_0) \\ &\quad \overset{(b)}{=} \prod_{i=1}^{n} p(x_i \,|\, y^{i-1}, z^{i-1}, x^{i-1}) \prod_{i=1}^{n} p(y_i, z_i, s_i \,|\, x_i, s_{i-1}) \end{aligned}$$

where s_0 is the initial channel state. Here, (a) follows because the transmitter is oblivious of the channel states but we do not rule out feedback and (b) captures the fact that S_{i-1} represents

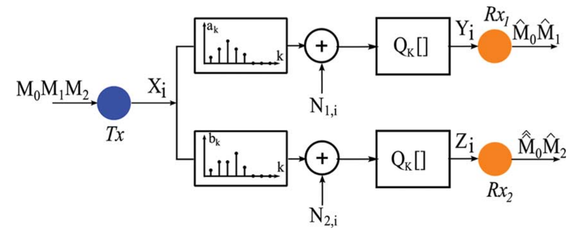

Fig. 2. A schematic description of the FSBC with ISI.

the past for time i. The FSBC models many practical communication scenarios. As an example, consider the Gaussian BC with intersymbol interference (ISI)

$$y_i = Q_K\left[\sum_{j=0}^{J} a_j x_{i-j} + n_{1,i}\right]$$
$$z_i = Q_K\left[\sum_{j=0}^{J} b_j x_{i-j} + n_{2,i}\right]$$

where $\{a_j\}_{j=0}^J$ and $\{b_j\}_{j=0}^J$ are the channel coefficients, $Q_K[\cdot]$ is a quantizer with K levels, $n_{1,i} \sim \mathcal{N}(0, \sigma_1^2)$ and $n_{2,i} \sim \mathcal{N}(0, \sigma_2^2)$ are each an independent and identically distributed (i.i.d.) bandlimited white Gaussian process, and $x_i \in \mathcal{X}$ is a finite signal constellation; all variables are real. Here the channel state is $S_{i-1} = (X_{i-J}, X_{i-J+1}, \ldots, X_{i-1})$, giving rise to a state space whose cardinality is $|\mathcal{S}| = |\mathcal{X}|^J$. This channel is depicted in Fig. 2.

This model is very common in communication [23], and its applications include wireless communications [24], digital subscriber lines [25, Ch. 11.2.2] and TV [26].

Definition 2: An (R_0, R_1, R_2, n) *deterministic code* for the FSBC for scenario Fb1-Fb2-RC (feedback and full unidirectional receiver cooperation), without knowledge of the channel states at the transmitter and receivers, consists of three message sets $\mathcal{M}_0 = \{1, 2, \ldots, 2^{nR_0}\}$, $\mathcal{M}_1 = \{1, 2, \ldots, 2^{nR_1}\}$, and $\mathcal{M}_2 = \{1, 2, \ldots, 2^{nR_2}\}$, and a collection of mappings $(\{f_i\}_{i=1}^n, g_{\mathbf{y}}, g_{\mathbf{z}})$ such that

$$f_i : \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{Z}^{i-1} \times \mathcal{Y}^{i-1} \mapsto \mathcal{X} \tag{3}$$

is the encoder at time i, and

$$g_{\mathbf{y}} : \mathcal{Z}^n \times \mathcal{Y}^n \mapsto \mathcal{M}_0 \times \mathcal{M}_1 \tag{4a}$$
$$g_{\mathbf{z}} : \mathcal{Z}^n \mapsto \mathcal{M}_0 \times \mathcal{M}_2 \tag{4b}$$

are the decoders. Here, $\mathcal{M}_0$ is the set of common messages and $\mathcal{M}_1$ and $\mathcal{M}_2$ are the sets of private messages to Rx_1 and Rx_2, respectively.

The following definitions extend the notions of average probability of error, achievable rate, and capacity [27], [5] to the FSBC.

Definition 3: The *average probability of error* of a code of blocklength n is defined as $\max_{s_0 \in \mathcal{S}} P_e^{(n)}(s_0)$, where for scenario Fb1-Fb2-RC

$$\begin{aligned} P_e^{(n)}(s_0) = \Pr(&\{g_{\mathbf{y}}(Z^n, Y^n) \neq (M_0, M_1)\} \\ &\text{or } \{g_{\mathbf{z}}(Z^n) \neq (M_0, M_2)\} \,|\, S_0 = s_0) \end{aligned} \tag{5}$$

and the messages $M_0 \in \mathcal{M}_0$, $M_1 \in \mathcal{M}_1$, and $M_2 \in \mathcal{M}_2$ are selected independently and uniformly over their message sets.

Definition 4: A rate triplet (R_0, R_1, R_2) is called *achievable* for the FSBC if for every $\epsilon > 0$ and $\delta > 0$ there exists an $n(\epsilon, \delta) \in \mathbb{N}$ such that $\forall n > n(\epsilon, \delta)$, an $(R_0 - \delta, R_1 - \delta, R_2 - \delta, n)$ code with $\max_{s_0 \in \mathcal{S}} P_e^{(n)}(s_0) \leq \epsilon$ can be constructed.

Definition 5: The *capacity region* of the FSBC is the convex hull of all achievable rate triplets.

Remark: The definitions above assume full unidirectional receiver cooperation (scenario Fb1-Fb2-RC). In scenario Fb1-PD, Z^n is not available at Rx_1 or at the transmitter. Thus, for scenario Fb1-PD, we modify Definitions 2 and 3 by replacing $g_{\mathbf{y}}(Z^n, Y^n)$ with $g_{\mathbf{y}}(Y^n)$ in (5) and using $g_{\mathbf{y}} : \mathcal{Y}^n \mapsto \mathcal{M}_0 \times \mathcal{M}_1$ in (4a) and $f_i : \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{Y}^{i-1} \mapsto \mathcal{X}$ in (3).

Sections III-A–III-C define the notion of degradedness [28] for channels with memory. Roughly speaking, degradedness implies that one receiver is stronger than the other receiver in the sense that when the weak receiver can reliably decode its message, the strong receiver can reliably decode the same message as well. It should be noted, however, that when the channel has memory, a feedback link can turn a degraded FSBC into a nondegraded one.

A. Physically Degraded FSBCs

Definition 6: The FSBC is called *physically degraded* if for every initial channel state $s_0 \in \mathcal{S}$ and symbol time i its p.m.f. satisfies

$$p(y_i \mid x^i, y^{i-1}, z^{i-1}, s_0) = p(y_i \mid x^i, y^{i-1}, s_0) \qquad (6a)$$
$$p(z_i \mid x^i, y^i, z^{i-1}, s_0) = p(z_i \mid y^i, z^{i-1}, s_0). \qquad (6b)$$

Here, Rx_1 (whose channel output is Y) is referred to as the "strong receiver" and Rx_2 (whose channel output Z is considered a degraded version of Y) is referred to as the "weak receiver."

Condition (6a) captures the intuitive notion of degradedness, namely that Z^{i-1} is a degraded version of Y^{i-1}, thus it does not add information (in the sense of statistical dependence) when Y^{i-1} is given. Note that in the memoryless case this condition is not necessary as, given X_i, Y_i is independent of the history. Condition (6b) follows from the second aspect of degradedness, that, when the history Z^{i-1} is given, Y^i makes Z_i independent of X^i. For memoryless channels, (6b) reduces to $p(z_i \mid y_i, x_i) = p(z_i \mid y_i)$. Thus, for memoryless channels, Definition 6 reduces to the standard definition of degradedness. Note that this definition does not involve the evolution of the states. It can be viewed as applied to the p.m.f. of the FSBC after averaging over all state sequences. From (6), it follows that for physically degraded channels:

$$Q(y^n \| z^{n-1}, x^n, s_0) = Q(y^n \| x^n, s_0)$$
$$Q(z^n \| y^n, x^n, s_0) = Q(z^n \| y^n, s_0).$$

This can be viewed as the equivalent of $p(y, z \mid x) = p(y \mid x) p(z \mid y)$ which characterizes the joint distribution of the outputs for degraded DMBCs.

B. The Physically Degraded FSBC Scenario Fb1-PD

As noted at the beginning of this section, the computation of $p(y_i \mid x^i, y^{i-1}, z^{i-1}, s_0)$ *depends on the feedback scheme*. For example, for scenario Fb1-PD, $x_i = f_i(x^{i-1}, y^{i-1})$ and $p(y_i \mid x^i, y^{i-1}, z^{i-1}, s_0)$ can be obtained from $p(y_i, z_i, s_i \mid x_i, s_{i-1})$ via the steps in (7)–(8), shown at the bottom of the page.

Using conditions (6a) and (6b) in scenario Fb1-PD, we obtain

$$\begin{aligned}
&p(z^n \mid y^n, x^n, s_0) \\
&\quad = \frac{p(z^n, y^n, x^n \mid s_0)}{p(y^n, x^n \mid s_0)} \\
&\quad = \frac{\prod_{i=1}^n p(z_i, y_i, x_i \mid z^{i-1}, y^{i-1}, x^{i-1}, s_0)}{p(y^n, x^n \mid s_0)} \\
&\quad = \frac{\prod_{i=1}^n p(x_i \mid z^{i-1}, y^{i-1}, x^{i-1})}{p(y^n, x^n \mid s_0)} \\
&\qquad \times \prod_{i=1}^n p(z_i, y_i \mid z^{i-1}, y^{i-1}, x^i, s_0) \\
&\quad \overset{(a)}{=} \frac{\prod_{i=1}^n p(x_i \mid y^{i-1}, x^{i-1})}{\prod_{i=1}^n p(x_i \mid y^{i-1}, x^{i-1})} \\
&\qquad \times \frac{\prod_{i=1}^n p(z_i, y_i \mid z^{i-1}, y^{i-1}, x^i, s_0)}{\prod_{i=1}^n p(y_i \mid y^{i-1}, x^i, s_0)} \\
&\quad \overset{(b)}{=} \frac{\prod_{i=1}^n p(y_i \mid y^{i-1}, x^i, s_0) \prod_{i=1}^n p(z_i \mid z^{i-1}, y^i, x^i, s_0)}{\prod_{i=1}^n p(y_i \mid y^{i-1}, x^i, s_0)} \\
&\quad \overset{(c)}{=} \prod_{i=1}^n p(z_i \mid z^{i-1}, y^i, s_0)
\end{aligned} \qquad (9)$$

$$p(y_i \mid x^i, y^{i-1}, z^{i-1}, s_0) = \frac{\sum_{\mathcal{S}^i} p(y^i, x^i, s^i, z^{i-1} \mid s_0)}{\sum_{\mathcal{S}^i} p(y^{i-1}, x^i, s^i, z^{i-1} \mid s_0)} \qquad (7)$$

$$\begin{aligned}
&= \frac{\sum_{\mathcal{S}^i} \prod_{j=1}^{i-1} p(y_j, x_j, s_j, z_j \mid y^{j-1}, x^{j-1}, s^{j-1}, z^{j-1}, s_0) p(y_i, x_i, s_i \mid y^{i-1}, x^{i-1}, s^{i-1}, z^{i-1}, s_0)}{\sum_{\mathcal{S}^i} \prod_{j=1}^{i-1} p(y_j, x_j, s_j, z_j \mid y^{j-1}, x^{j-1}, s^{j-1}, z^{j-1}, s_0) p(x_i, s_i \mid y^{i-1}, x^{i-1}, s^{i-1}, z^{i-1}, s_0)} \\
&= \frac{\sum_{\mathcal{S}^{i-1}} \prod_{j=1}^{i-1} p(x_j \mid y^{j-1}, x^{j-1}) p(y_j, s_j, z_j \mid x_j, s_{j-1}) p(x_i \mid y^{i-1}, x^{i-1}) p(y_i \mid x_i, s_{i-1})}{\sum_{\mathcal{S}^{i-1}} \prod_{j=1}^{i-1} p(x_j \mid y^{j-1}, x^{j-1}) p(y_j, s_j, z_j \mid x_j, s_{j-1}) p(x_i \mid y^{i-1}, x^{i-1})} \\
&= \frac{\sum_{\mathcal{S}^{i-1}} \prod_{j=1}^{i-1} p(y_j, s_j, z_j \mid x_j, s_{j-1}) p(y_i \mid x_i, s_{i-1})}{\sum_{\mathcal{S}^{i-1}} \prod_{j=1}^{i-1} p(y_j, s_j, z_j \mid x_j, s_{j-1})}.
\end{aligned} \qquad (8)$$

where (a) holds because Fb_2 is disconnected in Fb1-PD, (b) follows from (6a), and (c) follows from (6b). We conclude that when (6) holds, $p(z^n \mid y^n, x^n, s_0) = p(z^n \mid y^n, s_0)$. Hence

$$p(y^n, z^n, x^n \mid s_0) = p(y^n, x^n \mid s_0) p(z^n \mid y^n, s_0), \qquad \forall s_0 \in \mathcal{S} \tag{10}$$

or equivalently, $\forall s_0 \in \mathcal{S}$, $X^n \mid s_0 \leftrightarrow Y^n \mid s_0 \leftrightarrow Z^n \mid s_0$. Note that Z^n is a degraded version of Y^n but still depends on the state sequence (i.e., degradedness does not eliminate the memory). A special case of the physically degraded FSBC occurs when in (6b) we have $p(z_i \mid x^i, y^i, z^{i-1}, s_0) = p(z_i \mid y_i)$. Hence

$$p(z^n \mid y^n, s_0) = \prod_{i=1}^{n} p(z_i \mid y_i) = p(z^n \mid y^n). \tag{11}$$

Equation (11) is similar to the definition of degradedness for BCs with random parameters used in [13]. Condition (11) does not constitute only a mathematical relationship, but represents physical scenarios, as was demonstrated in [7].

C. Stochastically Degraded FSBCs

A weaker notion of degradedness is stochastic degradedness. Here, although the underlying physical process is not degraded, it is possible to find a physically degraded channel for which the marginal distributions of the channel outputs are the same as in the original channel. Since the capacity of the FSBC depends on the marginal distributions, the capacity of the two models is the same. Here, care should be taken to account for the effect of the feedback: it may be that a channel is stochastically degraded without feedback but stops being degraded when feedback is used (see scenario Fb2-PD in Section V-E). For this reason, the following definition uses the joint distribution $p(y^n, x^n \mid s_0)$ instead of the common practice of using the marginal $p(y^n \mid x^n, s_0)$.

Definition 7: The FSBC is called *stochastically degraded* if there exists a p.m.f. $\tilde{p}(z \mid y)$ such that

$$\begin{aligned} p(z^n, x^n \mid s_0) &= \sum_{\mathcal{Y}^n} p(z^n, y^n, x^n \mid s_0) \\ &= \sum_{\mathcal{Y}^n} p(y^n, x^n \mid s_0) \prod_{i=1}^{n} \tilde{p}(z_i \mid y_i). \end{aligned} \tag{12}$$

Equation (12) implies the Markov chain $X^n \mid s_0 \leftrightarrow Y^n \mid s_0 \leftrightarrow Z^n \mid s_0$, thus from the data processing inequality [29, Th. 2.8.1] $I(X^n; Y^n \mid s_0) \geq I(X^n; Z^n \mid s_0)$. We note that the way we defined stochastic degradedness for the FSBC is restrictive in the sense that the class of degrading channels is constrained to only memoryless channels, e.g., $\tilde{p}(z^n \mid y^n, x^n, s_0) = \prod_{i=1}^{n} \tilde{p}(z_i \mid y_i)$. It is possible to provide an n-letter definition for stochastic degradedness, e.g., define a channel to be stochastically degraded if for *every* blocklength n we can find a $\tilde{p}(z^n \mid y^n, s_0)$ such that

$$p(z^n, x^n \mid s_0) = \sum_{\mathcal{Y}^n} p(y^n, x^n \mid s_0) \tilde{p}(z^n \mid y^n, s_0).$$

However, such a definition does not give insight into the causal operation of the degrading channel (see [8, Sec. II-C]), and we therefore prefer to avoid it.

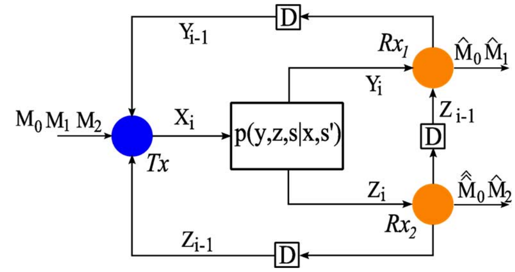

Fig. 3. Scenario Fb1-Fb2-RC: The general FSBC with feedback from both receivers and unidirectional receiver cooperation.

IV. Characterization of the Capacity Regions for Scenarios Fb1-Fb2-RC and Fb1-PD

A. Scenario Fb1-Fb2-RC: The General FSBC With Switches Fb_1, Fb_2, and RC Connected

In this scenario, Rx_1 sends its channel output to the transmitter and Rx_2 sends its channel output to both Rx_1 and the transmitter; see Fig. 3. This results in a physically degraded channel $X^n \mid s_0 \leftrightarrow (Y^n, Z^n) \mid s_0 \leftrightarrow Z^n \mid s_0$, but as the channel is not degraded without receiver cooperation it does not satisfy the Markov chain $X^n \mid s_0 \leftrightarrow Y^n \mid s_0 \leftrightarrow Z^n \mid s_0$. The scenario in which the channel output at one receiver is available to the other receiver is also called the augmented BC [10]. Note that the one-symbol delay from Rx_2 to Rx_1 is of no significance as the receivers can wait until the end of the entire block before beginning the decoding procedure. The capacity region for scenario Fb1-Fb2-RC is stated in the following theorem.

Theorem 1: Let $\mathcal{Q}_n^{SC1}$ be the set of all distributions of the form

$$\begin{aligned} \mathcal{Q}_n^{SC1} = \Big\{ & q(U^n, X^n \mid Y^n, Z^n) : \\ & q(u^n, x^n \mid y^n, z^n) \\ & \quad = Q(u^n \| z^{n-1}) Q(x^n \| y^{n-1}, z^{n-1}, u^n) \\ & \text{for all length } n \text{ sequences that belong to} \\ & \mathcal{U}^n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n \Big\}, \qquad \forall n \in \mathbb{N} \end{aligned}$$

where $u^n \in \mathcal{U}^n$ and $|\mathcal{U}| < \infty$. Define the region $\mathcal{R}_n^{SC1}$ as

$$\begin{aligned} \mathcal{R}_n^{SC1} = \mathrm{co} \bigcup_{q_n \in \mathcal{Q}_n^{SC1}} \Big\{ & (R_0, R_1, R_2) : \\ & R_0 \geq 0, R_1 \geq 0, R_2 \geq 0, \\ & R_1 \leq \min_{s_0' \in \mathcal{S}} \frac{1}{n} I(X^n \to Z^n, Y^n \| U^n, s_0')_{q_n}, \\ & R_0 + R_2 \leq \min_{s_0'' \in \mathcal{S}} \frac{1}{n} I(U^n \to Z^n \mid s_0'')_{q_n} \Big\}. \end{aligned}$$

For the general FSBC with feedback such that switches Fb_1, Fb_2, and RC are connected, the capacity region is given by

$$\mathcal{C}_{fb}^{SC1} = \lim_{n \to \infty} \mathcal{R}_n^{SC1}$$

and the limit exists and is finite.

Proof: The proof of Theorem 1 is provided in Appendix A. □

Comment 1: Since in scenario Fb1-Fb2-RC the channel is effectively a physically degraded BC, a superposition *codetree* achieves capacity. Interpreting superposition coding in terms of

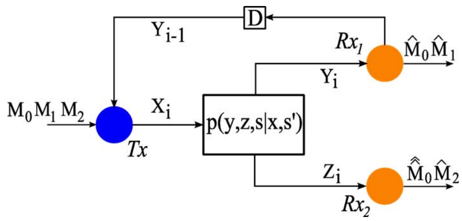

Fig. 4. Scenario Fb1-PD: The physically degraded FSBC with feedback only from the strong receiver Rx_1.

"cloud centers" and "cloud elements" we note that in scenario Fb1-Fb2-RC the cloud centers are in fact codetrees generated according to $Q(u^n \| z^{n-1})$ on which cloud elements, which are also codetrees, generated according to $Q(x^n \| y^{n-1}, z^{n-1}, u^n)$ are superimposed. This construction extends the superposition codebook to the case with memory and feedback. The focus of the achievability proof is to show how the properties of the superposition construction translate to this case.

Comment 2: Note that both receivers decode the cloud center based on Z^n, as Z^n is available also at Rx_1 through the cooperation link RC. Rx_1 now proceeds to decode the cloud element. However, since the channel is not physically degraded, then $Q(z^n \| y^n, x^n, s_0) \neq Q(z^n \| y^n, s_0)$ and/or $Q(y^n \| z^{n-1}, x^n, s_0) \neq Q(y^n \| x^n, s_0)$. Therefore, after decoding the cloud center using Z^n, Rx_1 uses both (Z^n, Y^n) rather than only Y^n to decode the cloud element. It should be noted that letting Rx_1 know Z^n is not equivalent to letting Rx_1 know M_2 since, when the channel is not degraded, Z^n may contain information on M_1 in addition to that contained in Y^n.

B. *Scenario Fb1-PD: The Physically Degraded FSBC With Only Switch Fb_1 Connected*

Here, Rx_2 does not send feedback hence the codebook is generated without knowledge of the channel output Z^n. The scenario is depicted in Fig. 4. For scenario Fb1-PD, we assume that the degradedness condition $X^n \,|\, s_0 \leftrightarrow Y^n \,|\, s_0 \leftrightarrow Z^n \,|\, s_0$ holds. Note that degradedness (10) is maintained also with feedback, as a consequence of (9). As noted in [11], feedback from just one user can sometimes help both receivers increase their rates. This is an important benefit of feedback in multiuser scenarios.

For scenario Fb1-PD, we obtain the following capacity region.

Theorem 2: Let $\mathcal{Q}_n^{SC2}$ be the set of all distributions of the form

$$\mathcal{Q}_n^{SC2} = \Big\{ q(U^n, X^n \,|\, Y^n) : \; q(u^n, x^n \,|\, y^n) = p(u^n) Q(x^n \| y^{n-1}; u^n) \text{ for all length } n \text{ sequences that belong to } \times_{i=1}^n \mathcal{U}_i \times \mathcal{X}^n \times \mathcal{Y}^n \text{ and } |\times_{i=1}^n \mathcal{U}_i| \leq \min\big\{ |\mathcal{X}|^n \cdot |\mathcal{Y}|^{n-1}, |\mathcal{Z}|^n \cdot |\mathcal{S}| \big\} + 2|\mathcal{S}| + 1 \Big\}, \qquad \forall n \in \mathbb{N}.$$

Define the region $\mathcal{R}_n^{SC2}$ as

$$\mathcal{R}_n^{SC2} = co \bigcup_{q_n \in \mathcal{Q}_n^{SC2}} \Big\{ (R_0, R_1, R_2) : R_0 \geq 0, R_1 \geq 0, R_2 \geq 0, \; R_1 \leq \min_{s_0' \in \mathcal{S}} \frac{1}{n} I(X^n \rightarrow Y^n \,|\, U^n, s_0')_{q_n}, \; R_0 + R_2 \leq \min_{s_0'' \in \mathcal{S}} \frac{1}{n} I(U^n; Z^n \,|\, s_0'')_{q_n} \Big\}.$$

For the physically degraded FSBC such that switch Fb_1 is connected and switches Fb_2 and RC are disconnected, the capacity region is given by

$$\mathcal{C}_{fb}^{SC2} = \lim_{n \to \infty} \mathcal{R}_n^{SC2}$$

and the limit exists and is finite.

Proof: The proofs of the achievability and the converse are similar to the proof of Theorem 1, only that decoding M_2 in Rx_1 is done as in [8, App. B-C]. The proof of the cardinality bound is provided in Appendix B. In Appendix C-A we provide a sketch of the converse proof. □

Since the capacity of the BC depends only on the conditional marginals $p(y^n, x^n \,|\, s_0)$ and $p(z^n, x^n \,|\, s_0)$ (see [29, Ch. 14.6]), the capacity region of the stochastically degraded FSBC for scenario Fb1-PD is the same as the corresponding physically degraded FSBC.

Corollary 1: For the stochastically degraded FSBC of Definition 7 with switch Fb_1 connected and switches Fb_2 and RC disconnected, the capacity region is given by Theorem 2.

Proof: We need to show that decoding M_2 at the strong receiver does not limit the rate, i.e., $\forall s_0 \in \mathcal{S}$ and for any $p_{U^n}(u^n)$, $I(U^n; Y^n \,|\, s_0) \geq I(U^n; Z^n \,|\, s_0)$. It is enough to show that

$$p(z^n \,|\, y^n, u^n, s_0) = p(z^n \,|\, y^n)$$

and the conclusion follows directly from the data processing inequality [29, Th. 2.8.1]. To that aim, we write

$$\begin{aligned} p(z^n, y^n, u^n \,|\, s_0) &= \sum_{\mathcal{X}^n} p(z^n, y^n, x^n, u^n \,|\, s_0) \\ &= \sum_{\mathcal{X}^n} p(y^n, x^n, u^n \,|\, s_0) p(z^n \,|\, y^n, x^n, u^n, s_0) \\ &\overset{(a)}{=} \sum_{\mathcal{X}^n} p(y^n, x^n, u^n \,|\, s_0) p(z^n \,|\, y^n, x^n, s_0) \\ &\overset{(b)}{=} \sum_{\mathcal{X}^n} p(y^n, x^n, u^n \,|\, s_0) \prod_{i=1}^n \tilde{p}(z_i \,|\, y_i) \\ &= p(y^n, u^n \,|\, s_0) \prod_{i=1}^n \tilde{p}(z_i \,|\, y_i) \end{aligned}$$

which gives the desired result. Here (a) follows from causality [22, eq. (8)] and (b) follows from (12). □

Due to the physically degraded structure, it is easy to see why feedback from the strong receiver helps both receivers: due to degradedness, the sum rate is bounded by the maximal rate at Rx_1. When this sum rate is increased, then for the same rate R_2 it is possible to obtain a higher rate R_1. Letting $s_0'^*$ be some arbitrary initial state, then

$$\begin{aligned}
&\min_{s_0'\in\mathcal{S}} I(X^n \to Y^n \,|\, U^n, s_0') + \min_{s_0''\in\mathcal{S}} I(U^n; Z^n \,|\, s_0'') \\
&\quad \le I(X^n \to Y^n \,|\, U^n, s_0'^*) + I(U^n; Z^n \,|\, s_0'^*) \\
&\quad \le I(X^n \to Y^n \,|\, U^n, s_0'^*) + I(U^n; Y^n \,|\, s_0'^*) \\
&\quad = \sum_{i=1}^{n} I(X^i; Y_i \,|\, Y^{i-1}, U^n, s_0'^*) + \sum_{i=1}^{n} I(U^n; Y_i \,|\, Y^{i-1}, s_0'^*) \\
&\quad = \sum_{i=1}^{n} I(U^n, X^i; Y_i \,|\, Y^{i-1}, s_0'^*) \\
&\quad = \sum_{i=1}^{n} I(X^i; Y_i \,|\, Y^{i-1}, s_0'^*) \\
&\quad = I(X^n \to Y^n \,|\, s_0'^*).
\end{aligned}$$

The maximum sum rate is always achievable by setting U^n atomic, $s_0'^*$ is the minimizing initial state in $\max_{Q(x^n\|y^{n-1})} \min_{s_0\in\mathcal{S}} I(X^n \to Y^n \,|\, s_0)$, and by continuity, we conclude that the entire region is increased.

It should be noted that when feedback is present it is not straightforward that the superposition construction achieves capacity, even when without feedback the channel is physically degraded. Generally speaking this is because feedback affects the degradedness of the channel. This is in contrast to the DMBC. We will discuss this in more details in Section V-A which considers feedback from both receivers without receiver cooperation.

C. Additional Comments

We make the following observations.

1) Note that in both scenarios we effectively have physically degraded situations (i.e., given the effective channel output at Rx_1, the channel output at Rx_2 is independent of the channel input from the transmitter). In scenario Fb1-Fb2-RC, physical degradedness is due to the cooperation link (switch RC). In scenario Fb1-PD, the channel is physically degraded by definition of the scenario. Physical degradedness implies that if the average probability of error for decoding information at Rx_2 can be made arbitrarily small, then the average probability of error for decoding the same information at Rx_1 can be made arbitrarily small as well. Therefore, in the proofs, it is enough to consider only the two private messages case as the common message can be incorporated by splitting the rate to Rx_2 into private and common rates, as in [29, Th. 14.6.4].
2) Computability of the regions: both capacity regions discussed above share the property that they are described in terms of a limiting region obtained by taking the block-length to infinity (the limiting region for each scenario exists and is finite). A natural question that arises in these situations is the feasibility of evaluating the capacity regions. The subject of computing the capacity of FSCs was considered by several authors. For example, Vontobel *et al.* [30] generalized the Blahut–Arimoto algorithm to indecomposable FSCs without feedback. In [31], the authors evaluated the capacity of finite-state ISI channels without feedback. These channels are strongly indecomposable (see [32]). We also note [33] in which an algorithm for computing the capacity of indecomposable FSCs with feedback under the assumption that the channel states can be computed from the channel inputs (e.g., ISI channels) was developed.
Clearly, for scenario Fb1-Fb2-RC, as we cannot provide a bound on the cardinality of the auxiliary random vector U^n, a finite computation of the capacity region is not possible. However, for scenario Fb1-PD, we provide a bound on $|\times_{i=1}^n \mathcal{U}_i|$, thus such a computation should be possible. We note that as in [34, Remark 1], the fact that the achievable regions are sup-additive implies that $\mathcal{R}_n^{SC1} \subseteq \mathcal{R}_{kn}^{SC1}$ and $\mathcal{R}_n^{SC2} \subseteq \mathcal{R}_{kn}^{SC2}$ for any positive integer k. This gives a finite-letter inner bound on the capacity region. As in [34, Remark 2], the existence of the limits implies that the capacity regions are computable but due to the lack of a finite-letter outer bound we do not know how close is the finite-letter achievable region to the capacity region. The difficulty in deriving an upper bound is that sub-additivity cannot be proved due to the memory in the codebook. It is possible to upper bound the capacity region by nontrivial finite-letter outer bounds, e.g., the sum-rate outer bound, but we were not able to show that such bounds converge to the capacity region.
3) Information-theoretic feedback represents the best possible feedback, thus it allows to identify the maximum possible gains from feedback. In practical systems, the feedback link has a limited capacity and may suffer from a substantial delay. We note that it is straightforward to extend the current achievability results to time-invariant deterministic feedback along the same lines considered in [5]. Such a deterministic function may incorporate quantization and delay, and provide an inner bound on the performance achievable under these constraints.
4) We note that the superposition construction provides inner bounds on the capacity region of nondegraded scenarios. For example, a codebook constructed according to Q_n^{SC2} can be used in conjunction with two decoders, one that decodes only the cloud centers and another that decodes both messages, to provide inner bounds for scenarios Fb1-Fb2 and Fb1 (see also Section V-A). Decoding both messages at one receiver will impose an additional rate constraint on the rate of the message that is decoded twice, resulting in an achievable region that is likely to be smaller than capacity.

V. Additional Feedback Configurations

In Fig. 1, there are eight possible configurations of feedback and receiver cooperation. The case where all switches are disconnected was considered in previous work [7], [8], and in the previous section, we described in detail the capacity regions of two additional scenarios. In the following, we address the capacity regions of the remaining scenarios. We derive explicitly the capacity for the cases in which it can be achieved with a superposition codebook. For each scenario, we modify Definitions

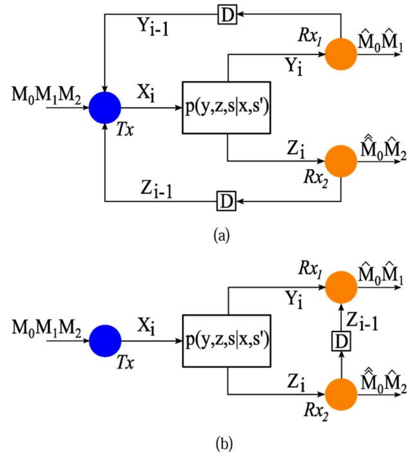

Fig. 5. Schematic description of scenarios (a) Fb1-Fb2 and (b) RC.

2 and 3 according to the configuration of feedback and receiver cooperation.

A. Scenario Fb1-Fb2: The General FSBC With Switches Fb_1 and Fb_2 Connected

In this scenario, the receivers do not cooperate directly, but each one sends its channel output back to the transmitter, as shown in Fig. 5(a). It turns out that using a superposition codebook with feedback in this case may actually result in lower rates compared to those achieved using a superposition codebook without feedback, even if the channel is degraded. This follows as a degraded channel with memory may become nondegraded when feedback is introduced. To see why, note that Z^n is not available at Rx_1. Therefore, if Rx_1 wishes to decode M_2 it must use a maximum-likelihood (ML) decoder without knowledge of feedback from Rx_2. Thus, the rate constraint on R_2 resulting from decoding M_2 at Rx_1 may be more constraining than the rate constraint for decoding M_2 at Rx_2. Decoding M_2 at Rx_2 takes place using the ML rule (see Appendix A-C)

$$\hat{m}_2 = \arg\max_{\tilde{m}_2} \sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} \prod_{i=1}^{n} p(z_i \,|\, z^{i-1}, u^i(\tilde{m}_2, z^{i-1}), s_0). \tag{13}$$

But, as Rx_1 does not know Z^{n-1} it has to consider all $|\,\mathcal{Z}\,|^{n-1}$ possible $\mathbf{u}$ sequences associated with each message $\tilde{m}_2 \in \mathcal{M}_2$ if it wishes to decode M_2. Therefore, the ML decoding rule will have to be

$$\begin{aligned}
\hat{m}_2 &= \arg\max_{\tilde{m}_2} \sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} p(y^n \,|\, \tilde{m}_2, s_0) \\
&= \arg\max_{\tilde{m}_2} \sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} \sum_{z^n \in \mathcal{Z}^n} p(y^n, z^n \,|\, \tilde{m}_2, s_0) \\
&= \arg\max_{\tilde{m}_2} \sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} \sum_{z^n \in \mathcal{Z}^n} \prod_{i=1}^{n} p(y_i, z_i \,|\, y^{i-1}, z^{i-1}, \tilde{m}_2, s_0) \\
&= \arg\max_{\tilde{m}_2} \sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} \sum_{z^n \in \mathcal{Z}^n} \prod_{i=1}^{n} p(y_i, z_i \,|\, y^{i-1}, z^{i-1}, \\
&\qquad\qquad u^i(\tilde{m}_2, z^{i-1}), s_0).
\end{aligned}$$

Thus, the decoder handles its lack of knowledge of Z^{n-1} in the same way it handles its lack of knowledge of the initial state—by averaging over all possibilities. As all $z^n \in \mathcal{Z}^n$ except the actual one result in a sequence u^n different from the one transmitted (with probability asymptotically approaching 1), the average detector may have a larger average probability of error than the detector (13). Specifically, the bounding technique employed in Appendix A will result in a lower rate bound. It should be noted that Rx_1 may employ other decoding rules that do not require decoding m_2 first, but since the feedback z^n is used in generating x^n, then decoding m_1 cannot be carried out without obtaining some information on z^n as well.

One possible way for using feedback in this scenario would be to find a set of $n-1$ pairs of functions $\{\alpha_k(\cdot)\}_{k=1}^{n-1}$ and $\{\beta_k(\cdot)\}_{k=1}^{n-1}$ with the property that for every pair of correlated vectors $(Z^n, Y^n) \sim p(z^n, y^n \,|\, s_0)$ we have $\alpha_k(Y^k) = \beta_k(Z^k) = a^k(Z^k, Y^k)$, $k = 1, 2, \ldots, n-1$. In this case, both the transmitter and Rx_2 generate the "cloud center" codetree with $a^k = \beta_k(z^k)$ instead of z^k and Rx_1 uses $\alpha_k(y^k)$, $k = 1, 2, \ldots, n-1$, when decoding m_2. The information $\alpha(\mathbf{y}) = \beta(\mathbf{z})$ is usually referred to as common information in the sense of Gács and Körner [35]. For *memoryless correlated sources*, such that there is a single pair of equivalence classes, one for each variable, Gács and Körner showed in [35] that unless there is some deterministic interdependence between the sources, then as n increases the common information per symbol decreases to zero, thus such a pair of functions cannot be used for asymptotically large blocklengths. The question whether the same conclusion holds also when there is memory, to the best of our knowledge, is an open one, and the answer is likely to depend on the structure of the memory.

B. Scenario RC: the General FSBC With Switch RC Connected

In this scenario, Rx_2 sends its channel output to Rx_1, but neither is sending feedback to the transmitter, as shown in Fig. 5(b). This type of cooperation is also called full cooperation; see [16]. For this scenario, we have the Markov chain

$$X^n | s_0 \leftrightarrow (Y^n, Z^n) \,|\, s_0 \leftrightarrow Z^n | s_0. \tag{14}$$

Therefore, whenever Rx_2 can decode M_2 with an arbitrarily small probability of error then Rx_1 can decode M_2 with an arbitrarily small probability of error as well. As the resulting channel is effectively physically degraded, the capacity region can be obtained from the capacity region of the discrete, physically degraded FSBC derived in [8, Th. 1].

Proposition 1 (Capacity Region for Scenario RC): Let $\mathcal{Q}_n^{SC4}$ be the set of all joint distributions on $\times_{i=1}^{n} \mathcal{U}_i \times \mathcal{X}^n$ such that $|\times_{i=1}^{n} \mathcal{U}_i| \leq \min\{|\mathcal{X}|^n, |\mathcal{Z}|^n \cdot |\mathcal{S}|\} + 2|\mathcal{S}| + 1$ and define the region $\mathcal{R}_n^{SC4}$ as

$$\begin{aligned}
\mathcal{R}_n^{SC4} = co \bigcup_{q_n \in \mathcal{Q}_n^{SC4}} \Big\{ &(R_0, R_1, R_2) : \\
&R_0 \geq 0, R_1 \geq 0, R_2 \geq 0, \\
&R_1 \leq \min_{s_0' \in \mathcal{S}} \frac{1}{n} I(X^n; Y^n, Z^n \,|\, U^n, s_0')_{q_n}, \\
&R_0 + R_2 \leq \min_{s_0'' \in \mathcal{S}} \frac{1}{n} I(U^n; Z^n \,|\, s_0'')_{q_n} \Big\}.
\end{aligned}$$

For the general FSBC such that switch RC is connected and Fb_1 and Fb_2 are disconnected, the capacity region is given by

$$\mathcal{C}_{fb}^{SC4} = \lim_{n\to\infty} \mathcal{R}_n^{SC4}$$

and the limit exists and is finite.

As in Fb1-Fb2-RC, the channel here is not physically degraded per symbol time [i.e., (6) does not hold for Y^n and Z^n] but the effective channel defined with $\tilde{Y}_i = (Y_i, Z_i)$ and $\tilde{Z}_i = Z_i$ is physically degraded.

C. Scenario Fb2-RC: the General FSBC With Switches Fb_2 and RC Connected

In this scenario, Rx_2 sends its channel output to both the transmitter and Rx_1, as shown in Fig. 6(a). Therefore, the resulting channel is again effectively a physically degraded one [(14) holds] and the transmitter receives feedback from the weak receiver. As Z^n is available to both receivers the transmitter can use the feedback to enhance the rates to both receivers. This is stated in the following proposition.

Proposition 2 (Capacity Region for Scenario Fb2-RC): Let $\mathcal{Q}_n^{SC5}$ be the set of all joint distributions on $\mathcal{U}^n \times \mathcal{X}^n$ for every given $\mathbf{z} \in \mathcal{Z}^n$, of the form

$$Q(u^n \| z^{n-1}) Q(x^n \| z^{n-1}, u^n)$$

and define the region $\mathcal{R}_n^{SC5}$ as

$$\mathcal{R}_n^{SC5} = co \bigcup_{q_n \in \mathcal{Q}_n^{SC5}} \Big\{ (R_0, R_1, R_2) : R_0 \geq 0, R_1 \geq 0, R_2 \geq 0, \quad R_1 \leq \min_{s_0' \in \mathcal{S}} \frac{1}{n} I(X^n \to Y^n, Z^n \| U^n, s_0')_{q_n}, \quad R_0 + R_2 \leq \min_{s_0'' \in \mathcal{S}} \frac{1}{n} I(U^n \to Z^n \mid s_0'')_{q_n} \Big\}. \tag{15}$$

For the general FSBC with feedback such that switches RC and Fb_2 are connected and Fb_1 is disconnected, the capacity region is given by

$$\mathcal{C}_{fb}^{SC5} = \lim_{n\to\infty} \mathcal{R}_n^{SC5}$$

and the limit exists and is finite.

In the achievability scheme, Rx_1 first decodes m_2 and then proceeds to decode m_1. The decoding rule for m_1, assuming correct decoding of m_2 at Rx_1, is

$$\begin{aligned} \hat{m}_1 &= \arg\max_{\tilde{m}_1} \sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} p(y^n, z^n \mid \tilde{m}_1, m_2, s_0) \\ &= \arg\max_{\tilde{x}^n(\tilde{m}_1, m_2; y^n, z^n)} \sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} \prod_{i=1}^{n} p(y_i, z_i \mid y^{i-1}, z^{i-1}, u^i, \tilde{x}^i, s_0) \\ &= \arg\max_{\tilde{x}^n(\tilde{m}_1, m_2; y^n, z^n)} \sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} Q(y^n, z^n \| \tilde{x}^n, s_0) \end{aligned}$$

leading to (15).

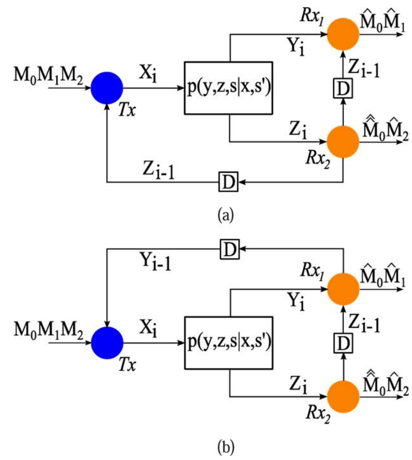

Fig. 6. Schematic description of scenarios (a) Fb2-RC and (b) Fb1-RC.

D. Scenario Fb1-RC: the General FSBC With Switches Fb_1 and RC Connected

Again due to RC connected the resulting channel is effectively physically degraded with *partial* feedback coming this time from the strong receiver, as shown in Fig. 6(b). This is because Rx_1 is restricted to sending only Y_{i-1} to the transmitter from Rx_2, thus $x_i = f_i(m_1, m_2, y^{i-1}, x^{i-1})$. The capacity region can be obtained from the result for Fb1-PD with (Y^n, Z^n) replacing Y^n in the rate constraint on R_1. The capacity is given by Theorem 2 with $\mathcal{R}_n^{SC2}$ replaced with $\mathcal{R}_n^{SC6}$ defined as

$$\mathcal{R}_n^{SC6} = \mathrm{co} \bigcup_{q_n \in \mathcal{Q}_n^{SC2}} \Big\{ (R_0, R_1, R_2) : R_0 \geq 0, R_1 \geq 0, R_2 \geq 0, \quad R_1 \leq \min_{s_0' \in \mathcal{S}} \frac{1}{n} I(X^n \to Y^n, Z^n \mid U^n, s_0')_{q_n}, \quad R_0 + R_2 \leq \min_{s_0'' \in \mathcal{S}} \frac{1}{n} I(U^n; Z^n \mid s_0'')_{q_n} \Big\}.$$

It should be noted that also without the constraint on Rx_1 this scenario is not equivalent to Fb1-Fb2-RC, due to the extra delay in which the transmitter is informed on Z_i. The scenario is also different from Fb1-PD since Y^n is not a degraded version of Z^n and therefore the feedback consists only of the y-part of the signal available at Rx_1.

E. Scenario Fb2-PD: the Physically Degraded FSBC With Switch Fb_2 Connected

As in the discussion on scenario Fb1-Fb2 in Section V-A, using feedback in conjunction with a superposition coding here may result in lower rates compared to a superposition codebook without feedback. This is because with feedback only from Rx_2 degradedness (6) does not lead to a relationship of the type (10).

VI. Conclusion

This paper extends our previous work on FSBCs without feedback or receiver cooperation [8]. We first derived the capacity regions for two FSBC configurations: Fb1-Fb2-RC and Fb1-PD. An important property of Fb1-Fb2-RC is that Z^n is also available to Rx_1. When this is not the case, i.e., Fb_1 and Fb_2 are connected but RC is disconnected (Fb1-Fb2), then

Rx_1 cannot use Z^n when decoding M_2. Therefore, despite the fact that feedback from Rx_2 is available at the transmitter, a superposition codetree may actually result in lower rates compared to a codebook that does not generate X^n using the feedback Z^n. This remains true even if the FSBC is physically degraded (Fb2-PD). In scenario Fb1-PD, as Z^n is not available at the transmitter and the channel is physically degraded, a superposition codebook in which the cloud centers are generated without feedback but *with memory* is again optimal. We also showed how to combine elements of the basic scenarios Fb1-Fb2-RC and Fb1-PD in order to obtain the capacity regions of three additional feedback configurations that include full unidirectional, receiver cooperation. In these configurations, a superposition codetree was shown to achieve capacity. Combined with [7] and [8], we obtained characterization of the capacity regions of four scenarios of the general FSBC and two scenarios of the physically degraded FSBC. Bounds on the cardinality of the auxiliary RVs were provided for some of these scenarios.

Appendix A
Proof of Theorem 1: Capacity Region of Scenario Fb1-Fb2-RC

In the proof, we will make frequent use of the following notations:

$$Q(z^n, y^n \| x^n, s_0) \triangleq \prod_{i=1}^{n} p(z_i, y_i \,|\, z^{i-1}, y^{i-1}, x^i, s_0) \quad \text{(A.1)}$$

$$Q(z^n \| u^n, s_0) \triangleq \prod_{i=1}^{n} p(z_i \,|\, u^i, z^{i-1}, s_0) \quad \text{(A.2)}$$

$$Q'_n(x^m \| y^{m-1}, z^{m-1}, u^m) \triangleq \prod_{i=1}^{m} p\left(x_{n+i} \,|\, x_{n+1}^{n+i-1}, y_{n+1}^{n+i-1}, z_{n+1}^{n+i-1}, u_{n+1}^{n+i}\right) \quad \text{(A.3)}$$

$$Q'_n(u^m \| z^{m-1}) \triangleq \prod_{i=1}^{m} p\left(u_{n+i} \,|\, u_{n+1}^{n+i-1}, z_{n+1}^{n+i-1}\right). \quad \text{(A.4)}$$

A. Outline of the Achievability Proof

1) Assume only two private messages, i.e., $R_0 = 0$. Fix the blocklength n and the distributions $\left\{p(u_i \,|\, u^{i-1}, z^{i-1}), p(x_i \,|\, x^{i-1}, y^{i-1}, z^{i-1}, u^i)\right\}_{i=1}^n \triangleq \{Q_1, Q_2\}$.

2) For every possible feedback sequence (y^{n-1}, z^{n-1}) construct a superposition codetree with (u^n, x^n) selected according to the distributions $\{Q_1, Q_2\}$.

3) Using ML decoding at Rx_2 according to

$$\hat{m}_2 = \arg\max_{\tilde{m}_2} \sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} \prod_{i=1}^{n} p(z_i \,|\, z^{i-1}, u^i(\tilde{m}_2, z^{i-1}), s_0)$$

we conclude that a positive error exponent for decoding M_2 can be obtained as long as $R_2 \le \min_{s_0 \in \mathcal{S}} \frac{1}{n} I(U^n \to Z^n \,|\, s_0) - \frac{\log_2 |\mathcal{S}|}{n} \triangleq I_2^n(Q_1, Q_2)$. As Rx_1 receives Z^n through the cooperation link, then also for decoding M_2 at Rx_1 the error exponent is positive.

4) Assuming correct decoding of M_2 at Rx_1, it decodes M_1 using the ML rule

$$\hat{m}_1 = \arg\max_{\tilde{m}_1} \sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} \prod_{i=1}^{n} p(y_i, z_i \,|\, x^i(y^{i-1}, z^{i-1}, \tilde{m}_1, u^i(m_2, z^{i-1})), y^{i-1}, z^{i-1}, s_0).$$

We conclude that a positive error exponent for decoding M_1 at Rx_1, given that M_2 was correctly decoded, can be achieved as long as $R_1 \le \min_{s_0 \in \mathcal{S}} \frac{1}{n} I(X^n \to Y^n, Z^n \| U^n, s_0) - \frac{\log_2 |\mathcal{S}|}{n} \triangleq I_1^n(Q_1, Q_2)$.

5) Next, for a fixed blocklength n and some positive integer k let $b = nk$. Construct extended distributions of length b by taking the product of Q_1 and Q_2, the basic distributions for a block of n symbols, k times, as described in (A.5) and (A.6), shown at the bottom of the page. Setting $R_1 = I_1^n(Q_1, Q_2)$, $R_2 = I_2^n(Q_1, Q_2)$, we show that taking k large enough the average probability of error can be made arbitrarily small, hence $(I_1^n(Q_1, Q_2), I_2^n(Q_1, Q_2))$ is achievable.

6) Finally, for $\lambda > 0$, define $C_{fb,SC1}^n(\lambda)$

$$C_{fb,SC1}^n(\lambda) = \max_{\substack{Q_1(u^n \| z^{n-1}) \\ Q_2(x^n \| z^{n-1}, y^{n-1}, u^n)}} F_n(\lambda, Q_1, Q_2) \quad \text{(A.7)}$$

$$F_n(\lambda, Q_1, Q_2) = \left\{ \min_{s''_0 \in \mathcal{S}} \frac{1}{n} I(U^n \to Z^n \,|\, s''_0)_{Q_1 Q_2} + \lambda \min_{s'_0 \in \mathcal{S}} \frac{1}{n} I(X^n \to Y^n, Z^n \| U^n, s'_0)_{Q_1 Q_2} \right\} - (1+\lambda) \frac{\log_2 |\mathcal{S}|}{n}.$$

We show that the achievable region for blocklength n is uniquely determined by $C_{fb,SC1}^n(\lambda)$, $0 \le \lambda \le 1$, and that $C_{fb,SC1}^n(\lambda)$ is sup-additive. Hence, $C_{fb,SC1}^\infty(\lambda) \triangleq$

$$\tilde{Q}_2(x^b \| y^{b-1}, z^{b-1}, u^b) = \prod_{k'=1}^{k} \prod_{i=1}^{n} p_{X_i \,|\, X^{i-1}, Y^{i-1}, Z^{i-1}, U^i} \left(x_{(k'-1)n+i} \,\middle|\, x_{(k'-1)n+1}^{(k'-1)n+i-1}, y_{(k'-1)n+1}^{(k'-1)n+i-1}, z_{(k'-1)n+1}^{(k'-1)n+i-1}, u_{(k'-1)n+1}^{(k'-1)n+i} \right) \quad \text{(A.5)}$$

$$\tilde{Q}_1(u^b \| z^{b-1}) = \prod_{k'=1}^{k} \prod_{i=1}^{n} p_{U_i \,|\, U^{i-1}, Z^{i-1}} \left(u_{(k'-1)n+i} \,\middle|\, u_{(k'-1)n+1}^{(k'-1)n+i-1}, z_{(k'-1)n+1}^{(k'-1)n+i-1} \right). \quad \text{(A.6)}$$

$\lim_{n\to\infty} C^n_{fb,SC1}(\lambda) = \sup_n C^n_{fb,SC1}(\lambda) < \infty$. We therefore conclude that the boundary of the largest achievable region can be written as

$$R_2(R_1) = \inf_{0\le\lambda\le 1} \left[C^{\infty}_{fb,SC1}(\lambda) - \lambda R_1\right] \tag{A.8}$$

completing the achievability part. Appendixes A-B–A-G present the achievability result and the converse is provided in Appendix A-H.

7) Finally, the common rate R_0 is incorporated by splitting the rate to the weak receiver Rx_2 into private and common parts as in [29, Th. 14.6.4].

B. Code Construction

Fix the blocklength n and the conditional distributions $\{p(u_i \mid u^{i-1}, z^{i-1})\}_{i=1}^n$ and $\{p(x_i \mid x^{i-1}, y^{i-1}, z^{i-1}, u^i)\}_{i=1}^n$.

- For every message $m_2 \in \mathcal{M}_2$ and feedback sequence $\mathbf{z} \in \mathcal{Z}^n$ generate a codetree of symbols $\mathbf{u}$, denoted $\mathbf{u}(m_2;\mathbf{z})$, according to the distribution $p(\mathbf{u}(m_2;\mathbf{z}) \mid \mathbf{z}) = \prod_{i=1}^n p(u_i \mid u^{i-1}, z^{i-1})$.
- For every message m_2 and feedback sequence from Rx_2, $(m_2, \mathbf{z}) \in \mathcal{M}_2 \times \mathcal{Z}^n$, every message $m_1 \in \mathcal{M}_1$ and feedback sequence $\mathbf{y} \in \mathcal{Y}^n$ generate a codetree of channel inputs $\mathbf{x}$ denoted $\mathbf{x}(m_1, m_2; \mathbf{y}, \mathbf{z})$ according to

$$\begin{aligned} &p(\mathbf{x}(m_1, m_2; \mathbf{z}, \mathbf{y}) \mid \mathbf{z}, \mathbf{y}, m_2) \\ &\quad = \prod_{i=1}^n p(x_i \mid x^{i-1}, y^{i-1}, z^{i-1}, u^i(m_2, z^{i-1})). \end{aligned} \tag{A.9}$$

For transmitting the message pair (m_1, m_2) the encoder outputs a sequence of n symbols generated according to the feedback sequence. The symbol at time i is $x_i(m_1, m_2, y^{i-1}, z^{i-1}) \equiv x_i(m_1, u^i, y^{i-1}, z^{i-1})$.

The average probability of error when the initial channel state is s_0 can be upper bounded by

$$\begin{aligned} P_e^{(n)}(s_0) &\le P_{e1}^{(n)}(s_0) + P_{e2}^{(n)}(s_0) \\ &\le \tilde{P}_{e1}^{(n)}(s_0) + P_{e2}^{(n)}(s_0) \\ &= P_{e12}^{(n)}(s_0) + P_{e11}^{(n)}(s_0) + P_{e2}^{(n)}(s_0) \end{aligned} \tag{A.10}$$

where $P_{ei}^{(n)}(s_0)$, $i = 1, 2$ is the average probability of error for decoding M_i at Rx_i when the initial state is s_0, $\tilde{P}_{e1}^{(n)}(s_0)$ is the probability of error for decoding *both* M_1 and M_2 at Rx_1, $P_{e12}^{(n)}(s_0)$ is the probability of error for decoding M_2 at Rx_1, and $P_{e11}^{(n)}(s_0)$ is the probability of error for decoding M_1 at Rx_1 assuming M_2 was correctly decoded at Rx_1.

C. Decoding at Rx_2

Rx_2 uses the "averaged ML" decoding rule

$$\hat{m}_2 = \arg\max_{\tilde{m}_2} \sum_{s_0\in\mathcal{S}} \frac{1}{|\mathcal{S}|} p(\mathbf{z} \mid \tilde{m}_2, s_0). \tag{A.11}$$

Note that this is not the exact ML rule as Rx_2 does not use available knowledge on the channel codewords $\mathbf{x}$, thus the distribution in (A.12) is derived by averaging over all codebooks generated according to $Q(x^n \| y^{n-1}, z^{n-1}, u^n)$. This is derived explicitly in (A.15). Expanding (A.11), we get

$$\hat{m}_2 = \arg\max_{\tilde{m}_2} \sum_{s_0\in\mathcal{S}} \frac{1}{|\mathcal{S}|} \prod_{i=1}^n p(z_i \mid z^{i-1}, u^i(\tilde{m}_2, z^{i-1}), s_0) \tag{A.12}$$

$$= \arg\max_{u^n(\tilde{m}_2; z^n)} \sum_{s_0\in\mathcal{S}} \frac{1}{|\mathcal{S}|} Q(z^n \| u^n, s_0). \tag{A.13}$$

The transition (A.12) follows since u^i represents the effect of the message $\tilde{m}_2$ on x^i and hence on z_i at time i (namely the contribution of $\tilde{m}_2$ to generating x_i is only through u^i). This implies the Markov chain $M_2 \leftrightarrow (U^i, Z^{i-1}) \leftrightarrow X^i$. Since from the causality of the channel we have $U^i \mid s_0 \leftrightarrow X^i \mid s_0 \leftrightarrow Z^i \mid s_0$, then averaging over X^i, we get $M_2 \mid s_0 \leftrightarrow (U^i, Z^{i-1}) \mid s_0 \leftrightarrow Z_i \mid s_0$ as we show formally in

$$\begin{aligned} &p(z_i \mid z^{i-1}, u^i, m_2, s_0) \\ &\quad = \frac{p(z^i, u^i \mid m_2, s_0)}{p(z^{i-1}, u^i \mid m_2, s_0)} \end{aligned} \tag{A.14}$$

$$\begin{aligned} &p(z^i, u^i \mid m_2, s_0) \\ &\quad = \sum_{\mathcal{Y}^i, \mathcal{S}^i, \mathcal{X}^i} p(z^i, y^i, s^i, x^i, u^i \mid m_2, s_0) \\ &\quad = \sum_{\mathcal{Y}^i, \mathcal{S}^i, \mathcal{X}^i} \prod_{j=1}^i p(u_j \mid z^{j-1}, y^{j-1}, s^{j-1}, x^{j-1}, u^{j-1}, m_2, s_0) \\ &\qquad \times \prod_{j=1}^i p(x_j \mid z^{j-1}, y^{j-1}, s^{j-1}, x^{j-1}, u^j, m_2, s_0) \\ &\qquad \times \prod_{j=1}^i p(z_j, y_j, s_j \mid s_{j-1}, x_j) \\ &\quad \overset{(a)}{=} \prod_{j=1}^i p(u_j \mid z^{j-1}, u^{j-1}, m_2) \\ &\qquad \times \sum_{\mathcal{Y}^i, \mathcal{S}^i, \mathcal{X}^i} \left[\prod_{j=1}^i p(x_j \mid z^{j-1}, y^{j-1}, x^{j-1}, u^j) \right. \\ &\qquad\qquad \left. \times \prod_{j=1}^i p(z_j, y_j, s_j \mid s_{j-1}, x_j) \right] \end{aligned} \tag{A.15}$$

$$\begin{aligned} &p(z^{i-1}, u^i \mid m_2, s_0) \\ &\quad = \sum_{z_i\in\mathcal{Z}} p(z^i, u^i \mid m_2, s_0). \end{aligned} \tag{A.16}$$

In (A.14)–(A.16), (a) follows from (A.9). We see that the dependence of the right-hand side of (A.14) on m_2 is only through an expression of the form $\prod_{j=1}^i p(u_j \mid z^{j-1}, u^{j-1}, m_2)$ which is outside the summation, both at the numerator and the denominator. Hence, when evaluating (A.14), this expression cancels out, giving $p(z_i \mid z^{i-1}, u^i, m_2, s_0) = p(z_i \mid z^{i-1}, u^i, s_0)$. The derivations above also show how to obtain $p(z_i \mid z^{i-1}, u^i, s_0)$ from the channel and the conditional p.m.f.s fixed in Appendix A-B.

We now analyze the probability of error of the decoder in (A.12), averaged over all codebooks and channel outputs. For given $m_2, m_2', z^n, u^n(m_2; z^n)$ and s_0 define the pairwise error event

$$E_{m_2'} = \Big\{ \text{Rx}_2 \text{ decodes } m_2' \text{ when } z^n \text{ is received, } m_2 \neq m_2' \text{ is transmitted using } u^n(m_2; z^n) \text{ and the initial channel state is } s_0 \Big\}.$$

The probability of $E_{m_2'}$ averaged over all possible codewords $u'^n(m_2'; z^n)$, when using the decoder (A.12)[1] is given by

$$\begin{aligned}\Pr\left(E_{m_2'}\right) &\overset{(a)}{=} \sum_{u'^n(m_2';z^n)\in\times_{i=1}^n\mathcal{U}_i} Q(u'^n\|z^{n-1}) \\ &\quad\times \text{ind}\left(\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}p(z^n\,|\,m_2', s_0) > \sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}p(z^n\,|\,m_2, s_0)\right) \qquad \text{(A.17)}\\ &\overset{(b)}{\leq} \sum_{u'^n(m_2';z^n)\in\times_{i=1}^n\mathcal{U}_i} Q(u'^n\|z^{n-1}) \\ &\quad\times\left(\frac{\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}p(z^n\,|\,m_2', s_0)}{\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}p(z^n\,|\,m_2, s_0)}\right)^t \qquad \text{(A.18)}\end{aligned}$$

where (a) follows since $p(u'^n(m_2'; z^n)\,|\,z^n, m_2, s_0) = p(u'^n(m_2'; z^n)\,|\,z^n) = Q(u'^n\|z^{n-1})$ is the probability that the codeword u'^n was constructed for transmission of m_2' when the feedback is z^n, and $\text{ind}(A)$ is the indicator function for the event A. In (b), we use $t > 0$. Averaging over x^n, y^n and s^n is implied in $p(z_i\,|\,z^{i-1}, u^i, s_0)$; see (A.15). Next, the probability of error, averaged over all codebooks, when m_2 is transmitted, z^n is the feedback sequence and $u^n(m_2; z^n)$ is used for transmission, is given by

$$\begin{aligned}P_e^{(n)}(m_2\,|\,z^n, u^n, s_0) &= \Pr\left(\bigcup_{m_2'\neq m_2} E_{m_2'}\right)\\ &\overset{(a)}{\leq} (\,|\mathcal{M}_2|\,-1)^\rho\Bigg[\sum_{u'^n(m_2';z^n)\in\times_{i=1}^n\mathcal{U}_i} Q(u'^n\|z^{n-1}) \\ &\qquad\times\left(\frac{\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}Q(z^n\|u'^n, s_0)}{\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}Q(z^n\|u^n, s_0)}\right)^t\Bigg]^\rho\end{aligned}$$

where in (a) we have $0 < \rho \leq 1$ [2, Lemma, p. 136] and we also used (A.13). Finally, the probability of error averaged over all channel outputs and codebooks is given by

$$\begin{aligned}P_e^{(n)}(m_2\,|\,s_0) &= \sum_{\mathcal{X}^n\times\mathcal{Y}^n\times\mathcal{S}^n\times\mathcal{Z}^n\times_{i=1}^n\mathcal{U}_i}\Big[p(u^n, z^n, y^n, x^n, s^n\,|\,m_2, s_0) \\ &\qquad\times P_e^{(n)}(m_2\,|\,z^n, u^n, s_0)\Big]\\ &= \sum_{(u^n,z^n)\in\times_{i=1}^n\mathcal{U}_i\times\mathcal{Z}^n} p(u^n, z^n\,|\,m_2, s_0)\\ &\quad\times P_e^{(n)}(m_2\,|\,z^n, u^n, s_0).\end{aligned}$$

[1]This decoder averages over all x^n, y^n and s^n sequences; see (A.15).

Now write

$$\begin{aligned}p(u^n, z^n\,|\,m_2, s_0) &= \prod_{i=1}^n p(u_i, z_i\,|\,u^{i-1}, z^{i-1}, m_2, s_0)\\ &= \prod_{i=1}^n p(u_i\,|\,u^{i-1}, z^{i-1}, m_2, s_0)\prod_{i=1}^n p(z_i\,|\,u^i, z^{i-1}, m_2, s_0)\\ &= \prod_{i=1}^n p(u_i\,|\,u^{i-1}, z^{i-1}, m_2)\prod_{i=1}^n p(z_i\,|\,u^i, z^{i-1}, s_0)\\ &= Q(u^n\|z^{n-1})Q(z^n\|u^n, s_0).\end{aligned}$$

Thus

$$\begin{aligned}P_{e2}^{(n)}(m_2\,|\,s_0) &\leq (\,|\mathcal{M}_2|\,-1)^\rho \\ &\quad\times\sum_{(u^n,z^n)\in\times_{i=1}^n\mathcal{U}_i\times\mathcal{Z}^n} Q(u^n\|z^{n-1})Q(z^n\|u^n, s_0)\\ &\quad\times\Bigg[\sum_{u'^n(m_2';z^n)\in\times_{i=1}^n\mathcal{U}_i} Q(u'^n\|z^{n-1})\\ &\qquad\times\left(\frac{\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}Q(z^n\|u'^n, s_0)}{\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}Q(z^n\|u^n, s_0)}\right)^t\Bigg]^\rho\\ &\overset{(a)}{\leq} |\mathcal{S}|^{1+\rho}(|\mathcal{M}_2|-1)^\rho\\ &\quad\times\max_{s_0\in\mathcal{S}}\sum_{z^n\in\mathcal{Z}^n}\left[\sum_{u^n\in\times_{i=1}^n\mathcal{U}_i} Q(u^n\|z^{n-1})Q(z^n\|u^n, s_0)^{\frac{1}{1+\rho}}\right]^{1+\rho}\end{aligned}$$

where in (a) we used $t = \frac{1}{1+\rho}$ and followed the standard transitions see [2, Ch. 5.9]. Let

$$\begin{aligned}&E_{2,n}(\rho, Q(u^n\|z^{n-1}), Q(x^n\|y^{n-1}, z^{n-1}, u^n), s_0)\\ &\quad= -\frac{1}{n}\log_2\sum_{\mathcal{Z}^n}\left[\sum_{\times_{i=1}^n\mathcal{U}_i} Q(u^n\|z^{n-1})Q(z^n\|u^n, s_0)^{\frac{1}{1+\rho}}\right]^{1+\rho}. \qquad \text{(A.19)}\end{aligned}$$

Since the bound on $P_{e2}^{(n)}(m_2\,|\,s_0)$ is independent of the actual values of s_0 and m_2, we can bound the probability of error averaged over all messages and codebooks by

$$\bar{P}_{e2}^{(n)} \leq |\mathcal{S}|2^{-n(F_{2,n}(\rho, Q(u^n\|z^{n-1}), Q(x^n\|y^{n-1}, z^{n-1}, u^n))-\rho R_2)} \qquad \text{(A.20)}$$

$$\begin{aligned}&F_{2,n}(\rho, Q(u^n\|z^{n-1}), Q(x^n\|y^{n-1}, z^{n-1}, u^n))\\ &\quad= \min_{s_0\in\mathcal{S}} E_{2,n}\Big(\rho, Q(u^n\|z^{n-1}),\\ &\qquad\qquad Q(x^n\|y^{n-1}, z^{n-1}, u^n), s_0\Big) - \rho\frac{\log_2|\mathcal{S}|}{n}.\end{aligned}$$

Using the derivation for the point-to-point FSC [5], we conclude from (A.19) that as long as

$$R_2 < \min_{s_0\in\mathcal{S}}\frac{1}{n}I(U^n \to Z^n\,|\,s_0) - \frac{\log_2|\mathcal{S}|}{n} \qquad \text{(A.21)}$$

there exists a $\rho^* > 0$ for which the error exponent is positive, i.e.,

$$F_{2,n}(\rho^*, Q(u^n\|z^{n-1}), Q(x^n\|y^{n-1}, z^{n-1}, u^n)) - \rho^* R_2 > 0. \tag{A.22}$$

We will give a more detailed derivation when establishing the rate constraint on R_1 in Appendix A-E. The rate constraint (A.21) can be obtained by essentially repeating the same steps.

Now we show a key property of convergence.

Lemma A.1: For all positive integers m and l, let the distribution of the codetrees u^{m+l}, x^{m+l} for given feedback sequences z^{m+l} and y^{m+l} be given by

$$Q(u^m\|z^{m-1})Q'_m(u^l\|z^{l-1}) \times Q(x^m\|y^{m-1}, z^{m-1}, u^m)Q'_m(x^l\|y^{l-1}, z^{l-1}, u^l).$$

Let $Q(u^l\|z^{l-1})$ be a downconverted version of $Q'_m(u^l\|z^{l-1})$ (i.e., subtract m from all indices) and similarly let $Q(x^l\|y^{l-1}, z^{l-1}, u^l)$ be the downconverted version of $Q'_m(x^l\|y^{l-1}, z^{l-1}, u^l)$. Then

$$\begin{aligned}(m+l)&F_{2,m+l}(\rho, Q(u^m\|z^{m-1})Q'_m(u^l\|z^{l-1}),\\ &Q(x^m\|y^{m-1}, z^{m-1}, u^m)Q'_m(x^l\|y^{l-1}, z^{l-1}, u^l))\\ &\geq mF_{2,m}(\rho, Q(u^m\|z^{m-1}), Q(x^m\|y^{m-1}, z^{m-1}, u^m))\\ &\quad + lF_{2,l}(\rho, Q(u^l\|z^{l-1}), Q(x^l\|y^{l-1}, z^{l-1}, u^l)).\end{aligned} \tag{A.23}$$

Proof: First, note that

$$\begin{aligned}Q&(z^{m+l}\|u^{m+l}, s_0)\\ &= \prod_{i=1}^{m+l} p(z_i \,|\, z^{i-1}, u^i, s_0)\\ &= Q(z^m\|u^m, s_0)\prod_{j=m+1}^{m+l} p(z_j \,|\, z^{j-1}, u^j, m_2, s_0)\\ &= Q(z^m\|u^m, s_0)\prod_{j=m+1}^{m+l} p(z_j \,|\, z^{j-1}, m_2, s_0)\\ &= Q(z^m\|u^m, s_0)\sum_{s_m\in\mathcal{S}} \big[p(s_m \,|\, z^m, m_2, s_0)\\ &\qquad\qquad \times p\left(z_{m+1}^{m+l} \,\big|\, z^m, s_m, m_2, s_0\right)\big]\\ &= Q(z^m\|u^m, s_0)\sum_{s_m\in\mathcal{S}} \Bigg[p(s_m \,|\, z^m, u^m, m_2, s_0)\\ &\qquad\qquad \times \prod_{i=1}^{l} p(z_{m+i}|z^{m+i-1}, s_m, m_2, s_0)\Bigg]\\ &\overset{(a)}{=} Q(z^m\|u^m, s_0)\sum_{s_m\in\mathcal{S}} \Bigg[p(s_m \,|\, z^m, u^m, s_0)\\ &\qquad\qquad \times \prod_{i=1}^{l} p(z_{m+i}|z^{m+i-1}, s_m, u^{m+i}, s_0)\Bigg]\\ &= Q(z^m\|u^m, s_0)\sum_{s_m\in\mathcal{S}} p(s_m \,|\, z^m, u^m, s_0)Q'_m(z^l\|u^l, s_m).\end{aligned}$$

To see (a), write

$$p(s_m \,|\, z^m, u^m, m_2, s_0) = \frac{p(s_m, z^m, u^m \,|\, m_2, s_0)}{p(z^m, u^m \,|\, m_2, s_0)} \tag{A.24}$$

$$\begin{aligned}p&(s_m, z^m, u^m \,|\, m_2, s_0)\\ &= \sum_{\mathcal{X}^m, \mathcal{Y}^m, \mathcal{S}^{m-1}} p(s^m, y^m, x^m, z^m, u^m \,|\, m_2, s_0)\\ &= \sum_{\mathcal{X}^m, \mathcal{Y}^m, \mathcal{S}^{m-1}} \prod_{i=1}^{m} p(u_i \,|\, y^{i-1}, x^{i-1}, z^{i-1}, u^{i-1}, m_2)\\ &\qquad \times p(x_i \,|\, y^{i-1}, x^{i-1}, z^{i-1}, u^i, m_2)\\ &\qquad \times \prod_{i=1}^{m} p(s_i, y_i, z_i \,|\, s_{i-1}, x_i)\\ &= \prod_{i=1}^{m} p(u_i \,|\, z^{i-1}, u^{i-1}, m_2)\\ &\quad \times \sum_{\mathcal{X}^m, \mathcal{Y}^m, \mathcal{S}^{m-1}} \Bigg[\prod_{i=1}^{m} p(x_i \,|\, y^{i-1}, x^{i-1}, z^{i-1}, u^i)\\ &\qquad\qquad \times \prod_{i=1}^{m} p(s_i, y_i, z_i \,|\, s_{i-1}, x_i)\Bigg].\end{aligned} \tag{A.25}$$

Note that m_2 appears outside the summation. Therefore, also in $p(z^m, u^m \,|\, m_2, s_0) = \sum_{s_m\in\mathcal{S}} p(s_m, z^m, u^m \,|\, m_2, s_0)$, m_2 appears outside the summation. Hence, evaluating the division in (A.24), we conclude that $p(s_m \,|\, z^m, u^m, m_2, s_0) = p(s_m \,|\, z^m, u^m, s_0)$, independent of m_2. In step (a), we also used the fact that by construction, $z_{m+i} \perp\!\!\!\perp u^m$ when s_m is given. To see this, consider the transitions in (A.26) and (A.27), shown on the next page. Note that z^m, u^m, m_2, and s_0 appear only in the first two terms, thus they will cancel out in the division in (A.26), resulting in

$$p(z_{m+i} \,|\, z^{m+i-1}, u^{m+i}, s_m, m_2, s_0) = p(z_{m+i} \,|\, z_{m+1}^{m+i-1}, u_{m+1}^{m+i}, s_m). \tag{A.28}$$

Denote with $s_{0,m}$ the state that achieves the minimum in $F_{2,m}(\rho, Q(u^m\|z^{m-1}), Q(x^m\|y^{m-1}, z^{m-1}, u^m))$, with $s_{0,l}$ the state that achieves the minimum in $F_{2,l}(\rho, Q(u^l\|z^{l-1}), Q(x^l\|y^{l-1}, z^{l-1}, u^l))$, and with $s_{0,l+m}$ the state that minimizes

$$F_{2,l+m}(\rho, Q(u^m\|z^{m-1})Q'_m(u^l\|z^{l-1}), Q(x^m\|y^{m-1}, z^{m-1}, u^m)Q'_m(x^l\|y^{l-1}, z^{l-1}, u^l)).$$

We now have the transition in (A.29)–(A.30), shown on the next page, where in (a) we used $(\sum_i a_i)^r \leq \sum_i a_i^r, a_i \geq 0, 0 < r \leq 1$ [36, Sec. 2.10, Th. 19] and $(\sum_i P_i a_i)^r \leq \sum_i P_i a_i^r, a_i \geq 0, r \geq 1$ and $\{P_i\}$ is a p.m.f. [36, Sec. 2.9, Th. 16]. Finally, the transitions in (A.31)–(A.32), shown at the bottom of next page, provide the desired result. In (A.31)–(A.32), (a) follows from Minkowski's inequality: $\sum_k (\sum_j Q_j a_{jk}^{\frac{1}{r}})^r \leq (\sum_j Q_j (\sum_k a_{jk})^{\frac{1}{r}})^r, r \geq 1, a_{jk} \geq 0, \{Q_j\}$ is a p.m.f. [36, Sec. 2.11, Th. 24]. □

$$p(z_{m+i}\,|\,z^{m+i-1}, s_m, u^{m+i}, m_2, s_0) = \frac{p(z^{m+i}, u^{m+i}, s_m\,|\,m_2, s_0)}{p(z^{m+i-1}, u^{m+i}, s_m\,|\,m_2, s_0)} \tag{A.26}$$

$$\begin{aligned} p(z^{m+i}, u^{m+i}, s_m\,|\,m_2, s_0) &= \sum_{\mathcal{Y}^{m+i}, \mathcal{S}^{m+i}\setminus\mathcal{S}_m, \mathcal{X}^{m+i}} p(z^{m+i}, y^{m+i}, s^{m+i}, x^{m+i}, u^{m+i}\,|\,m_2, s_0)\\ &= \sum_{\mathcal{Y}^{m+i}, \mathcal{S}^{m+i}\setminus\mathcal{S}_m, \mathcal{X}^{m+i}} p(z^m, y^m, s^m, x^m, u^m\,|\,m_2, s_0)\prod_{j=1}^{i} p\left(u_{m+j}\,|\,z_{m+1}^{m+j-1}, u_{m+1}^{m+j-1}, m_2\right)\\ &\quad\times\prod_{j=1}^{i} p\left(x_{m+j}\,|\,z_{m+1}^{m+j-1}, y_{m+1}^{m+j-1}, x_{m+1}^{m+j-1}, u_{m+1}^{m+j}\right)\prod_{j=1}^{i} p(z_{m+j}, y_{m+j}, s_{m+j}\,|\,s_{m+j-1}, x_{m+j})\\ &= p(z^m, s_m, u^m\,|\,m_2, s_0)\prod_{j=1}^{i} p(u_{m+j}\,|\,z_{m+1}^{m+j-1}, u_{m+1}^{m+j-1}, m_2)\\ &\quad\times\sum_{\mathcal{Y}_{m+1}^{m+i}, \mathcal{S}_{m+1}^{m+i}, \mathcal{X}_{m+1}^{m+i}}\prod_{j=1}^{i} p\left(x_{m+j}\,|\,z_{m+1}^{m+j-1}, y_{m+1}^{m+j-1}, x_{m+1}^{m+j-1}, u_{m+1}^{m+j}\right)\prod_{j=1}^{i} p(z_{m+j}, y_{m+j}, s_{m+j}\,|\,s_{m+j-1}, x_{m+j}). \end{aligned} \tag{A.27}$$

$$\begin{aligned} &2^{-(m+l)F_{2,m+l}(\rho, Q(u^m\|z^{m-1})Q'_m(u^l\|z^{l-1}), Q(x^m\|y^{m-1}, z^{m-1}, u^m)Q'_m(x^l\|y^{l-1}, z^{l-1}, u^l))}\\ &= |\mathcal{S}|^{\rho}\sum_{z^{m+l}\in\mathcal{Z}^{m+l}}\Bigg[\sum_{u^{m+l}\in\times_{i=1}^{m+l}\mathcal{U}_i} Q(u^m\|z^{m-1})Q'_m(u^l\|z^{l-1})\\ &\qquad\times\Big[Q(z^m\|u^m, s_{0,l+m})\sum_{s_m\in\mathcal{S}} p(s_m\,|\,z^m, u^m, s_{0,l+m})Q'_m(z^l\|u^l, s_m)\Big]^{\frac{1}{1+\rho}}\Bigg]^{1+\rho} \end{aligned} \tag{A.29}$$

$$\begin{aligned} &\overset{(a)}{\le} |\mathcal{S}|^{2\rho}\sum_{s_m\in\mathcal{S}}\sum_{z^{m+l}\in\mathcal{Z}^{m+l}}\Bigg[\sum_{u^{m+l}\in\times_{i=1}^{m+l}\mathcal{U}_i} Q(u^m\|z^{m-1})Q'_m(u^l\|z^{l-1})\\ &\qquad\times\Big[Q(z^m\|u^m, s_{0,l+m})p(s_m\,|\,z^m, u^m, s_{0,l+m})Q'_m(z^l\|u^l, s_m)\Big]^{\frac{1}{1+\rho}}\Bigg]^{1+\rho} \end{aligned} \tag{A.30}$$

$$\begin{aligned} &2^{-(m+l)F_{2,m+l}(\rho, Q(u^m\|z^{m-1})Q'_m(u^l\|z^{l-1}), Q(x^m\|y^{m-1}, z^{m-1}, u^m)Q'_m(x^l\|y^{l-1}, z^{l-1}, u^l))}\\ &\le |\mathcal{S}|^{2\rho}\sum_{s_m\in\mathcal{S}}\sum_{z^m\in\mathcal{Z}^m}\sum_{z_{m+1}^{m+l}\in\mathcal{Z}_{m+1}^{m+l}}\Bigg[\sum_{u^m\in\times_{i=1}^{m}\mathcal{U}_i} Q(u^m\|z^{m-1})\Big[Q(z^m\|u^m, s_{0,l+m})p(s_m\,|\,z^m, u^m, s_{0,l+m})\Big]^{\frac{1}{1+\rho}}\\ &\qquad\times\sum_{u_{m+1}^{m+l}\in\times_{i=m+1}^{m+l}\mathcal{U}_i} Q'_m(u^l\|z^{l-1})Q'_m(z^l\|u^l, s_m)^{\frac{1}{1+\rho}}\Bigg]^{1+\rho} \end{aligned} \tag{A.31}$$

$$\begin{aligned} &\le |\mathcal{S}|^{\rho}\sum_{s_m\in\mathcal{S}}\sum_{z^m\in\mathcal{Z}^m}\Bigg[\sum_{u^m\in\times_{i=1}^{m}\mathcal{U}_i} Q(u^m\|z^{m-1})\Big[Q(z^m\|u^m, s_{0,l+m})p(s_m\,|\,z^m, u^m, s_{0,l+m})\Big]^{\frac{1}{1+\rho}}\Bigg]^{1+\rho}\\ &\quad\times 2^{-lF_{2,l}(\rho, Q(u^l\|z^{l-1}), Q(x^l\|y^{l-1}, z^{l-1}, u^l))}\\ &\overset{(a)}{\le} |\mathcal{S}|^{\rho}\sum_{z^m\in\mathcal{Z}^m}\Bigg[\sum_{u^m\in\times_{i=1}^{m}\mathcal{U}_i} Q(u^m\|z^{m-1})\Big[\sum_{s_m\in\mathcal{S}} Q(z^m\|u^m, s_{0,l+m})p(s_m\,|\,z^m, u^m, s_{0,l+m})\Big]^{\frac{1}{1+\rho}}\Bigg]^{1+\rho}\\ &\quad\times 2^{-lF_{2,l}(\rho, Q(u^l\|z^{l-1}), Q(x^l\|y^{l-1}, z^{l-1}, u^l))}\\ &\le 2^{-mF_{2,m}(\rho, Q(u^m\|z^{m-1}), Q(x^m\|y^{m-1}, z^{m-1}, u^m))}2^{-lF_{2,l}(\rho, Q(u^l\|z^{l-1}), Q(x^l\|y^{l-1}, z^{l-1}, u^l))}. \end{aligned} \tag{A.32}$$

Now we have the following: let

$$Q(u^n\|z^{n-1})Q(x^n\|z^{n-1},y^{n-1},u^n)$$

be a distribution of blocklength n. Let k be some positive integer and consider blocklength $b=kn$. Construct a distribution on $\mathcal{U}^b\times\mathcal{X}^b$ for every feedback $(z^{b-1},y^{b-1})\in\mathcal{Z}^{b-1}\times\mathcal{Y}^{b-1}$ by multiplying the basic distribution n times

$$\begin{aligned}\tilde{Q}(u^b\|z^{b-1})\tilde{Q}(x^b\|z^{b-1},y^{b-1},u^b)\\ =\prod_{j=1}^{k}Q'_{(j-1)n}(u^n\|z^{n-1})Q'_{(j-1)n}(x^n\|z^{n-1},y^{n-1},u^n)\end{aligned}$$

as done in Appendix A-A, item 5. Then, applying Lemma A.1 k times, we have

$$\begin{aligned}bF_{2,b}(\rho,\tilde{Q}(u^b\|z^{b-1}),\tilde{Q}(x^b\|z^{b-1},y^{b-1},u^b))\\ \geq bF_{2,n}(\rho,Q(u^n\|z^{n-1}),Q(x^n\|z^{n-1},y^{n-1},u^n)).\end{aligned}\quad\text{(A.33)}$$

This property will be used when verifying achievability in Appendix A-F.

D. Decoding m_2 at Rx_1

As the channel output at $\mathrm{Rx}_2,\mathbf{z}$, is made available to Rx_1 through the unidirectional cooperation link, Rx_1 can decode m_2 by applying the same decoding method used by Rx_2. Hence, (A.21) guarantees that there exists a $\rho>0$ for which the error exponent for decoding m_2 at Rx_1 is positive. The error exponent is in fact stated in (A.22).

E. Decoding m_1 at Rx_1

Decoding m_1 takes place after m_2 was decoded at Rx_1. Assuming correct decoding of m_2,Rx_1 uses both $\mathbf{y}$ and $\mathbf{z}$ when decoding m_1 via the ML rule

$$\begin{aligned}\hat{m}_1&=\arg\max_{\tilde{m}_1}\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}p(y^n,z^n\,|\,\tilde{m}_1,m_2,s_0)\\ &\overset{(a)}{=}\arg\max_{\tilde{m}_1}\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}\prod_{i=1}^{n}p(y_i,z_i\,|\,y^{i-1},z^{i-1},\tilde{m}_1,m_2,s_0)\\ &\overset{(b)}{=}\arg\max_{\tilde{m}_1}\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}\prod_{i=1}^{n}p(y_i,z_i\,|\,y^{i-1},z^{i-1},\\ &\qquad\qquad x^i(\tilde{m}_1,m_2;y^{i-1},z^{i-1}),s_0)\\ &=\arg\max_{\tilde{x}^n(\tilde{m}_1,m_2;y^n,z^n)}\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}Q(z^n,y^n\|\tilde{x}^n,s_0).\end{aligned}$$

In (a), we note that since the channel $X^n\,|\,s_0-(Y^n,Z^n)\,|\,s_0$ is *not* degraded, $\mathbf{z}$ adds information on $\mathbf{y}$, and therefore, it is used in the decoder, and (b) follows from the causality of the channel. Assuming correct decoding at Rx_2 we need to evaluate $\bar{P}_{e11}^{(n)}(m_1\,|\,m_2,s_0)$, the average probability of error for decoding m_1 when the initial state is s_0. We begin by considering the pairwise error event, evaluated given $m_1,m'_1,z^n,y^n,x^n(m_1,m_2;y^n,z^n)$ and $u^n(m_2;z^n)$

$$\begin{aligned}E_{m'_1}\triangleq\Big\{&\mathrm{Rx}_1\text{ decodes }m'_1\text{ when }y^n\text{ and }z^n\text{ are received,}\\ &(m_1,m_2)\text{ is transmitted using }x^n(m_1,m_2;y^n,z^n)\\ &\text{and }u^n(m_2;z^n)\text{, the correct }m_2\text{ is available}\\ &\text{at }\mathrm{Rx}_1\text{ and the initial channel state is }s_0\Big\}.\end{aligned}$$

The probability of $E_{m'_1}$ averaged over all possible selections of $x'^n(m'_1,m_2;y^n,z^n)$ is given by (A.34), shown at the bottom of the page. Note that

$$\begin{aligned}\mathrm{ind}\Bigg(&\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}p(z^n,y^n\,|\,m'_1,m_2,s_0)\\ &>\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}p(z^n,y^n\,|\,m_1,m_2,s_0)\Bigg)\\ &\leq\frac{\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}Q(z^n,y^n\|x'^n,s_0)}{\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}Q(z^n,y^n\|x^n,s_0)}\end{aligned}$$

and that

$$\begin{aligned}&\Pr(x'^n(m'_1,m_2;y^n,z^n)\text{ is the codetree for}\\ &\quad(m'_1,m_2,y^n,z^n)\,|\,m_2,y^n,z^n,u^n(m_2;z^n))\\ &\quad=Q(x'^n\|z^{n-1},y^{n-1},u^n).\end{aligned}$$

Thus, for any $t>0$, we obtain the bound

$$\begin{aligned}\Pr(E_{m'_1}\,|\,&m_1,m_2,y^n,z^n,x^n(m_1,m_2;y^n,z^n),u^n(m_2;z^n),s_0)\\ \leq&\sum_{x'^n\in\mathcal{X}^n}Q(x'^n\|z^{n-1},y^{n-1},u^n)\\ &\times\left(\frac{\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}Q(z^n,y^n\|x'^n,s_0)}{\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}Q(z^n,y^n\|x^n,s_0)}\right)^t\end{aligned}\quad\text{(A.35)}$$

as we note that x'^n is a dummy variable for summation [2, p. 137]. Note that u^n is fixed since m_2 is assumed to be correctly decoded. Therefore, the average probability of error given

$$\begin{aligned}&\Pr(E_{m'_1}\,|\,m_1,m_2,y^n,z^n,x^n(m_1,m_2;y^n,z^n),u^n(m_2;z^n),s_0)\\ &=\sum_{x'^n(m'_1,m_2;y^n,z^n)\in\mathcal{X}^n}\Pr\left(x'^n(m'_1,m_2;y^n,z^n)\text{ is the codetree for }(m'_1,m_2,y^n,z^n)\,|\,m_2,y^n,z^n,u^n(m_2;z^n)\right)\\ &\quad\times\mathrm{ind}\left(\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}p(z^n,y^n\,|\,m'_1,m_2,s_0)>\sum_{s_0\in\mathcal{S}}\frac{1}{|\mathcal{S}|}p(z^n,y^n\,|\,m_1,m_2,s_0)\right).\end{aligned}\quad\text{(A.34)}$$

$m_1, m_2, y^n, z^n, x^n(m_1, m_2; y^n, z^n), u^n(m_2; z^n)$ and s_0 can be bounded by the union bound

$$\Pr(\text{error}|m_1, m_2, y^n, z^n, x^n(m_1, m_2; y^n, z^n), u^n(m_2; z^n), s_0) \stackrel{(a)}{\le} \left[(|\mathcal{M}_1| - 1) \sum_{x'^n \in \mathcal{X}^n} Q(x'^n \| z^{n-1}, y^{n-1}, u^n) \times \left(\frac{\sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} Q(z^n, y^n \| x'^n, s_0)}{\sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} Q(z^n, y^n \| x^n, s_0)} \right)^t \right]^\rho \quad \text{(A.36)}$$

where $0 < \rho \le 1$, and (a) follows from [2, Lemma, p. 136]. The average probability of error, averaged over all selections of "cloud centers" $\mathbf{u}$, channel codewords $\mathbf{x}$, and received signals $\mathbf{y}$ and $\mathbf{z}$ is given by (A.37), shown at the bottom of the page. Note that

$$\begin{aligned} p(y^n, z^n, u^n, x^n \,|\, m_1, m_2, s_0) &\stackrel{(a)}{=} \prod_{i=1}^n p(u_i, x_i \,|\, y^{i-1}, z^{i-1}, u^{i-1}, x^{i-1}, m_1, m_2, s_0) \times p(y_i, z_i \,|\, y^{i-1}, z^{i-1}, u^i, x^i, s_0) \\ &\stackrel{(b)}{=} \prod_{i=1}^n p(u_i \,|\, y^{i-1}, z^{i-1}, u^{i-1}, x^{i-1}, m_2) \times p(x_i \,|\, y^{i-1}, z^{i-1}, u^i, x^{i-1}, m_1) \times p(y_i, z_i \,|\, y^{i-1}, z^{i-1}, x^i, s_0) \\ &\stackrel{(c)}{=} Q(u^n \| z^{n-1}) Q(x^n \| z^{n-1}, y^{n-1}, u^n) Q(y^n, z^n \| x^n, s_0) \end{aligned}$$

where (a) follows from causality: once the sequences $u^i, x^i, y^{i-1}, z^{i-1}$ are fixed, the outputs y_i, z_i are independent of the messages, (b) is because when x^i is given, the channel outputs are independent of u^i, and (c) follows from the code construction procedure. Therefore

$$\begin{aligned} \bar{P}_{e11}^{(n)}(m_1 \,|\, m_2, s_0) &\le |\mathcal{S}| (|\mathcal{M}_1| - 1)^\rho \sum_{\mathbf{u}(m_2; \mathbf{z}), \mathbf{z}, \mathbf{y}} Q(u^n \| z^{n-1}) \times \sum_{\mathcal{X}^n} \left[Q(x^n \| z^{n-1}, y^{n-1}, u^n) \times \left(\sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} Q(y^n, z^n \| x^n, s_0) \right)^{1 - t\rho} \right] \\ &\quad \times \left[\sum_{x'^n \in \mathcal{X}^n} Q(x'^n \| z^{n-1}, y^{n-1}, u^n) \times \left(\sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} Q(z^n, y^n \| x'^n, s_0) \right)^t \right]^\rho. \end{aligned}$$

Letting $t = \frac{1}{1+\rho}$, we have

$$\bar{P}_{e11}^{(n)}(m_1 \,|\, m_2, s_0) \stackrel{(a)}{\le} (|\mathcal{M}_1| - 1)^\rho |\mathcal{S}|^{\rho+1} \max_{s_0 \in \mathcal{S}} \sum_{\times_{i=1}^n \mathcal{U}_i, \mathcal{Z}^n, \mathcal{Y}^n} Q(u^n \| z^{n-1}) \times \left[\sum_{\mathcal{X}^n} Q(x^n \| z^{n-1}, y^{n-1}, u^n) Q(z^n, y^n \| x^n, s_0)^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

where (a) follows the same steps leading to [8, eq. (B.6)], and $0 \le \rho \le 1$. Note that the bound is independent of s_0. Defining

$$\begin{aligned} &E_{n,11}(\rho, Q(u^n \| z^{n-1}), Q(x^n \| z^{n-1}, y^{n-1}, u^n), s_0) \\ &= -\frac{1}{n} \log_2 \sum_{\times_{i=1}^n \mathcal{U}_i, \mathcal{Z}^n, \mathcal{Y}^n} Q(u^n \| z^{n-1}) \times \left[\sum_{\mathcal{X}^n} Q(x^n \| z^{n-1}, y^{n-1}, u^n) \times Q(z^n, y^n \| x^n, s_0)^{\frac{1}{1+\rho}} \right]^{1+\rho} - \rho \frac{\log_2 |\mathcal{S}|}{n} \end{aligned} \quad \text{(A.38)}$$

$$\begin{aligned} &F_{n,11}(\rho, Q(u^n \| z^{n-1}), Q(x^n \| z^{n-1}, y^{n-1}, u^n)) \\ &= \min_{s_0 \in \mathcal{S}} E_{n,11}(\rho, Q(u^n \| z^{n-1}), Q(x^n \| z^{n-1}, y^{n-1}, u^n), s_0) \end{aligned} \quad \text{(A.39)}$$

the upper bound on the probability of error can be written as

$$\bar{P}_{e11}^{(n)}(m_1 \,|\, m_2) \le |\mathcal{S}| 2^{-n(F_{n,11}(\rho, Q(u^n \| z^{n-1}), Q(x^n \| z^{n-1}, y^{n-1}, u^n)) - \rho R_1)}. \quad \text{(A.40)}$$

As in [8, Lemma B.1], we can show that if $R_1 < \frac{\partial E_{n,11}(\rho, Q(u^n \| z^{n-1}), Q(x^n \| z^{n-1}, y^{n-1}, u^n), s_0)}{\partial \rho} |_{\rho=0}, \forall s_0 \in \mathcal{S}$, then there exists $0 < \rho \le 1$ for which

$$\bar{P}_{e11}^{(n)}(m_1 \,|\, m_2, s_0) \le |\mathcal{S}| (|\mathcal{M}_1| - 1)^\rho \sum_{\mathbf{u}(m_2; \mathbf{z}), \mathbf{z}, \mathbf{y}, \mathbf{x}(m_1, m_2; \mathbf{z}, \mathbf{y})} \sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} p(y^n, z^n, u^n, x^n \,|\, m_1, m_2, s_0) \times \left[\sum_{x'^n \in \mathcal{X}^n} Q(x'^n \| z^{n-1}, y^{n-1}, u^n) \left(\frac{\sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} Q(z^n, y^n \| x'^n, s_0)}{\sum_{s_0 \in \mathcal{S}} \frac{1}{|\mathcal{S}|} Q(z^n, y^n \| x^n, s_0)} \right)^t \right]^\rho. \quad \text{(A.37)}$$

$F_{n,11}(\rho, Q(u^n\|z^{n-1}), Q(x^n\|z^{n-1}, y^{n-1}, u^n)) - \rho R_1 > 0$. We now consider

$$\begin{aligned}
&\sum_{\mathcal{X}^n} Q(x^n\|z^{n-1}, y^{n-1}, u^n) Q(z^n, y^n\|x^n, s_0) \\
&= \sum_{\mathcal{X}^n} \prod_{i=1}^{n} \frac{p(z_i, y_i, x_i, u_i \,|\, y^{i-1}, z^{i-1}, x^{i-1}, u^{i-1}, m_2, s_0)}{p(u_i \,|\, u^{i-1}, y^{i-1}, z^{i-1}, x^{i-1}, m_2, s_0)} \quad \text{(A.41)}\\
&= \frac{p(z^n, y^n, u^n \,|\, m_2, s_0)}{\prod_{i=1}^{n} p(u_i \,|\, u^{i-1}, z^{i-1}, m_2)} \\
&= \frac{\prod_{i=1}^{n} p(z_i, y_i, u_i \,|\, z^{i-1}, y^{i-1}, u^{i-1}, m_2, s_0)}{\prod_{i=1}^{n} p(u_i \,|\, u^{i-1}, z^{i-1}, m_2)} \\
&\overset{(a)}{=} \prod_{i=1}^{n} p(z_i, y_i \,|\, z^{i-1}, y^{i-1}, u^i, s_0) \\
&\triangleq Q(z^n, y^n\|u^n, s_0) \quad \text{(A.42)}
\end{aligned}$$

where (a) follows from (A.14). Now we have from (A.41) and (A.42)

$$Q(u^n\|z^{n-1}) Q(z^n, y^n\|u^n, s_0) = p(u^n, y^n, z^n \,|\, s_0)$$

hence

$$\left.\frac{\partial E_{n,11}(\rho, Q(u^n\|z^{n-1}), Q(x^n\|z^{n-1}, y^{n-1}, u^n), s_0)}{\partial \rho}\right|_{\rho=0} + \frac{\log_2|\mathcal{S}|}{n} = \frac{1}{n} I(X^n \to Z^n, Y^n\|U^n, s_0)$$

and we therefore conclude that as long as

$$R_1 < \frac{1}{n} I(X^n \to Z^n, Y^n\|U^n, s_0) - \frac{\log_2|\mathcal{S}|}{n}$$

we can find $0 < \rho \leq 1$ such that $E_{n,11}(\rho, Q(u^n\|z^{n-1}), Q(x^n\|z^{n-1}, y^{n-1}, u^n), s_0) - \rho R_1 > 0$. Finally, taking the minimum over all $s_0 \in \mathcal{S}$, i.e.,

$$R_1 < \min_{s_0 \in \mathcal{S}} \frac{1}{n} I(X^n \to Z^n, Y^n\|U^n, s_0) - \frac{\log_2|\mathcal{S}|}{n}$$

allows obtaining a positive error exponent for every initial state s_0, and therefore also for $F_{n,11}(\rho, Q(u^n\|z^{n-1}), Q(x^n\|z^{n-1}, y^{n-1}, u^n)) - \rho R_1$.

The following lemma states an important property of $F_{n,11}(\rho, Q(u^n\|z^{n-1}), Q(x^n\|z^{n-1}, y^{n-1}, u^n))$.

Lemma A.2: For any positive integers m and l, it holds that

$$\begin{aligned}
&(m+l) F_{m+l,11}\Big(\rho, Q(u^m\|z^{m-1}) Q'_m(u^l\|z^{l-1}), \\
&\qquad Q(x^m\|z^{m-1}, y^{m-1}, u^m) Q'_m(x^l\|z^{l-1}, y^{l-1}, u^l)\Big) \\
&\geq m F_{m,11}(\rho, Q(u^m\|z^{m-1}), Q(x^m\|z^{m-1}, y^{m-1}, u^m)) \\
&\quad + l F_{l,11}(\rho, Q(u^l\|z^{l-1}), Q(x^l\|z^{l-1}, y^{l-1}, u^l)), \quad \text{(A.43)}
\end{aligned}$$

where $Q(u^l\|z^{l-1})$ and $Q(x^l\|z^{l-1}, y^{l-1}, u^l)$ are the downconverted versions of $Q'_m(u^l\|z^{l-1})$ and $Q'_m(x^l\|z^{l-1}, y^{l-1}, u^l)$, respectively.

Proof: Let s'_0 be the state that achieves the minimum in $F_{m+l,11}$. Thus we can write transitions (A.44)–(A.45), shown at the bottom of the next page, where step (b) follows similarly to step (a) in (A.30), and (c) is because in evaluating $F_{l,11}(\rho, Q(u^l\|z^{l-1}), Q(x^l\|z^{l-1}, y^{l-1}, u^l))$ the minimizing initial state is used. Step (a) follows from

$$\begin{aligned}
&Q(z^{m+l}, y^{m+l}\|x^{m+l}, s'_0) \\
&\overset{(a')}{=} Q(z^m, y^m\|x^m, s'_0) \\
&\quad \times \prod_{i=m+1}^{m+l} p(z_i, y_i \,|\, y^{i-1}, z^{i-1}, x^i, m_1, m_2, s'_0) \\
&\overset{(b')}{=} Q(z^m, y^m\|x^m, s'_0) \\
&\quad \times \sum_{s_m \in \mathcal{S}} p(s_m \,|\, y^m, z^m, x^m, s'_0) Q'_m(z^l, y^l\|x^l, s_m)
\end{aligned}$$

where (a') is because given $(X^i, Y^{i-1}, Z^{i-1}, s_0)$ then $(Y_i, Z_i) \,|\, X^i, Y^{i-1}, Z^{i-1}, s_0 \perp\!\!\!\perp (M_1, M_2)$ by causality of the channel [22, eq. (8)] and (b') follows from [5, Lemma 21].[2]

Continuing with (A.46), we obtain (A.47), shown at the bottom of next page, where (a) follows from Minkowski's inequality [36, Sec. 2.11, Th. 24]. This verifies (A.43). □

F. Achievability

For brevity of notation, we set $Q_1 \triangleq Q(u^n\|z^{n-1})$ and $Q_2 \triangleq Q(x^n\|z^{n-1}, y^{n-1}, u^n)$. Using (A.43), we show that for the general FSBC scenario Fb1-Fb2-RC, any rate pair (R_1, R_2) satisfying for some n

$$\begin{aligned}
R_1 &< \min_{s'_0 \in \mathcal{S}} \frac{1}{n} I(X^n \to Y^n, Z^n\|U^n, s'_0)_{Q_1 Q_2} - \frac{\log_2|\mathcal{S}|}{n} \\
&\triangleq R_{1,n}(Q_1, Q_2) \quad \text{(A.48a)}\\
R_2 &< \min_{s''_0 \in \mathcal{S}} \frac{1}{n} I(U^n \to Z^n \,|\, s''_0)_{Q_1 Q_2} - \frac{\log_2|\mathcal{S}|}{n} \\
&\triangleq R_{2,n}(Q_1, Q_2) \quad \text{(A.48b)}
\end{aligned}$$

[2] We can also show $p(s_m \,|\, y^m, z^m, x^m, m_1, m_2, s'_0) = p(s_m \,|\, y^m, z^m, x^m, s'_0)$

$$\begin{aligned}
&p(s_m \,|\, y^m, z^m, x^m, m_1, m_2, s'_0) \\
&\quad = \frac{p(s_m, y^m, z^m, x^m \,|\, m_1, m_2, s'_0)}{p(y^m, z^m, x^m \,|\, m_1, m_2, s'_0)} \\
&p(s_m, y^m, z^m, x^m \,|\, m_1, m_2, s'_0) \\
&\quad = \sum_{\mathcal{S}^{m-1}} p(s^m, y^m, z^m, x^m \,|\, m_1, m_2, s'_0) \\
&\quad = \prod_{i=1}^{m} p(x_i \,|\, y^{i-1}, z^{i-1}, x^{i-1}, m_1, m_2) \\
&\qquad \times \sum_{\mathcal{S}^{m-1}} \prod_{i=1}^{m} p(s_i, y_i, z_i \,|\, s_{i-1}, x_i).
\end{aligned}$$

Therefore, the division is independent of (m_1, m_2).

is achievable. It should be clear that if any of the rate expressions $R_{1,n}(Q_1,Q_2)$ or $R_{2,n}(Q_1,Q_2)$ is negative, we set it to zero.

To show achievability of $(R_{1,n}(Q_1,Q_2), R_{2,n}(Q_1,Q_2))$, we apply steps similar to those used in [6, Sec. VI] for the FS-MAC. Let $Q_1(u^n\|z^{n-1})Q_2(x^n\|z^{n-1},y^{n-1},u^n)$ be an input distribution for blocklength n, let k be some integer, and consider codewords of length $b=kn$ symbols. Construct a distribution on $\mathcal{U}^b\times\mathcal{X}^b$ by extending the basic distribution n times

$$\tilde{Q}(u^b\|z^{b-1})\tilde{Q}(x^b\|z^{b-1},y^{b-1},u^b) = \prod_{j=1}^{k} Q'_{(j-1)n}(u^n\|z^{n-1})Q'_{(j-1)n}(x^n\|z^{n-1},y^{n-1},u^n)$$

where

$$Q'_{(j-1)n}(u^n\|z^{n-1}) = \prod_{d=1}^{n} p_{U_d\,|\,U^{d-1},Z^{d-1}}\left(u_{(j-1)n+d}|u_{(j-1)n+1}^{(j-1)n+d-1}, z_{(j-1)n+1}^{(j-1)n+d-1}\right)$$

and similarly, $Q'_{(j-1)n}(x^n\|z^{n-1},y^{n-1},u^n)$ is given by (A.3). Then, applying (A.43) k times, we have

$$bF_{n,11}(\rho,\tilde{Q}(u^b\|z^{b-1}),\tilde{Q}(x^b\|z^{b-1},y^{b-1},u^b)) \geq bF_{n,11}\big(\rho,Q(u^n\|z^{n-1}),Q(x^n\|z^{n-1},y^{n-1},u^n)\big). \tag{A.49}$$

$$\begin{aligned}
&2^{-(m+l)F_{m+l,11}(\rho,Q(u^m\|z^{m-1})Q'_m(u^l\|z^{l-1}),Q(x^m\|z^{m-1},y^{m-1},u^m)Q'_m(x^l\|z^{l-1},y^{l-1},u^l))}\\
&= |\mathcal{S}|^\rho \sum_{\times_{k=1}^{m+l}\mathcal{U}_k,\mathcal{Z}^{m+l},\mathcal{Y}^{m+l}} Q(u^m\|z^{m-1})Q'_m(u^l\|z^{l-1})\\
&\quad\times\left[\sum_{\mathcal{X}^{m+l}} Q(x^m\|z^{m-1},y^{m-1},u^m)Q'_m(x^l\|z^{l-1},y^{l-1},u^l)Q(z^{m+l},y^{m+l}\|x^{m+l},s'_0)^{\frac{1}{1+\rho}}\right]^{1+\rho} &\text{(A.44)}\\
&\overset{(a)}{=} |\mathcal{S}|^\rho \sum_{\times_{k=1}^{m+l}\mathcal{U}_k,\mathcal{Z}^{m+l},\mathcal{Y}^{m+l}} Q(u^m\|z^{m-1})Q'_m(u^l\|z^{l-1})\\
&\quad\times\left[\sum_{\mathcal{X}^{m+l}} Q(x^m\|z^{m-1},y^{m-1},u^m)Q'_m(x^l\|z^{l-1},y^{l-1},u^l)\right.\\
&\quad\quad\left.\times\left[Q(z^m,y^m\|x^m,s'_0)\sum_{s_m\in\mathcal{S}} p(s_m\,|\,y^m,z^m,x^m,s'_0)Q'_m(z^l,y^l\|x^l,s_m)\right]^{\frac{1}{1+\rho}}\right]^{1+\rho}\\
&\overset{(b)}{\leq} |\mathcal{S}|^{2\rho}\sum_{s_m\in\mathcal{S}}\sum_{\times_{k=1}^{m+l}\mathcal{U}_k,\mathcal{Z}^{m+l},\mathcal{Y}^{m+l}} Q(u^m\|z^{m-1})Q'_m(u^l\|z^{l-1})\\
&\quad\times\left[\sum_{\mathcal{X}^{m+l}} Q(x^m\|z^{m-1},y^{m-1},u^m)Q'_m(x^l\|z^{l-1},y^{l-1},u^l)\right.\\
&\quad\quad\left.\times\left[Q(z^m,y^m\|x^m,s'_0)p(s_m\,|\,y^m,z^m,x^m,s'_0)Q'_m(z^l,y^l\|x^l,s_m)\right]^{\frac{1}{1+\rho}}\right]^{1+\rho}\\
&\overset{(c)}{\leq} |\mathcal{S}|^{\rho}\sum_{s_m\in\mathcal{S}}\sum_{\times_{k=1}^{m}\mathcal{U}_k,\mathcal{Z}^{m},\mathcal{Y}^{m}} Q(u^m\|z^{m-1})\\
&\quad\times\left[\sum_{\mathcal{X}^{m}} Q(x^m\|z^{m-1},y^{m-1},u^m)\Big[Q(z^m,y^m\|x^m,s'_0)p(s_m\,|\,y^m,z^m,x^m,s'_0)\Big]^{\frac{1}{1+\rho}}\right]^{1+\rho} 2^{-lF_{l,11}} &\text{(A.45)}
\end{aligned}$$

$$\begin{aligned}
&2^{-(m+l)F_{m+l,11}(\rho,Q(u^m\|z^{m-1})Q'_m(u^l\|z^{l-1}),Q(x^m\|z^{m-1},y^{m-1},u^m)Q'_m(x^l\|z^{l-1},y^{l-1},u^l))}\\
&\overset{(a)}{\leq} |\mathcal{S}|^\rho \sum_{\times_{k=1}^{m}\mathcal{U}_k,\mathcal{Z}^{m},\mathcal{Y}^{m}} Q(u^m\|z^{m-1})\\
&\quad\times\left[\sum_{\mathcal{X}^{m}} Q(x^m\|z^{m-1},y^{m-1},u^m)\left[\sum_{s_m\in\mathcal{S}} Q(z^m,y^m\|x^m,s'_0)p(s_m\,|\,y^m,z^m,x^m,s'_0)\right]^{\frac{1}{1+\rho}}\right]^{1+\rho} 2^{-lF_{l,11}} &\text{(A.46)}\\
&= |\mathcal{S}|^\rho \sum_{\times_{k=1}^{m}\mathcal{U}_k,\mathcal{Z}^{m},\mathcal{Y}^{m}} Q(u^m\|z^{m-1})\left[\sum_{\mathcal{X}^{m}} Q(x^m\|z^{m-1},y^{m-1},u^m)Q(z^m,y^m\|x^m,s'_0)^{\frac{1}{1+\rho}}\right]^{1+\rho} 2^{-lF_{l,11}}\\
&\leq 2^{-mF_{m,11}}2^{-lF_{l,11}} &\text{(A.47)}
\end{aligned}$$

We now bound the average probability of error for the length b codebook. Recalling that decoding m_2 at Rx_1 is done exactly as in Rx_2, we have

$$\begin{aligned} &P_e^{(n)} \\ &\le \Pr\Big(\{\text{Error decoding } m_2 \text{ at } \mathrm{Rx}_2\} \\ &\qquad \bigcup\{\text{Error decoding } m_2 \text{ at } \mathrm{Rx}_1\} \\ &\qquad \bigcup\{\text{Error decoding } m_1 \text{ at } \mathrm{Rx}_1\}\Big) \\ &\le 2\max_{s_0''\in\mathcal{S}} \bar{P}_{e2}^{(n)}(s_0'') + \max_{s_0'\in\mathcal{S}} \bar{P}_{e11}^{(n)}(s_0') \\ &\overset{(a)}{\le} 2|\mathcal{S}|2^{-b(F_{2,b}(\rho_2,\tilde{Q}(u^b\|z^{b-1}),\tilde{Q}(x^b\|z^{b-1},y^{b-1},u^b))-\rho_2 R_{2,n})} \\ &\quad + |\mathcal{S}|2^{-b(F_{b,11}(\rho_1,\tilde{Q}(u^b\|z^{b-1}),\tilde{Q}(x^b\|z^{b-1},y^{b-1},u^b))-\rho_1 R_{1,n})} \\ &\overset{(b)}{\le} 2|\mathcal{S}|2^{-kn(F_{2,n}(\rho_2,Q(u^n\|z^{n-1}),Q(x^n\|z^{n-1},y^{n-1},u^n))-\rho_2 R_2)} \\ &\quad + |\mathcal{S}|2^{-kn(F_{n,11}(\rho_1,Q(u^n\|z^{n-1}),Q(x^n\|z^{n-1},y^{n-1},u^n))-\rho_1 R_1})\end{aligned}$$

which can be made arbitrarily small by taking k large enough. Here the bounds in (a) come from (A.20) and (A.40), as these bounds are independent of the values of m_1, m_2m and s_0. ρ_1 and ρ_2 are selected such that error exponents for $b = n$ (i.e., $k = 1$) are positive. Such ρ's exist following the analysis in Appendixes A-C and A-E, as long as the rate bounds (A.48) are satisfied. The inequality in (b) follows from (A.33) and (A.49).

Finally note that since the average probability of error, averaged over all codebooks, can be made arbitrarily small by increasing k, there exists a sequence of codes whose average probability of error can be made arbitrarily small as the blocklength b increases. This complete the proof of achievability of the rates in (A.48).

The envelope of the achievable region for a fixed n can be written as a function $R_2(R_1)$ given by (see [37] and [38, Lemma 3])

$$R_2(R_1) = \inf_{\lambda\ge 0}\left(C_{fb,SC1}^n(\lambda) - \lambda R_1\right)$$

$$\begin{aligned} C_{fb,SC1}^n(\lambda) = \max_{Q_1,Q_2}\Big\{&\min_{s_0''\in\mathcal{S}} \frac{1}{n} I\left(U^n \to Z^n \,|\, s_0''\right)_{Q_1Q_2} \\ &+\lambda \min_{s_0'\in\mathcal{S}}\frac{1}{n} I(X^n \to Y^n, Z^n \| U^n, s_0')_{Q_1Q_2}\Big\} \\ &- (1+\lambda)\frac{\log_2|\mathcal{S}|}{n}. \end{aligned} \tag{A.50}$$

In the following, we assume that the channel supports positive rate pairs. Let

$$R_2^\lambda(R_1) \triangleq C_{fb,SC1}^n(\lambda) - \lambda R_1.$$

When $\lambda = 0$, we obtain

$$\begin{aligned} &R_2^0(R_1) \\ &= \max_{\substack{Q(u^n\|z^{n-1}),\\ Q(x^n\|u^n,z^{n-1},y^{n-1})}} \min_{s_0''\in\mathcal{S}} \frac{1}{n} I\left(U^n \to Z^n \,|\, s_0''\right) - \frac{\log_2|\mathcal{S}|}{n}. \end{aligned}$$

Consider $\lambda \ge 1$

$$\begin{aligned} &\max_{Q_1,Q_2}\Big\{\min_{s_0''\in\mathcal{S}}\frac{1}{n} I(U^n \to Z^n \,|\, s_0'')_{Q_1Q_2} \\ &\qquad +\lambda\min_{s_0'\in\mathcal{S}}\frac{1}{n} I(X^n \to Z^n, Y^n \| U^n, s_0')_{Q_1Q_2}\Big\} \\ &\le \max_{Q_1,Q_2}\Big\{\min_{s_0\in\mathcal{S}}\Big[\frac{1}{n} I(U^n \to Z^n \,|\, s_0)_{Q_1Q_2} \\ &\qquad +\lambda\frac{1}{n} I(X^n \to Z^n, Y^n \| U^n, s_0)_{Q_1Q_2}\Big]\Big\} \\ &\le \max_{Q_1,Q_2}\Big\{\min_{s_0\in\mathcal{S}}\lambda\Big[\frac{1}{n} I(U^n \to Z^n \,|\, s_0)_{Q_1Q_2} \\ &\qquad +\frac{1}{n} I(X^n \to Z^n, Y^n \| U^n, s_0)_{Q_1Q_2}\Big]\Big\} \\ &\overset{(a)}{\le} \max_{Q_1,Q_2}\Big\{\min_{s_0\in\mathcal{S}}\lambda\Big[\frac{1}{n} I\left(U^n \to Z^n, Y^n \,|\, s_0\right)_{Q_1Q_2} \\ &\qquad +\frac{1}{n} I(X^n \to Z^n, Y^n \| U^n, s_0)_{Q_1Q_2}\Big]\Big\} \\ &\overset{(b)}{=} \max_{Q_1,Q_2}\Big\{\min_{s_0\in\mathcal{S}}\lambda\sum_{i=1}^{n}\frac{1}{n} I(X^i; Z_i, Y_i \,|\, Y^{i-1}, Z^{i-1}, s_0)_{Q_1Q_2}\Big\} \\ &= \lambda\max_{Q_1,Q_2}\min_{s_0\in\mathcal{S}}\frac{1}{n} I(X^n \to Z^n, Y^n \,|\, s_0)_{Q_1Q_2} \end{aligned}$$

where (a) follows from

$$\begin{aligned} &I(U^n \to Z^n, Y^n \,|\, \tilde{s}_0) - I(U^n \to Z^n \,|\, \tilde{s}_0) \\ &= \sum_{i=1}^{n}(I(U^i; Z_i, Y_i \,|\, Z^{i-1}, Y^{i-1}, \tilde{s}_0) \\ &\qquad - I(U^i; Z_i \,|\, Z^{i-1}, \tilde{s}_0)) \\ &= \sum_{i=1}^{n}(H(U^i \,|\, Z^{i-1}, Y^{i-1}, \tilde{s}_0) - H(U^i \,|\, Z^i, Y^i, \tilde{s}_0) \\ &\qquad - H(U^i \,|\, Z^{i-1}, \tilde{s}_0) + H(U^i \,|\, Z^i, \tilde{s}_0)) \\ &= \sum_{i=1}^{n}(H(U^{i-1} \,|\, Z^{i-1}, Y^{i-1}, \tilde{s}_0) \\ &\qquad + H(U_i \,|\, U^{i-1}, Z^{i-1}, Y^{i-1}, \tilde{s}_0) \\ &\qquad - H(U^i \,|\, Z^i, Y^i, \tilde{s}_0) - H(U^{i-1} \,|\, Z^{i-1}, \tilde{s}_0) \\ &\qquad - H(U_i \,|\, U^{i-1}, Z^{i-1}, \tilde{s}_0) + H(U^i \,|\, Z^i, \tilde{s}_0)) \\ &\overset{(a')}{=} \sum_{i=1}^{n}(I(U^i; Y^i \,|\, Z^i, \tilde{s}_0) - I(U^{i-1}; Y^{i-1} \,|\, Z^{i-1}, \tilde{s}_0)) \\ &= I(U^n; Y^n \,|\, Z^n, \tilde{s}_0) \ge 0 \end{aligned}$$

where (a') holds since $H(U_i \,|\, U^{i-1}, Z^{i-1}, Y^{i-1}, \tilde{s}_0) = H(U_i \,|\, U^{i-1}, Z^{i-1}, \tilde{s}_0)$, and step (b) is because X^i is the channel input thus, by causality, we have the Markov chain

$$\begin{aligned} U^i \,|\, z^{i-1}, y^{i-1}, s_0 \leftrightarrow X^i \,|\, z^{i-1}, y^{i-1}, s_0 \\ \leftrightarrow (Z_i, Y_i) \,|\, z^{i-1}, y^{i-1}, s_0. \end{aligned}$$

Therefore, when $\lambda \ge 1$, the maximum over all distributions is independent of U^n and we can set U^n atomic

$$\begin{aligned} &C_{fb,SC1}^n(\lambda) \\ &\quad = \lambda\max_{Q(x^n\|z^{n-1},y^{n-1})}\min_{s_0'\in\mathcal{S}}\frac{1}{n} I(X^n \to Z^n, Y^n \,|\, s_0')_Q \\ &\qquad - (1+\lambda)\frac{\log_2|\mathcal{S}|}{n} \\ &R_2^\lambda(R_1) \\ &\quad = \lambda\left(C_{fb,SC1}^n(1) + \frac{\log_2|\mathcal{S}|}{n} - R_1\right) - \frac{\log_2|\mathcal{S}|}{n}. \end{aligned}$$

Clearly, when $\lambda > 1$, the bounding line passes to the right of the bounding line for $\lambda = 1$ (see [8, Fig. 5], and we conclude that it is enough to consider only $0 \le \lambda \le 1$, i.e., the envelope of the achievable region is given by

$$R_2(R_1) = \inf_{0\le\lambda\le1}\Big(C^n_{fb,SC1}(\lambda) - \lambda R_1\Big).$$

It is interesting to note that the *slope* of the achievable region at the intersections with the R_1-axis and the R_2-axis for the case with feedback is the same as in the case without feedback [8, App. B-F]. Thus, feedback does not change the slope of the region at the end points where either R_1 or R_2 is zero.

G. Convergence of the Achievable Region

In the last part of the achievability proof, we address the question of convergence of the achievable region: if the region does not converge to a limiting region then it is not possible to determine if the channel supports reliable communication in the standard sense of Definition 4.

To this aim, define $G_n(\lambda) \triangleq C^n_{fb,SC1}(\lambda)$. We will show that for $0 < \lambda \le 1, G_n(\lambda)$ is sup-additive and bounded, i.e., for every $n = l + m, l, m > 0$

$$nG_n(\lambda) \ge lG_l(\lambda) + mG_m(\lambda) \quad \text{(A.51a)}$$
$$G_n(\lambda) < M < \infty \quad \text{(A.51b)}$$

for some finite constant M. W.l.o.g. we can fix the cardinalities of all $\{\mathcal{U}_i\}_{i=1}^{m+l}$ to be the same, setting $\mathcal{U}_i = \mathcal{U}$.

- Let $\{Q_{1m}(u^m\|z^{m-1})Q_{2m}(x^m\|z^{m-1}, y^{m-1}, u^m), s''_{0,m}, s'_{0,m}\}$ achieve the max–min solution for $G_m(\lambda)$.
- Let $\{Q_{1l}(u^l\|z^{l-1})Q_{2l}(x^l\|z^{l-1}, y^{l-1}, u^l), s''_{0,l}, s'_{0,l}\}$ achieve the max–min solution for $G_l(\lambda)$.
- Define

$$\begin{aligned}G_n(\tilde{Q}_1\tilde{Q}_2, \lambda) &\triangleq \min_{s''_0\in\mathcal{S}} \frac{1}{n} I(U^n \to Z^n \,|\, s''_0)_{\tilde{Q}_1\tilde{Q}_2}\\ &+ \lambda \min_{s'_0\in\mathcal{S}} \frac{1}{n} I(X^n \to Y^n, Z^n\|U^n, s'_0)_{\tilde{Q}_1\tilde{Q}_2}\\ &- (1+\lambda)\frac{\log_2|\mathcal{S}|}{n}.\end{aligned}$$

We denote with $s''_{0,n}$ the initial state that minimizes $I(U^n \to Z^n \,|\, s_0)_{\tilde{Q}_1\tilde{Q}_2}$ and with $s'_{0,n}$ the initial state that minimizes $I(X^n \to Y^n, Z^n\|U^n, s_0)_{\tilde{Q}_1\tilde{Q}_2}$.
- Let

$$\tilde{Q}_1(u^n\|z^{n-1}) = Q_{1m}(u^m\|z^{m-1})Q'_{1l,m}(u^l\|z^{l-1})$$

where

$$Q'_{1l,m}(u^l\|z^{l-1}) = \prod_{i=1}^{l} p_{1l,U_i\,|\,U^{i-1},Z^{i-1}}(u_{m+i}\,|\,u^{m+i-1}_{m+1}, z^{m+i-1}_{m+1})$$

and $\{p_{1l}(\cdot)\}$ are the conditional distributions used in $Q_{1l}(u^l\|z^{l-1})$.
- Let

$$\begin{aligned}&\tilde{Q}_2(x^n\|u^n, z^{n-1}, y^{n-1})\\ &\quad = Q_{2m}(x^m\|u^m, z^{m-1}, y^{m-1})Q'_{2l,m}(x^l\|u^l, z^{l-1}, y^{l-1})\end{aligned}$$

where $Q'_{2l,m}(x^l\|u^l, z^{l-1}, y^{l-1})$ is defined in (A.52), shown at the bottom of the page, and $\{p_{2l}(\cdot)\}$ are the conditional distributions used in $Q_{2l}(x^l\|u^l, z^{l-1}, y^{l-1})$.
- Let $(s'_{0,n}, s''_{0,n})$ be the minimizing states in $G_n(\tilde{Q}_1\tilde{Q}_2, \lambda)$.

We now have

$$\begin{aligned}&(1+\lambda)\log_2|\mathcal{S}| + nG_n(\lambda)\\ &\quad\ge I(U^n \to Z^n \,|\, s''_{0,n})_{\tilde{Q}_1\tilde{Q}_2}\\ &\qquad + \lambda I(X^n \to Z^n, Y^n\|U^n, s'_{0,n})_{\tilde{Q}_1\tilde{Q}_2}\\ &\quad = \sum_{i=1}^{m} I(U^i; Z_i \,|\, Z^{i-1}, s''_{0,n})_{Q_{1m}Q_{2m}}\\ &\qquad + \lambda\sum_{i=1}^{m} I(X^i; Z_i, Y_i \,|\, Z^{i-1}, Y^{i-1}, U^i, s'_{0,n})_{Q_{1m}Q_{2m}}\\ &\qquad + \sum_{i=m+1}^{n} I(U^i; Z_i \,|\, Z^{i-1}, s''_{0,n})_{\tilde{Q}_1\tilde{Q}_2}\\ &\qquad + \lambda\sum_{i=m+1}^{n} I(X^i; Z_i, Y_i \,|\, Z^{i-1}, Y^{i-1}, U^i, s'_{0,n})_{\tilde{Q}_1\tilde{Q}_2}\\ &\quad \ge \sum_{i=1}^{m} I(U^i; Z_i \,|\, Z^{i-1}, s''_{0,m})_{Q_{1m}Q_{2m}}\\ &\qquad + \lambda\sum_{i=1}^{m} I(X^i; Z_i, Y_i \,|\, Z^{i-1}, Y^{i-1}, U^i, s'_{0,m})_{Q_{1m}Q_{2m}}\\ &\qquad + \sum_{i=m+1}^{n} I(U^i; Z_i \,|\, Z^{i-1}, s''_{0,n})_{\tilde{Q}_1\tilde{Q}_2}\\ &\qquad + \lambda\sum_{i=m+1}^{n} I(X^i; Z_i, Y_i \,|\, Z^{i-1}, Y^{i-1}, U^i, s'_{0,n})_{\tilde{Q}_1\tilde{Q}_2}.\end{aligned}$$

Thus

$$\begin{aligned}&nG_n(\lambda)\\ &\quad\ge mG_m(\lambda) + \sum_{i=m+1}^{n} I(U^i; Z_i \,|\, Z^{i-1}, s''_{0,n})_{\tilde{Q}_1\tilde{Q}_2}\\ &\qquad + \lambda\sum_{i=m+1}^{n} I(X^i; Z_i, Y_i \,|\, Z^{i-1}, Y^{i-1}, U^i, s'_{0,n})_{\tilde{Q}_1\tilde{Q}_2}\\ &\quad\overset{(a)}{\ge} mG_m(\lambda) + \sum_{i=m+1}^{n} I(U^i; Z_i \,|\, Z^{i-1}, S_m, s''_{0,n})_{\tilde{Q}_1\tilde{Q}_2}\end{aligned}$$

$$Q'_{2l,m}(x^l\|u^l, z^{l-1}, y^{l-1}) = \prod_{i=1}^{l} p_{2l,X_i\,|\,U^i,X^{i-1},Z^{i-1},Y^{i-1}}\left(x_{m+i}|\, u^{m+i}_{m+1}, x^{m+i-1}_{m+1}, z^{m+i-1}_{m+1}, y^{m+i-1}_{m+1}\right) \quad \text{(A.52)}$$

$$
\begin{aligned}
&+\lambda \sum_{i=m+1}^{n} I(X^i;Z_i,Y_i \,|\, Z^{i-1},Y^{i-1},U^i,S_m,s'_{0,n})_{\tilde{Q}_1\tilde{Q}_2} \\
&-(1+\lambda)\log_2|\mathcal{S}| \\
&\overset{(b)}{\geq} mG_m(\lambda) + \sum_{i=m+1}^{n} I(U^i_{m+1};Z_i \,|\, Z^{i-1},S_m,s''_{0,n})_{\tilde{Q}_1\tilde{Q}_2} \\
&+\lambda \sum_{i=m+1}^{n} I\big(X^i_{m+1};Z_i,Y_i|Z^{i-1},Y^{i-1},U^i,S_m,s'_{0,n}\big)_{\tilde{Q}_1\tilde{Q}_2} \\
&-(1+\lambda)\log_2|\mathcal{S}| \\
&\overset{(c)}{=} mG_m(\lambda) + \sum_{i=m+1}^{n} I\left(U^i_{m+1};Z_i \,|\, Z^{i-1}_{m+1},S_m,s''_{0,n}\right)_{\tilde{Q}_1\tilde{Q}_2} \\
&+\lambda \sum_{i=m+1}^{n} I\left(X^i_{m+1};Z_i,Y_i \,|\, Z^{i-1}_{m+1},Y^{i-1}_{m+1},\right. \\
&\qquad\qquad \left. U^i_{m+1},S_m,s'_{0,n}\right)_{\tilde{Q}_1\tilde{Q}_2} - (1+\lambda)\log_2|\mathcal{S}| \\
&\geq mG_m(\lambda) \\
&+ \sum_{i=m+1}^{n} I\left(U^i_{m+1};Z_i \,|\, Z^{i-1}_{m+1},S_m=s''_{0,l}\right)_{Q'_{1l,m}Q'_{2l,m}} \\
&+\lambda \sum_{i=m+1}^{n} I\left(X^i_{m+1};Z_i,Y_i \,|\, Z^{i-1}_{m+1},Y^{i-1}_{m+1},\right. \\
&\qquad\qquad \left. U^i_{m+1},S_m=s'_{0,l}\right)_{Q'_{1l,m}Q'_{2l,m}} \\
&-(1+\lambda)\log_2|\mathcal{S}| \\
&= mG_m(\lambda) + lG_l(\lambda)
\end{aligned}
$$

where (a) follows from Lemma A.3, (b) is because the number of RVs on the left-hand side of each of the mutual information expressions was reduced, and (c) follows from (A.28), [5, Lemma 21] and also from the fact that by construction $(U^i_{m+1},X^i_{m+1}) \perp\!\!\!\perp (U^m,X^m)$ when $(Z^{i-1}_{m+1},Y^{i-1}_{m+1})$ are given, and hence when $(Z^{i-1}_{m+1},Y^{i-1}_{m+1},S_m,U^i_{m+1})$ are given, $(Y_i,Z_i) \perp\!\!\!\perp (U^m,X^m)$, for $i>m$

$$
\begin{aligned}
&p(y_i,z_i \,|\, y^{i-1},z^{i-1},u^i,s_m,s_0) \\
&\qquad = p(y_i,z_i \,|\, y^{i-1}_{m+1},z^{i-1}_{m+1},u^i_{m+1},s_m,s_0).
\end{aligned}
$$

Finally, we give the statement of Lemma A.3.

Lemma A.3: Let (U^n,Z^n,X^n,Y^n,S) be a collection of RVs with an arbitrary joint distribution and $S\in\mathcal{S}, |\mathcal{S}|<\infty$. Then, the following holds.

R1)
$$
\left|\sum_{i=i_0}^{n} I(U^i;Z_i \,|\, Z^{i-1}) - \sum_{i=i_0}^{n} I(U^i;Z_i \,|\, Z^{i-1},S)\right| \leq \log_2|\mathcal{S}|.
$$

R2)
$$
\left|\sum_{i=i_0}^{n} I(X^i;Y_i,Z_i \,|\, U^i,Y^{i-1},Z^{i-1}) - \sum_{i=i_0}^{n} I(X^i;Y_i,Z_i \,|\, U^i,Y^{i-1},Z^{i-1},S)\right| \leq \log_2|\mathcal{S}|.
$$

Proof: The proof is similar to that of [5, Lemma 4] and will not be repeated here. □

The last step is to show that for $0\leq\lambda\leq 1, G_n(\lambda)\leq M<\infty$ for every n. This holds since

$$
\begin{aligned}
&\frac{1}{n}I(U^n\to Z^n \,|\, s''_0) + \lambda\frac{1}{n}I(X^n\to Z^n,Y^n \| U^n,s'_0) \\
&\quad = \frac{1}{n}\sum_{i=1}^{n} I(U^i;Z_i \,|\, Z^{i-1},s''_0) \\
&\qquad + \lambda\frac{1}{n}\sum_{i=1}^{n} I(X^i;Z_i,Y_i \,|\, Z^{i-1},Y^{i-1},U^i,s'_0) \\
&\quad \leq \log_2|\mathcal{Z}| + \lambda\log_2(|\mathcal{Z}|\cdot|\mathcal{Y}|) \\
&\quad \leq \log_2|\mathcal{Z}| + \log_2(|\mathcal{Z}|\cdot|\mathcal{Y}|).
\end{aligned}
$$

Thus, we obtain that (A.51) holds. The implication of (A.51) is that $nG_n(\lambda)$ is sup-additive, hence[3] $\lim_{n\to\infty} G_n(\lambda) = \sup_n G_n(\lambda) \leq M < \infty$, for $0\leq\lambda\leq 1$, completing the proof of convergence.

We therefore conclude that the envelope of the largest achievable region for scenario Fb1-Fb2-RC is given by

$$
R_2(R_1) = \inf_{0\leq\lambda\leq 1}[C^{\infty}_{fb,SC1}(\lambda) - \lambda R_1]
$$

where $C^{\infty}_{fb,SC1}(\lambda) \triangleq \lim_{n\to\infty} C^n_{fb,SC1}(\lambda)$. In Appendix A-H, we provide a converse that shows that this achievable region is in fact the capacity region.

H. Converse

Theorem A.1: For any (R_1,R_2,n) code such that $R_2+\lambda R_1 > C^{\infty}_{fb,SC1}(\lambda)+\epsilon$ for some $\lambda>0$, there exist initial states $s'_{0,n},s''_{0,n}\in\mathcal{S}$ for which

$$
P^{(n)}_{e2}(s''_{0,n})R_2 + \lambda P^{(n)}_{e1}(s'_{0,n})R_1 > \epsilon - (1+\lambda)\frac{1+\log_2|\mathcal{S}|}{n}. \tag{A.53}
$$

The implication of (A.53), as explained in [37], is that for n large enough the probability of error cannot be made arbitrarily small outside the region whose boundary is given by (A.8).

Proof: Recall that $P^{(n)}_{e2}(s_0)$ and $P^{(n)}_{e1}(s_0)$ denote probabilities of error for initial state s_0, when the encoder and decoders are ignorant of the initial state. From Fano's inequality [29, Th. 2.11.1], we have that for initial state s_0

$$
H(M_2 \,|\, Z^n,s_0) \leq P^{(n)}_{e2}(s_0)nR_2 + 1 \tag{A.54a}
$$
$$
H(M_1 \,|\, Z^n,Y^n,s_0) \leq P^{(n)}_{e1}(s_0)nR_1 + 1. \tag{A.54b}
$$

Consider a code of length n and denote with $s''_{0,n}$ the initial state that maximizes $H(M_2 \,|\, Z^n,s_0)$ and with $s'_{0,n}$ the initial state that maximizes $H(M_1 \,|\, Z^n,Y^n,s_0)$. Now, note that

$$
\begin{aligned}
&\min_{s_0\in\mathcal{S}} I(M_2;Z^n \,|\, s_0) && \text{(A.55)}\\
&\qquad = nR_2 - \max_{s_0\in\mathcal{S}} H(M_2 \,|\, Z^n,s_0) \\
&\min_{s_0\in\mathcal{S}} I(M_1;Z^n,Y^n \,|\, M_2,s_0) \\
&\qquad = nR_1 - \max_{s_0\in\mathcal{S}} H(M_1 \,|\, Y^n,Z^n,M_2,s_0) \\
&\qquad \geq nR_1 - \max_{s_0\in\mathcal{S}} H(M_1 \,|\, Y^n,Z^n,s_0). && \text{(A.56)}
\end{aligned}
$$

[3]Here we use [2, Lemma 4A.1] which states that for a sequence $\{a_N\}_{N\in\mathbb{N}}$, with $\bar{a}=\sup_N a_N<\infty$, if for all $n\geq 1$ and all $N>n, Na_N \geq na_n + (N-n)a_{N-n}$, then $\lim_{N\to\infty} a_N = \bar{a}$.

We next have

$$\begin{aligned}I(M_2; Z^n \mid s_0) &= \sum_{i=1}^{n}(H(Z_i \mid Z^{i-1}, s_0) - H(Z_i \mid Z^{i-1}, M_2, s_0)) \\ &= I(U^n \to Z^n \mid s_0)\end{aligned}$$

where $U_i = (M_2, Z^{i-1}), i = 1, 2, \ldots, n$. Note that $U_i \leftrightarrow (U^{i-1}, Z^{i-1}) \leftrightarrow (Y^{i-1}, X^{i-1}, S^{i-1}, s_0)$ and also

$$\begin{aligned}M_1, M_2, U^i \mid Z^{i-1}, Y^{i-1}, s_0 &\leftrightarrow X^i \mid Z^{i-1}, Y^{i-1}, s_0 \\ &\leftrightarrow Y_i, Z_i \mid Z^{i-1}, Y^{i-1}, s_0. \quad \text{(A.57)}\end{aligned}$$

Therefore, the selection of the auxiliary RV U^n induces an input distribution that satisfies

$$\begin{aligned}&p(u_i, x_i \mid u^{i-1}, x^{i-1}, z^{i-1}, y^{i-1}, s^{i-1}, s_0) \\ &\quad = p(u_i \mid u^{i-1}, z^{i-1})p(x_i \mid u^i, x^{i-1}, z^{i-1}, y^{i-1}).\end{aligned}$$

We also have that

$$\begin{aligned}I(M_1; Z^n, Y^n \mid M_2, s_0) &= \sum_{i=1}^{n}\big(H(Z_i, Y_i \mid Z^{i-1}, Y^{i-1}, M_2, s_0) \\ &\quad - H(Z_i, Y_i \mid Z^{i-1}, Y^{i-1}, M_1, M_2, s_0)\big) \\ &\overset{(a)}{\leq} \sum_{i=1}^{n}(H(Z_i, Y_i \mid Z^{i-1}, Y^{i-1}, U^i, s_0) \\ &\quad - H(Z_i, Y_i \mid Z^{i-1}, Y^{i-1}, X^i, U^i, s_0)) \\ &= I(X^n \to Z^n, Y^n \| U^n, s_0)\end{aligned}$$

where in (a) we used $U^i = (M_2, Z^{i-1})$, the fact that conditioning reduces entropy and the fact that the channel is causal (A.57).

Combining both derivations, we have that for our choice of U^n

$$\begin{aligned}&\min_{s_0'' \in \mathcal{S}} I(M_2; Z^n \mid s_0'') + \lambda \min_{s_0' \in \mathcal{S}} I(M_1; Z^n, Y^n \mid M_2, s_0') \\ &\quad \leq \min_{s_0'' \in \mathcal{S}} I(U^n \to Z^n \mid s_0'') \\ &\qquad + \lambda \min_{s_0' \in \mathcal{S}} I(X^n \to Z^n, Y^n \| U^n, s_0') \\ &\quad \overset{(a)}{\leq} nC_{fb,SC1}^{n}(\lambda) + (1+\lambda)\log_2|\mathcal{S}| \\ &\quad \overset{(b)}{\leq} nC_{fb,SC1}^{\infty}(\lambda) + (1+\lambda)\log_2|\mathcal{S}| \qquad \text{(A.58)}\end{aligned}$$

where (a) holds since $C_{fb,SC1}^{n}(\lambda)$ is obtained by maximizing over all joint distributions of the form $Q(u^n \| z^{n-1})Q(x^n \| z^{n-1}, y^{n-1}, u^n)$ and (b) is because $C_{fb,SC1}^{n}(\lambda)$ is sup-additive. It is clear from the derivation above that it is enough to consider $|\mathcal{U}^n| < \infty$.

Plugging (A.55) and (A.56) into (A.58) yields

$$\begin{aligned}&nR_2 - H(M_2 \mid Z^n, s_{0,n}'') + \lambda\big(nR_1 - H(M_1 \mid Z^n, Y^n, s_{0,n}')\big) \\ &\quad \leq nC_{fb,SC1}^{\infty}(\lambda) + (1+\lambda)\log_2|\mathcal{S}| \\ &\Rightarrow H(M_2 \mid Z^n, s_{0,n}'') + \lambda H(M_1 \mid Z^n, Y^n, s_{0,n}') \\ &\qquad + (1+\lambda)\log_2|\mathcal{S}| \\ &\quad \geq n\left(R_2 + \lambda R_1 - C_{fb,SC1}^{\infty}(\lambda)\right) > n\epsilon.\end{aligned}$$

Combined with Fano's inequalities (A.54), we obtain (A.53). This means that at least one of the states $s_{0,n}'', s_{0,n}'$ results in a probability of error (at the respective receiver) that is bounded away from zero, completing the proof of the converse. □

APPENDIX B
PROOF OF CARDINALITY BOUNDS FOR THEOREM 2 (SCENARIO FB1-PD)

Using the method of Ahlswede and Körner [38], [39], we bound the cardinality of the auxiliary RV U^n for Fb1-PD. The bound results from the following two lemmas.

Lemma B.1: Let

$$\begin{aligned}C_{fb,SC2}^{n}(\lambda) &= \max_{p(u^n)Q(x^n \| y^{n-1}; u^n)} \Bigg[\min_{s_0'' \in \mathcal{S}} \frac{1}{n} I\left(U^n; Z^n \mid s_0''\right) \\ &\qquad + \lambda \min_{s_0' \in \mathcal{S}} \frac{1}{n} I(X^n \to Y^n \mid U^n, s_0')\Bigg] \\ &\quad - (1+\lambda)\frac{\log_2|\mathcal{S}|}{n}. \qquad \text{(B.1)}\end{aligned}$$

Then, the following holds.

1) $\lim_{n\to\infty} C_{fb,SC2}^{n}(\lambda), \lambda \in [0,1]$ uniquely characterizes the capacity region for Fb1-PD.
2) $\forall \lambda \in [0,1], C_{fb,SC2}^{n}(\lambda)$ can be completely characterized with a RV U^n having a cardinality that is bounded by $|\times_{i=1}^{n} \mathcal{U}_i| \leq |\mathcal{X}|^n \cdot |\mathcal{Y}|^{n-1} + 2|\mathcal{S}| + 1$.

Proof: From [38, Lemma 3], similarly to the argument in Appendix A-F, it follows directly that $C_{fb,SC2}^{n}(\lambda)$ uniquely characterizes the achievable region for a given blocklength n. Combined with the converse provided in Appendix C-A and the fact that the achievable region is sup-additive, which follows along the lines of Appendix A-G, we obtain the first statement of the lemma.

We now show the second statement. Let $\tilde{\mathcal{X}}$ contain all the elements in the set $\mathcal{X}$ except one. Define the vector $\mathbf{p}(u^n)$ that contains all distributions in $Q(x^n \| y^{n-1}; u^n)$ as in (B.2), shown at the bottom of the page. Note that $\mathbf{p}(u^n)$ uniquely characterizes the causal distribution $\prod_{i=1}^{n} p(x_i \mid x^{i-1}, y^{i-1}, u^n)$. Let the

$$\begin{aligned}\mathbf{p}(u^n) = \Big(&\{p(x_1 = i \mid u^n)\}_{i \in \tilde{\mathcal{X}}}, \{p(x_2 = i \mid x_1 = j_1, y_1 = k_1; u^n)\}_{i \in \tilde{\mathcal{X}}, j_1 \in \mathcal{X}, k_1 \in \mathcal{Y}}, \ldots, \\ &\{p(x_n = i \mid x^{n-1} = j^{n-1}, y^{n-1} = k^{n-1}; u^n)\}_{i \in \tilde{\mathcal{X}}, (j_1, j_2, \ldots, j_{n-1}) \in \mathcal{X}^{n-1}, (k_1, k_2, \ldots, k_{n-1}) \in \mathcal{Y}^{n-1}}\Big). \qquad \text{(B.2)}\end{aligned}$$

set $\mathcal{P}$ be the collection of all causal distribution vectors $\mathbf{p}$. The set $\mathcal{P}$ is compact[4].

Recall that $C^n_{fb,SC2}(\lambda)$ is characterized by the quantities $\{I(X^n \to Y^n \mid U^n, s_0), I(U^n; Z^n \mid s_0)\}_{s_0 \in \mathcal{S}}$. Denote the maximizing distribution of $C^n_{fb,SC2}(\lambda)$ with $p^*(u^n)Q^*(x^n \| y^{n-1}; u^n)$, where $p^*(u^n)$ is a measure on U^n. Then, $p^*(u^n)$ induces a measure on $\mathcal{P}$. Denote this measure with $\mu^*(d\mathbf{p})$. Now, note that from $p^*(u^n)$ and $Q^*(x^n \| y^{n-1}; u^n)$, we can derive

$$\begin{aligned}
&p^*(y^n, z^n, s^n, x^n, u^n \mid s_0)\\
&\quad = p^*(u^n) p^*(y^n, z^n, s^n, x^n \mid u^n, s_0)\\
&\quad = p^*(u^n) \prod_{i=1}^n p^*(y_i, z_i, s_i, x_i | y^{i-1}, z^{i-1}, s^{i-1}, x^{i-1}, u^n, s_0)\\
&\quad = p^*(u^n) \prod_{i=1}^n p^*(x_i \mid y^{i-1}, z^{i-1}, s^{i-1}, x^{i-1}, u^n, s_0)\\
&\qquad \times p(y_i, z_i, s_i \mid y^{i-1}, z^{i-1}, s^{i-1}, x^i, u^n, s_0)\\
&\quad \overset{(a)}{=} p^*(u^n) Q^*(x^n \| y^{n-1}; u^n) \prod_{i=1}^n p(y_i, z_i, s_i \mid s_{i-1}, x_i)\\
&\quad \equiv \mu^*(d\mathbf{p}) Q(x^n \| y^{n-1}; \mathbf{p}) \prod_{i=1}^n p(y_i, z_i, s_i \mid s_{i-1}, x_i) \qquad \text{(B.3)}
\end{aligned}$$

where (a) follows from causality [22, eq. (8)], by viewing u^n as a sequence on which the receivers obtain information only through the channel inputs $\{x_i\}_{i=1}^n$. Note that

$$\begin{aligned}
&p^*(y^n, z^n, x^n \mid s_0)\\
&\quad = \sum_{\mathcal{S}^n} \int_{\times_{i=1}^n \mathcal{U}_i} p^*(y^n, z^n, s^n, x^n \mid u^n, s_0) dp^*(u^n)\\
&\quad = \left(\int_{\mathcal{P}} \mu^*(d\mathbf{p}) Q(x^n \| y^{n-1}; \mathbf{p}) \right) \sum_{\mathcal{S}^n} \prod_{i=1}^n p(y_i, z_i, s_i | s_{i-1}, x_i)
\end{aligned} \qquad \text{(B.4)}$$

$$\begin{aligned}
&p^*(z^n \mid s_0)\\
&\quad = \sum_{\mathcal{X}^n \times \mathcal{Y}^n} p^*(y^n, z^n, x^n \mid s_0). \qquad \text{(B.5)}
\end{aligned}$$

Therefore, knowledge of the distribution

$$\int_{\mathcal{P}} \mu^*(d\mathbf{p}) Q(x^n \| y^{n-1}; \mathbf{p})$$

[4] $\mathcal{P}$ is compact by the Krein–Milman theorem [40] since every element in $\mathcal{P}$ can be written as a convex combination of the extreme points of $\mathcal{P}$. To explain the meaning of an extreme point of $\mathcal{P}$ consider as an example a situation with $n = 2$, $|\mathcal{X}| = |\mathcal{Y}| = 2$. The set $\mathcal{P}$ of probability distributions is characterized by vectors of the form

$$\begin{aligned}
\mathbf{p} = \Big(&p(X_1 = 0), p(X_2 = 0 \mid Y_1 = 0, X_1 = 0),\\
&p(X_2 = 0 \mid Y_1 = 0, X_1 = 1),\\
&p(X_2 = 0 \mid Y_1 = 1, X_1 = 0),\\
&p(X_2 = 0 \mid Y_1 = 1, X_1 = 1) \Big).
\end{aligned}$$

The extreme points of $\mathcal{P}$ are all binary vectors of length 5 excluding the all-zero vector: $(1,0,0,0,0), (0,1,0,0,0), (0,0,1,0,0), (0,0,0,1,0), (0,0,0,0,1), (1,1,0,0,0), (0,1,1,0,0,), (1,0,1,0,0), \ldots, (1,1,1,1,1)$. Since every point in $\mathcal{P}$ can be obtained as a convex combination of the extreme points of $\mathcal{P}$, $\mathcal{P}$ is compact. Note that every convex combination of the extreme elements leads to a valid point in $\mathcal{P}$.

is enough to characterize $Z^n \mid s_0$ for any $s_0 \in \mathcal{S}$. We next define for each $\mathbf{p} \in \mathcal{P}$ the set of $|\mathcal{X}|^n \cdot |\mathcal{Y}|^{n-1}$ functions

$$f_{x^n, y^{n-1}}(\mathbf{p}) = Q(x^n \| y^{n-1}; \mathbf{p}), x^n \in \mathcal{X}^n, y^{n-1} \in \mathcal{Y}^{n-1} \qquad \text{(B.6)}$$

namely it is the product of elements in $\mathbf{p}$ (or complements of sums of these elements) corresponding to the vectors (x^n, y^{n-1}). Note that a set of $|\mathcal{X}|^n \cdot |\mathcal{Y}|^{n-1}$ causally conditioned distributions is required to evaluate (B.4) via

$$\int_{\mathcal{P}} f_{x^n, y^{n-1}}(\mathbf{p}) \mu^*(d\mathbf{p}), \quad x^n \in \mathcal{X}^n, y^{n-1} \in \mathcal{Y}^{n-1}. \qquad \text{(B.7)}$$

Next, for every initial state $s_0 \in \mathcal{S}$ and a causally conditioned set of distributions $\mathbf{p}$, let

$$t_Z(\mathbf{p}, s_0) \triangleq -\frac{1}{n} H(Z^n \mid s_0)_{\mathbf{p}} \qquad \text{(B.8)}$$

$$t_Y(\mathbf{p}, s_0) \triangleq \lambda \frac{1}{n} \sum_{i=1}^n I(X^i; Y_i \mid Y^{i-1}, s_0)_{\mathbf{p}}. \qquad \text{(B.9)}$$

This gives a total of $2|\mathcal{S}|$ equations. Note that

$$\begin{aligned}
&\int_{\mathcal{P}} t_Z(\mathbf{p}, s_0) \mu^*(d\mathbf{p})\\
&\quad = -\frac{1}{n} H(Z^n \mid U^n, s_0)_{p^*(u^n) Q^*(x^n \| y^{n-1}; u^n)}\\
&\quad \triangleq H^*(Z^n \mid U^n, s_0) \qquad \text{(B.10a)}\\
&\int_{\mathcal{P}} t_Y(\mathbf{p}, s_0) \mu^*(d\mathbf{p})\\
&\quad = \lambda \frac{1}{n} \sum_{i=1}^n I(X^i; Y_i \mid Y^{i-1}, U^n, s_0)_{p^*(u^n) Q^*(x^n \| y^{n-1}; u^n)}\\
&\quad = \lambda \frac{1}{n} I(X^n \to Y^n \mid U^n, s_0)_{p^*(u^n) Q^*(x^n \| y^{n-1}; u^n)}. \qquad \text{(B.10b)}
\end{aligned}$$

Finally, we note that we can obtain $\{p^*(z^n \mid s_0)\}_{Z^n \times \mathcal{S}}$ from the $|\mathcal{X}|^n \cdot |\mathcal{Y}|^{n-1}$ expressions (B.7) via (B.4) and (B.5). Therefore, we can derive $\{H^*(Z^n \mid s_0)\}_{s_0 \in \mathcal{S}}$, which in combination with the $2|\mathcal{S}|$ equations (B.10) provides a complete characterization of the optimal $C^n_{fb,SC2}(\lambda)$.

In conclusion, for any fixed measure $p^*(u^n)$, the $2|\mathcal{S}|$ expressions (B.8), (B.9) and the $|\mathcal{X}|^n \cdot |\mathcal{Y}|^{n-1}$ expressions (B.7) provide a complete characterization of $C^n_{fb,SC2}(\lambda)$. The total number of equations needed is

$$|\mathcal{X}|^n \cdot |\mathcal{Y}|^{n-1} + 2|\mathcal{S}|$$

thus by [39, Lemma 3],[5] the cardinality of $\times_{i=1}^n \mathcal{U}_i$ can be bounded by $|\mathcal{X}|^n \cdot |\mathcal{Y}|^{n-1} + 2|\mathcal{S}| + 1$. This completes the proof of the lemma. □

Finally, Lemma B.2 provides the second cardinality bound.

[5] Statement of [39, Lemma 3]: Let $\mathcal{P}_n$ be the set of all probability n-vectors $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ and let $f_j(\mathbf{p})$, $j = 1, 2, \ldots, k$, be continuous functions on $\mathcal{P}_n$. Then, to any probability measure μ on the Borel subsets of $\mathcal{P}_n$, there exist $(k+1)$ elements $\mathbf{p}_i$ of $\mathcal{P}_n$ and constants $\alpha_i \geq 0$, $i = 1, 2, \ldots, k+1$ with $\sum_{i=1}^{k+1} \alpha_i = 1$ such that

$$\int f_j(\mathbf{p}) d\mu = \sum_{i=1}^{k+1} \alpha_i f_j(\mathbf{p}_i), \qquad j = 1, 2, \ldots, k.$$

Lemma B.2: The cardinality of the auxiliary random vector U^n required to achieve the maximum of $C^n_{fb,SC2}(\lambda)$, defined in (B.1) is upper bounded by $|\mathcal{S}| \cdot |\mathcal{Z}|^n + 2|\mathcal{S}| + 1$.

Proof: Again let $\mathbf{p}(u^n)$ be the vector of probabilities defined in (B.2). Using the relationships (B.3), (B.4), and (B.5), we can derive $\{p^*(z^n \,|\, s_0)\}_{\mathcal{Z}^n \times \mathcal{S}}$ once a measure on U^n is defined. In fact, $|\mathcal{S}|(|\mathcal{Z}|^n - 1)$ expressions are enough to fully characterize $\{p^*(z^n \,|\, s_0)\}_{\mathcal{Z}^n \times \mathcal{S}}$. Together with the functions (B.8) and (B.9), a complete characterization of $C^n_{fb,SC2}(\lambda)$ is obtained. By [39, Lemma 3], this can be done with at most $|\mathcal{Z}|^n \cdot |\mathcal{S}| + 2|\mathcal{S}|$ probability vectors $\mathbf{p}(u^n)$, hence $|\times_{i=1}^n \mathcal{U}_i| \le |\mathcal{Z}|^n \cdot |\mathcal{S}| + 2|\mathcal{S}|$. □

Combining Lemmas B.1 and B.2, we conclude that the capacity region $SC2$ can be completely characterized while the cardinality of the auxiliary random vector U^n satisfies

$$|\times_{i=1}^n \mathcal{U}_i| \le \min\{|\mathcal{X}|^n \cdot |\mathcal{Y}|^{n-1}, |\mathcal{Z}|^n \cdot |\mathcal{S}|\} + 2|\mathcal{S}| + 1.$$

APPENDIX C
PROOF OUTLINES OF THE UPPER BOUNDS FOR SCENARIOS FB1-PD, RC, FB2-RC, AND FB1-RC

In this Appendix, we outline the proofs of the upper bounds for several scenarios considered in this work. We provide only the essence of the proofs, as they mostly follow the steps detailed in Appendix A-H.

A. Scenario Fb1-PD

In a parallel manner to (A.55) and (A.56), we have

$$\min_{s_0 \in \mathcal{S}} I(M_2; Z^n \,|\, s_0) = nR_2 - \max_{s_0 \in \mathcal{S}} H(M_2 \,|\, Z^n, s_0)$$
$$\min_{s_0 \in \mathcal{S}} I(M_1; Y^n \,|\, M_2, s_0) \ge nR_1 - \max_{s_0 \in \mathcal{S}} H(M_1 \,|\, Y^n, s_0).$$

Let $U_i = M_2, i = 1, 2 \dots, n$. Thus, we can write

$$\begin{aligned}
&I(M_2; Z^n \,|\, s_0) \\
&\quad = I(U^n; Z^n \,|\, s_0) \\
&I(M_1; Y^n \,|\, M_2, s_0) \\
&\quad = \sum_{i=1}^n I(M_1; Y_i \,|\, Y^{i-1}, M_2, s_0) \\
&\quad \overset{(a)}{\le} \sum_{i=1}^n \Big[H(Y_i \,|\, Y^{i-1}, U^n, s_0) \\
&\qquad\qquad - H(Y_i \,|\, Y^{i-1}, X^i, U^n, M_1, M_2, s_0)\Big] \\
&\quad \overset{(b)}{=} \sum_{i=1}^n \Big[H(Y_i \,|\, Y^{i-1}, U^n, s_0) - H(Y_i \,|\, Y^{i-1}, X^i, U^n, s_0)\Big] \\
&\quad = I(X^n \to Y^n \,|\, U^n, s_0)
\end{aligned}$$

where (a) is because conditioning reduces entropy, (b) follows from the causality of the channel [22, eq. (8)], which induces the Markov chain

$$M_1, M_2, U^n \,|\, Y^{i-1}, s_0 \leftrightarrow X^i \,|\, Y^{i-1}, s_0 \leftrightarrow Y_i \,|\, Y^{i-1}, s_0. \tag{C.1}$$

Note that the selection of U_i and $X_i = f_i(X^{i-1}, Y^{i-1}, M_1, M_2)$ induces the p.m.f

$$\begin{aligned}
&p(u_i, x_i \,|\, u^{i-1}, x^{i-1}, y^{i-1}, z^{i-1}, s^{i-1}, s_0) \\
&\qquad = p(u_i \,|\, u^{i-1}) p(x_i \,|\, u^n, x^{i-1}, y^{i-1})
\end{aligned}$$

which leads to the underlying distribution of Theorem 2.

Finally, we note that the causality relationship (C.1) is not trivial: the straightforward extension of Massey's formal causality condition [22] to the FSBC is $p(y_i, z_i \,|\, y^{i-1}, z^{i-1}, x^i, u^n, s_0) = p(y_i, z_i \,|\, y^{i-1}, z^{i-1}, x^i, s_0)$, i.e., the conditioning contains *both* channel output sequences. However, using the fact that feedback is only from Rx_1, we can write

$$p(y_i \,|\, y^{i-1}, x^i, u^n, s_0) = \frac{\sum_{\mathcal{Z}^i} p(y^i, z^i, x^i \,|\, u^n, s_0)}{\sum_{\mathcal{Z}^{i-1}} p(y^{i-1}, z^{i-1}, x^i \,|\, u^n, s_0)}$$

and use the code construction $p(u^n, x^n \,|\, y^n) = p(u^n) Q(x^n \| y^{n-1}; u^n)$ together with the FSBC causality condition above to show that (C.1) holds.

B. Scenario RC

This scenario is different from Fb1-PD in the rate bound on R_1 and the underlying distribution for X^n. We define the auxiliary RV $U_i = M_2, i = 1, 2, \dots, n$, and write

$$\begin{aligned}
&I(M_1; Y^n, Z^n \,|\, M_2, s_0) \\
&\quad = H(Y^n, Z^n \,|\, M_2, s_0) - H(Y^n, Z^n \,|\, M_1, M_2, s_0) \\
&\quad \le H(Y^n, Z^n \,|\, U^n, s_0) - H(Y^n, Z^n \,|\, X^n, U^n, s_0) \\
&\quad = I(X^n; Y^n, Z^n \,|\, U^n, s_0).
\end{aligned}$$

Clearly, as there is no feedback, $X^n = f(M_1, M_2)$ and $U^n \,|\, s_0 \leftrightarrow X^n \,|\, s_0 \leftrightarrow Y^n, Z^n \,|\, s_0$.

C. Scenario Fb2-RC

Here, we define $U_i = (M_2, Z^{i-1}), X_i = \tilde{f}_i(X^{i-1}, Z^{i-1}, M_1, M_2)$. Thus, $X^i = f_i(M_1, M_2, Z^{i-1})$ and as in Appendix A-H, $I(M_2; Z^n \,|\, s_0) = I(U^n \to Z^n \,|\, s_0)$, and

$$\begin{aligned}
&I(M_1; Y^n, Z^n \,|\, M_2, s_0) \\
&\quad = \sum_{i=1}^n I(M_1; Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, M_2, s_0) \\
&\quad \le \sum_{i=1}^n \Big[H(Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, U^i, s_0) \\
&\qquad\qquad - H(Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, X^i, U^i, s_0)\Big] \\
&\quad = I(X^n \to Y^n, Z^n \| U^n, s_0).
\end{aligned}$$

Note that the selection of auxiliary RV U_i induces the Markov chains

$$\begin{aligned}
M_1, M_2, U^i \,|\, Y^{i-1}, Z^{i-1}, s_0 &\leftrightarrow X^i \,|\, Y^{i-1}, Z^{i-1}, s_0 \\
&\leftrightarrow Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, s_0
\end{aligned}$$

which follows from causality, and

$$X^{i-1}, Y^{i-1}, S^{i-1} \leftrightarrow U^{i-1}, Z^{i-1} \leftrightarrow U_i$$
$$Y^{i-1}, S^{i-1}, s_0 \leftrightarrow U^i, Z^{i-1}, X^{i-1} \leftrightarrow X_i$$

where the last chain is also due to the code construction. The last two chains give rise to the probability distribution

$$p(u_i, x_i \,|\, u^{i-1}, z^{i-1}, x^{i-1}, y^{i-1}, s^{i-1}, s_0) = p(u_i \,|\, u^{i-1}, z^{i-1}) p(x_i \,|\, u^i, z^{i-1}, x^{i-1})$$

used in the achievability result.

D. Scenario Fb1-RC

Here we define $U_i = M_2$. We show only the rate constraint on R_1

$$\begin{aligned} I(M_1; Y^n, Z^n \,|\, M_2, s_0) &= H(Y^n, Z^n \,|\, M_2, s_0) - H(Y^n, Z^n \,|\, M_2, M_1, s_0) \\ &\le \sum_{i=1}^{n} \Big[H(Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, U^n, s_0) \\ &\quad - H(Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, X^i, U^n, M_2, M_1, s_0) \Big] \\ &= \sum_{i=1}^{n} \Big[H(Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, U^n, s_0) \\ &\quad - H(Y_i, Z_i \,|\, Y^{i-1}, Z^{i-1}, X^i, U^n, s_0) \Big] \\ &= I(X^n \to Y^n, Z^n \,|\, U^n, s_0). \end{aligned}$$

The code construction $X_i = f_i(X^{i-1}, Y^{i-1}, M_1, M_2)$ gives rise to the chain

$$S^{i-1}, Z^{i-1}, s_0 \leftrightarrow Y^{i-1}, X^{i-1}, U^n \leftrightarrow X_i.$$

Therefore, we obtain

$$p(u_i, x_i \,|\, u^{i-1}, x^{i-1}, z^{i-1}, y^{i-1}, s^{i-1}, s_0) = p(u_i \,|\, u^{i-1}) p(x_i \,|\, u^n, x^{i-1}, y^{i-1}).$$

References

[1] B. McMillan, "The basic theorems of information theory," *Ann. Math. Stat.*, vol. 24, no. 2, pp. 196–219, Jun. 1953.
[2] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[3] T. Holliday, A. Goldsmith, and P. Glynn, "Capacity of finite state channels based on Lyapunov exponents of random matrices," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3509–3532, Aug. 2006.
[4] H. Permuter, P. Cuff, B. Van Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3150–3165, Jul. 2008.
[5] H. H. Permuter, T. Weissman, and A. J. Goldsmith, "Finite-state channels with time-invariant deterministic feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 644–662, Feb. 2009.
[6] H. H. Permuter, T. Weissman, and J. Chen, "Capacity region of the finite-state multiple access channel with and without feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2455–2477, Jun. 2009.
[7] R. Dabora and A. Goldsmith, "The capacity region of the degraded finite-state broadcast channel," in *Proc. Inf. Theory Workshop*, Porto, Portugal, May 2008, pp. 11–15.
[8] R. Dabora and A. Goldsmith, "The capacity region of the degraded finite-state broadcast channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1828–1851, Apr. 2010.
[9] A. El-Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 379–381, May 1978.
[10] L. H. Ozarow, "Coding and capacity for additive white Gaussian noise multi-user channels with feedback," Ph.D. dissertation, Dept. Electr. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, 1979.
[11] S. R. Bhaskaran, "Broadcasting with feedback," in *Proc. Int. Symp. Inf. Theory*, Nice, France, Jun. 2007, pp. 406–410.
[12] A. Rosenzweig, "The capacity of Gaussian multi-user channels with state and feedback," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4349–4355, Nov. 2007.
[13] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2867–2877, Aug. 2005.
[14] Y. Steinberg and S. Shamai, "Achievable rates for the broadcast channel with states known at the transmitter," in *Proc. Int. Symp. Inf. Theory*, Adelaide, Australia, Sep. 2005, pp. 2184–2188.
[15] S. C. Draper, B. J. Frey, and F. R. Kschischang, "Interactive decoding of a broadcast message," in *Proc. Allerton Conf. Commun. Control Comput.*, Urbana, IL, Oct. 2003, pp. 170–180.
[16] R. Dabora and S. D. Servetto, "Broadcast channels with cooperating decoders," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5438–5454, Dec. 2006.
[17] R. Dabora and S. D. Servetto, "On the role of estimate-and-forward with time sharing in cooperative communication," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4409–4431, Oct. 2008.
[18] Y. Liang and V. V. Veeravalli, "Cooperative relay broadcast channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 900–928, Mar. 2007.
[19] Y. Liang and G. Kramer, "Rate regions for relay broadcast channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3517–3535, Oct. 2007.
[20] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 4–21, Jan. 2003.
[21] H. Marko, "The bidirectional communication theory—A generalization of information theory," *IEEE Trans. Commun.*, vol. 21, no. 12, pp. 1345–1351, Dec. 1973.
[22] J. L. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Its Appl.*, Waikiki, HI, Nov. 1990, pp. 303–305.
[23] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless channels," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, Mar. 1998.
[24] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Process. Mag.*, vol. 14, no. 6, pp. 49–83, Nov. 1997.
[25] , P. Golden, H. Dedieu, and K. S. Jacobsen, Eds., *Fundamentals of DSL Technology*. New York: Auerbach Publications, 2004.
[26] A. Hazmi, J. Rinne, and M. Renfors, "DVB-T signal over cable TV network and phase noise requirements," in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.*, Taoyuan, Taiwan, Mar. 2001, pp. 166–169.
[27] A. Lapidoth and I. E. Telatar, "The compound channel capacity of a class of finite-state channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 973–983, May 1998.
[28] T. M. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 2–14, Jan. 1972.
[29] T. M. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[30] P. O. Vontobel, A. Kavcic, D. Arnold, and H.-A. Loeliger, "A generalization of the Blahut-Arimoto algorithm to finite-state channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1887–1918, May 2008.
[31] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite-state ISI channels," in *Proc. IEEE Global Commun. Conf.*, San Antonio, TX, Nov. 2001, pp. 2992–2996.
[32] R. Dabora and A. Goldsmith, "On the capacity of indecomposable finite-state channels with feedback," in *Proc. Allerton Conf. Commun. Control Comput.*, Urbana, IL, Sep. 2008, pp. 1045–1052.
[33] S. Yang, A. Kavcic, and S. Tatikonda, "Feedback capacity of finite-state machine channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 799–810, Mar. 2005.
[34] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 648–657, Nov. 1980.
[35] P. Gács and J. Körner, "Common information is far less than mutual information," *Probl. Control Inf. Theory*, vol. 2, pp. 149–162, 1973.

[36] G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*. Cambridge, U.K.: Cambridge Univ. Press, 1964.

[37] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, vol. 10, no. 3, pp. 3–14, Jul.–Sep. 1974.

[38] M. Salehi, "Cardinality bounds on auxiliary variables in multiple-user theory via the method of Ahlswede and Körner," Dept. Stat., Stanford Univ., Stanford, CA, Tech. Rep. 33, 1978.

[39] R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Thoery*, vol. 21, no. 6, pp. 629–637, Nov. 1975.

[40] M. Krein and D. Milman, "On extreme points of regular convex sets," *Studia Mathematica*, vol. 9, pp. 133–138, 1940.

Ron Dabora received the B.Sc. and M.Sc. degrees from Tel-Aviv University, Tel-Aviv, Israel, in 1994 and 2000, respectively, and the Ph.D. degree from Cornell University, Ithaca, NY, in 2007, all in electrical engineering.

From 1994 to 2000, he was an Engineer at the Ministry of Defense of Israel, and from 2000 to 2003, he was with the Algorithms Group at Millimetrix Broadband Networks, Israel. From 2007 to 2009, he was a Postdoctoral Researcher at the Department of Electrical Engineering, Stanford University, Stanford, CA. Since 2009, he has been an Assistant Professor at the Department of Electrical and Computer Engineering, Ben-Gurion University, Be'er Sheva, Israel.

Andrea J. Goldsmith (S'90–M'95–SM'99–F'05) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of California Berkeley, Berkeley, in 1986, 1991, and 1994, respectively.

Currently, she is a Professor of Electrical Engineering at Stanford University, Stanford, CA, and was previously an Assistant Professor of Electrical Engineering at the California Institute of Technology (Caltech), Pasadena. She founded Quantenna Communications Inc., and has previously held industry positions at Maxim Technologies, Memorylink Corporation, and AT&T Bell Laboratories. Her research includes work on wireless information and communication theory, MIMO systems and multihop networks, cognitive radios, sensor networks, cross-layer wireless system design, wireless communications for distributed control, and communications for biomedical applications. She is the author of the book *Wireless Communications* (Cambridge, U.K.: Cambridge Univ. Press, 2005) and coauthor of the book *MIMO Wireless Communications* (Cambridge, U.K.: Cambridge Univ. Press, 2007).

Dr. Goldsmith is a Fellow of Stanford. She has received several awards for her research, including the National Academy of Engineering Gilbreth Lectureship, the IEEE Comsoc Wireless Communications Technical Committee Recognition Award, the Alfred P. Sloan Fellowship, the Stanford Terman Fellowship, the National Science Foundation CAREER Development Award, and the Office of Naval Research Young Investigator Award. In addition, she was a corecipient of the 2005 IEEE Communications Society and Information Theory Society joint paper award. She currently serves as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY and as an Editor for the *Foundations and Trends in Communications and Information Theory* and the *Foundations and Trends in Networking*. She previously served as an editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and for the IEEE WIRELESS COMMUNICATIONS MAGAZINE, as well as Guest Editor for several IEEE journal and magazine special issues. She participates actively in committees and conference organization for the IEEE Information Theory and Communications Societies and has served on the Board of Governors for both societies. She is a Distinguished Lecturer for both societies, the President of the IEEE Information Theory Society, and was the Technical Program Cochair for the 2007 IEEE International Symposium on Information Theory. She also founded the student committee of the IEEE Information Theory Society, is an inaugural recipient of Stanford's postdoc mentoring award, and serves as Stanford's faculty senate chair for the 2009/2010 academic year.