The Capacity Region of the Degraded Finite-State Broadcast Channel

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

***Abstract*—We introduce and study the discrete, finite-state broadcast channel (FSBC) with memory. For this class of channels we define *physical degradedness* and *stochastic degradedness*, and demonstrate these definitions with practical communication scenarios. We then show that a superposition codebook with memory achieves the capacity region of physically degraded FSBCs. This result is subsequently used to characterize the capacity region of stochastically degraded FSBCs. In both scenarios, we consider indecomposable as well as nonindecomposable channels.**

***Index Terms*—Broadcast channels, capacity, channels with memory, finite-state channels, indecomposable channels, network information theory, superposition codebook.**

I. Introduction

THE information-theoretic model for the broadcast channel (BC) was introduced by Cover in 1972 [1]. In this scenario, a single sender transmits three messages, one common and two private, to two receivers, over a channel defined by $\{\mathcal{X}, p(y,z|x), \mathcal{Y}\times\mathcal{Z}\}$. Here, X is the channel input from the transmitter, Y is the channel output at receiver 1 (Rx_1) and Z is the channel output at receiver 2 (Rx_2). In the years following its introduction, the study of the BC focused on memoryless scenarios, i.e., when the joint distribution of the channel outputs at time i is given by $p(y_i, z_i|x^i, y^{i-1}, z^{i-1}) = p(y_i, z_i|x_i)$. In recent years, capacity analysis of time-varying channels with memory has been the focus of considerable interest, especially in Gaussian scenarios. This has been motivated by the proliferation of mobile communications for which the channel is subject to multipath and correlated fading, both of which introduce memory into the channel.

A. Models for Time-Varying Channels

There are several approaches to modeling time-varying channels with memory. The model we use in this work is the finite-state channel (FSC) model, introduced as early as 1953 [2]. In this model, the channel memory for the current transmission is captured by the state of the channel after the transmission of the previous symbol. Letting S_{i-1} denote the state of the channel at the end of the $(i-1)$th transmission interval, the transition function of the FSC at the ith transmission interval satisfies $p(y_i, s_i|x^i, y^{i-1}, s^{i-1}) = p(y_i, s_i|x_i, s_{i-1})$. Note that both the channel output and the current state depend on the channel input as well as the previous state. The pair (X_i, S_{i-1}), makes (Y_i, S_i) independent of the entire history $(X^{i-1}, Y^{i-1}, S^{i-2})$. In this paper, we also consider a special class of FSCs called *indecomposable FSCs*. Loosely speaking, in indecomposable FSCs the effect of the initial state on the state transitions becomes negligible as time evolves. For nonindecomposable channels the initial state may affect the state transitions indefinitely, hence there may be states that will be visited only once.

Time-varying channels can also be modeled using the arbitrarily varying channel (AVC) approach. The AVC is characterized by the transition function $p(y^n|x^n, s^n) = \prod_{i=1}^{n} p(y_i|x_i, s_i)$. This model was first considered in [10]. The AVC models a memoryless channel whose law varies with time in an arbitrary and unknown manner (see also [11]). The state transitions in the AVC are independent of the channel input and output symbols. In some AVCs the conditional p.m.f $p(y|x,s)$ takes either 0 or 1. Such channels are referred to as deterministic AVCs [11]. In the FSC model, the states are random, thus $p(y|x)$, obtained by averaging over the states, is, in general, not a 0 – 1 p.m.f. Therefore FSCs are more appropriate for modeling wireless applications [11]. For a comprehensive discussion on models for time-varying channels see the survey paper [11]. In [13], the capacity of degraded arbitrarily varying broadcast channels with causal side information at the transmitter and noncausal side information at the good receiver is derived.

Related to the AVC is the model in which the channel states vary in an i.i.d. manner according to some known fixed distribution. This model is referred to as a channel with random parameters [12]. In the context of discrete, multiuser channels with random parameters, we note that the capacity of discrete, memoryless broadcast channels (DMBCs) with random parameters has been recently investigated in [14] and [30]. In [14], degraded DMBCs with random parameters with causal and noncausal side information at the transmitter were considered, and in [30] an achievable rate region for the general, memoryless BC with random parameters was derived.

B. Finite-State Channels

The capacity of FSCs without feedback was originally studied by Shannon in 1957 [3]. In his work, Shannon analyzed the capacity of FSCs subject to the assumptions that the transmitter can calculate the state sequence (i.e., the current state is a function of the previous state and the current channel input), the initial state is known at the transmitter, any state is accessible

Manuscript received April 30, 2008; revised September 08, 2009. Current version published March 17, 2010. This work was supported in part by the DARPA ITMANET program by Grant 1105741-1-TFIND and the ARO under MURI award W911NF-05-1-0246. The material in this paper was presented in part at the IEEE Information Theory Workshop (ITW), Porto, Portugal, May 2008.

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, Stanford, CA 94305 USA (e-mail: andrea@wsl.stanford.edu).

Communicated by G. Kramer, Associate Editor for Shannon Theory.

Color versions of Figures 1 and 2 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2010.2040974

from any other state in a finite number of time intervals, and the channel output depends only on the current state and the channel input. We note that this model is appropriate for, e.g., multipath channels with an ideal receiver. Without explicitly deriving the capacity, Shannon proved that for such channels capacity is independent of the initial channel state, by showing that appending the codeword with an appropriately selected sequence can drive the channel back to its initial state. Later, Blackwell, Breiman, and Thomasian derived the capacity of a class of FSCs in which all reachable output states can be reached with a finite number of time intervals from every input state, without Shannon's additional restrictions [4]. The general (i.e., nonindecomposable) FSC was studied by Gallager, whose book [5] contains an extensive treatment of the subject. [5]. The capacity of some special classes of FSCs has been studied by several other authors, see [6], [7], and references therein. In [9] a capacity analysis of finite-state (FS) Markov channels with feedback was carried out. In the models [4]–[9], the transmitter and receiver do not know the channel states, contrary to [3]. A channel model related to the FSC is the finite-*memory* channel [4]. In this model, the distribution of the current channel output depends on a finite number of previous channel inputs and outputs, as well as on the current channel input. When the alphabets are discrete, this channel can be modeled as an FSC.

C. Multiuser Channels With Memory

The multiple-access channel (MAC) with memory was studied in the framework of finite-memory channels in [31] and [32]. Here, the focus was on the effect of time synchronization on the capacity region. In [31] the channel output at time i depends on the current input pair as well as the previous $m-1$ channel input pairs. The effect of frame (codeword) synchronization on the capacity region was analyzed assuming symbol time synchronization was achieved. It was shown that with lack of frame synchronization, the capacity region is achieved by restricting the channel inputs to be stationary sequences. We note that when the symbol-time and the frame-time are synchronized, the channel becomes an indecomposable FS-MAC. It should also be noted that for broadcast channels the problem of frame synchronization does not arise, as there is a single transmitter. In a following work [32], the case in which the frames are synchronized but the symbols are not was studied for the Gaussian MAC, and a relationship between the capacity region and the waveforms was established.

The finite-state MAC was studied in [15]. This channel is characterized by the probability transition function $p(y,s|x_1,x_2,s')$, and the work in [15] also considered the effect of feedback on the achievable rate region. Lastly, the two-way finite-state channel was discussed in [34] for the special case in which each transmitter can generate a finite length signal based on its last pair of channel input and output, such that after transmission of this signal, the channel state goes back to be the state when transmission began. We also note that the finite-memory Gaussian BC was considered in [20], which derived the capacity region of continuous-alphabet BCs with intersymbol interference and colored Gaussian noise.

In this paper, we study the finite-state broadcast channel (FSBC), depicted in Fig. 1. In the FSBC model, the channel

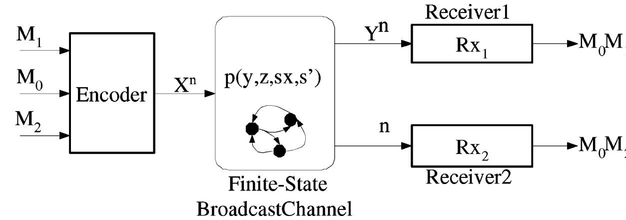

Fig. 1. The finite-state broadcast channel. The finite-state machine illustration emphasizes that the channel is governed by an "internal" state sequence. M_0 is a common message with rate R_0 intended for both receivers, M_1 and M_2 are private messages with rates R_1 and R_2, intended for Rx_1 and Rx_2, respectively.

from the transmitter to the receivers is governed by a state sequence that depends on the channel input, outputs and previous state. These symbols interact with each other according to the transition function $p(y,z,s|x,s')$, where S' and S denote the state of the channel at the end of the previous and current symbol intervals, respectively. As for the point-to-point FSC [5], we can divide the class of FSBCs into indecomposable channels and nonindecomposable channels. We will provide a precise definition of indecomposable FSBCs in Definition 2. We consider both types of channels in this paper.

The focus of this work is on *degraded* FSBCs. The capacity of discrete, memoryless, degraded BCs was characterized by Gallager [16] and Bergmans [17] in the 1970's. The key ingredient in the achievability theorem was the introduction of a superposition codebook, originally suggested by Cover in [1].

D. Main Contributions and Organization

In this paper, we introduce the finite-state broadcast channel. We first define the model and the notion of degradedness for BCs with memory. We consider the question of defining an achievable rate triplet based on the average probability of error going to zero with increasing blocklength. Here, there are two possible approaches. One approach is the compound channel approach (see [21]), that requires the average probability of error to go to zero for every initial state. In [31], the compound approach was applied to the MAC without frame synchronization. The MAC capacity region was defined as the largest region that guarantees a small probability of error for all timing offsets. The second approach is to define the average probability of error without explicitly considering the states (see [15]). In this paper we follow the compound channel approach. We expand on these ideas in the next section. We then study the capacity region of degraded FSBCs assuming that the receivers and transmitter operate without knowledge of the channel states. We prove an achievability result and an upper bound, with the average probability of error defined as in the compound channel approach. These results characterize the capacity region of nonindecomposable degraded FSBCs and of indecomposable degraded FSBCs, both for physically degraded and stochastically degraded channels.

The rest of this paper is organized as follows: in Section II we introduce the channel model and provide a formal definition of the scenario. In Section III we study the capacity of general degraded FSBCs as well as the special class of indecomposable FSBCs. Finally, in Section IV we give some concluding remarks. The proof of the converse and the achievability theorem are provided in Appendices I and II, respectively.

II. Channel Model and Definitions

A. Notations

In the following, we denote random variables with upper case letters, e.g., X, Y, and their realizations with lower case letters, x, y. A random variable (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 $\mathcal{X}$, 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 symbol x. We use $p_{X|Y}(x|y)$ to denote the conditional p.m.f. of X given 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 $i \le j$ 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 $\mathbf{x} \equiv x^n$. When $i > j$, $x_i^j = \emptyset$, where $\emptyset$ denotes the empty set. A vector of random variables 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 $i \le j$. We use $H(\cdot)$ to denote the entropy of a discrete random variable and $I(\cdot;\cdot)$ to denote the mutual information between two random variables, as defined in [22, Ch. 2]. $I(\cdot;\cdot)_q$ denotes the mutual information evaluated with a p.m.f. q on the channel inputs. Finally, *co* $\mathcal{R}$ denotes the convex hull of the set $\mathcal{R}$, $\mathbb{N}$ denotes the set of natural numbers, and we use $X \perp\!\!\!\perp Y$ to denote that X is statistically independent of Y.

B. Definition of the General FSBC Scenario

Definition 1: The discrete, finite-state broadcast channel 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. The p.m.f. of the FSBC satisfies

$$p(y_i, z_i, s_i | y^{i-1}, z^{i-1}, s^{i-1}, x^i, s_0) = p(y_i, z_i, s_i | x_i, s_{i-1}).$$

An example of an FSBC is the finite-duration intersymbol interference broadcast channel depicted in Fig. 2.

When the transmitter is oblivious of the channel states, the p.m.f. of a block of n transmissions can be written as

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

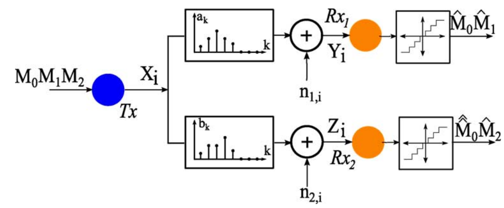

Fig. 2. A schematic description of the FSBC with ISI.

where s_0 is the initial state of the channel. Here, (a) holds since the transmitter is oblivious of the channel states, therefore X_i is independent of S_0^{i-1} when X^{i-1}, Y^{i-1}, and Z^{i-1} are given, and (b) captures the fact that given (X_i, S_{i-1}), the channel outputs and the channel state at time i are independent of the past. Equation (1) is the most general formulation as it does not rule out feedback. When the transmitter operates without feedback, then (see also [23])

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

and we arrive at

$$p(y^n, z^n, s^n | x^n, s_0) = \prod_{i=1}^{n} p(y_i, z_i, s_i | x_i, s_{i-1}). \quad (2)$$

In this paper, we focus on the FSBC *without* feedback.

Definition 2: The FSBC is called *indecomposable* if for every $\epsilon > 0$ there exists $N_0(\epsilon) \in \mathbb{N}$ such that for all $n > N_0(\epsilon)$

$$|p(s_n|\mathbf{x}, s_0) - p(s_n|\mathbf{x}, s_0')| \le \epsilon \quad (3)$$

for all s_n, $\mathbf{x}$, and initial states s_0 and s_0'.

Note that (3) is identical to the definition of indecomposable point-to-point (PtP) FSCs [5, eq. 4.6.26]. This is because indecomposability characterizes the interaction between the states and the channel inputs, averaging out the channel output(s), [5, Sec. 4.6]. As in both PtP-FSCs and FSBCs there is a single channel input, the criteria for the channel to be indecomposable is the same, though the models are different.

Definition 3: An (R_0, R_1, R_2, n) *deterministic code* for the FSBC consists of three message sets, $\mathcal{M}_k = \{1, 2, \ldots, 2^{nR_k}\}$, $k = 0, 1, 2$, and three mappings $(f, g_{\mathbf{y}}, g_{\mathbf{z}})$ such that

$$f: \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2 \mapsto \mathcal{X}^n$$

is the encoder, and

$$\begin{aligned} g_{\mathbf{y}} &: \mathcal{Y}^n \mapsto \mathcal{M}_0 \times \mathcal{M}_1 \\ g_{\mathbf{z}} &: \mathcal{Z}^n \mapsto \mathcal{M}_0 \times \mathcal{M}_2 \end{aligned}$$

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. We assume no knowledge of the states at the transmitter and receivers.

Definition 4: The maximum average probability of error of a code for the FSBC is defined as $\max_{s_0 \in \mathcal{S}} P_e^{(n)}(s_0)$, where

$$P_e^{(n)}(s_0) = \Pr\big(g_{\mathbf{y}}(Y^n) \neq (M_0, M_1) \quad \text{or} \quad g_{\mathbf{z}}(Z^n) \neq (M_0, M_2) | S_0 = s_0\big)$$

is the average probability of error when the initial state is s_0, 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 from their message sets.

Definition 5: 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 for all $n > n(\epsilon, \delta)$ it is possible to construct an $(R_0 - \delta, R_1 - \delta, R_2 - \delta, n)$ code with $\max_{s_0 \in \mathcal{S}} P_e^{(n)}(s_0) \leq \epsilon$.

Definition 6: The *capacity region* of the FSBC is the convex hull of all achievable rate triplets.

As mentioned in Section I-D, there are two possible approaches to the definition of the average probability of error. Each approach leads to a different characterization of the upper bound on the capacity region of the FSBC. In Definition 4, the probability of error is evaluated for every initial state s_0, and the maximum error probability is taken. This was done, for example, in [21]. Alternatively, one can define the average probability of error without explicitly considering the initial state, as was done, for example, in [15]

$$P_e^{(n)} = \Pr\big(g_{\mathbf{y}}(Y^n) \neq (M_0, M_1) \ \text{ or } g_{\mathbf{z}}(Z^n) \neq (M_0, M_2)\big). \tag{4}$$

Each approach leads to a different capacity region. The first approach is used in Definition 5: this is the *compound channel* approach (see also [11] and [21]), where it is required that the probability of error can be made arbitrarily small for every initial state. This definition is particularly useful for nonindecomposable channels, since for such channels the effect of s_0 may never fade away, hence all initial states must be accounted for. We will show that this approach allows us to obtain the capacity region of the general, degraded finite-state broadcast channel (the channel is general in the sense that no assumptions are made on the structure of the state transitions). The second approach does not explicitly involve the initial state, therefore it is suitable for scenarios where the initial state does not affect the behavior of the channel as the blocklength grows to infinity. This is the case with indecomposable FSBCs. When the channel is nonindecomposable, then this definition imposes some kind of an "averaging" and thus it leads to an optimistic error bound. However, when the channel is indecomposable, then both Definition 4 and the alternative definition (4) lead to the same capacity region. In this paper we take the compound approach.

We now proceed to the definition of the *degraded* FSBC. When discussing degradedness here, we use the convention that Z^n, the channel output at Rx_2, is a degraded version of Y^n, the channel output at Rx_1.

C. Definition of Physical and Stochastic Degradedness for the Broadcast Channel With Memory

For discrete, memoryless broadcast channels (DMBCs), physical degradedness is characterized via a "single-letter" relationship

$$p(z|y,x) = p(z|y). \tag{5}$$

This is sufficient for the memoryless case as the scenario is completely characterized by the joint distribution $p(y,z|x)$. Conceptually, the essence of physical degradedness is that one receiver is "weaker" than the other receiver in the sense that the signal received at the "weak" receiver is a noisy version of the signal received at the "strong" receiver. Intuitively, the concept of degradedness can be captured by the following two statements, which should hold for every initial state s_0 and symbol time i (these statements will be made mathematically precise shortly).

- S1. The channel output at the weak receiver, Z^i, does not "contain information" beyond what is "already present" in the channel output of the strong receiver, Y^i.
- S2. When the history Z^{i-1} is given, Y^i makes Z_i independent of X^i.

The single letter characterization (5) suffices to capture S1 and S2 for memoryless BCs. However, for BCs with memory a single-letter characterization is generally not possible. Thus, we introduce the following definition.

Definition 7: The FSBC is called *physically degraded* if for every $s_0 \in \mathcal{S}$ and symbol time i its p.m.f. satisfies

$$p(y_i|x^i, y^{i-1}, z^{i-1}, s_0) = p(y_i|x^i, y^{i-1}, s_0) \tag{6a}$$

$$p(z_i|x^i, y^i, z^{i-1}, s_0) = p(z_i|y^i, z^{i-1}, s_0). \tag{6b}$$

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 when Y^{i-1} is given (statement S1). 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 described in statement S2. For memoryless channel (6b) reduces to $p(z_i|y_i, x_i) = p(z_i|y_i)$. Thus, for memoryless channels Definition 7 reduces to the standard definition of (5). 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. Let $Q(y^n \| z^{n-1}, x^n, s_0) = \prod_{i=1}^n p(y_i|y^{i-1}, z^{i-1}, x^i, s_0)$ and $Q(z^n \| y^n, x^n, s_0) = \prod_{i=1}^n p(z_i|z^{i-1}, y^i, x^i, s_0)$; 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 seen as the equivalent of $p(y,z|x) = p(y|x)p(z|y)$ which characterizes the joint distribution of the outputs for degraded DMBCs.

Proposition 1: Definition 7 implies

$$p(z^n|y^n, x^n, s_0) = p(z^n|y^n, s_0), \qquad \forall s_0 \in \mathcal{S}. \tag{7}$$

Proof: Using (6a) and (6b) we obtain (when $p(y^n, x^n|s_0) > 0$)

$$\begin{aligned}
&p(z^n|y^n, x^n, s_0)\\
&= \frac{p(z^n, y^n, x^n|s_0)}{p(y^n, x^n|s_0)}\\
&= \frac{\prod_{i=1}^n p(z_i, y_i, x_i|z^{i-1}, y^{i-1}, x^{i-1}, s_0)}{\prod_{i=1}^n p(y_i, x_i|y^{i-1}, x^{i-1}, s_0)}\\
&= \frac{\prod_{i=1}^n p(x_i|z^{i-1}, y^{i-1}, x^{i-1})}{\prod_{i=1}^n p(x_i|y^{i-1}, x^{i-1}) \prod_{i=1}^n p(y_i|y^{i-1}, x^i, s_0)}\\
&\qquad\qquad \times \prod_{i=1}^n p(z_i, y_i|z^{i-1}, y^{i-1}, x^i, s_0)\\
&\overset{(a)}{=} \frac{\prod_{i=1}^n p(x_i|x^{i-1}) \prod_{i=1}^n p(z_i, y_i|z^{i-1}, y^{i-1}, x^i, s_0)}{\prod_{i=1}^n p(x_i|x^{i-1}) \prod_{i=1}^n p(y_i|y^{i-1}, x^i, s_0)}\\
&\overset{(b)}{=} \frac{\prod_{i=1}^n p(y_i|y^{i-1}, x^i, s_0) \prod_{i=1}^n p(z_i|z^{i-1}, y^i, x^i, s_0)}{\prod_{i=1}^n p(y_i|y^{i-1}, x^i, s_0)}\\
&\overset{(c)}{=} \prod_{i=1}^n p(z_i|z^{i-1}, y^i, s_0) \qquad (8)
\end{aligned}$$

where (a) holds since there is no feedback, (b) follows from (6a) and (c) follows from (6b). We therefore conclude that when (6) holds then we obtain (7). ■

From Proposition 1 it follows that Definition 7 implies the chain $p(y^n, z^n|x^n, s_0) = p(y^n|x^n, s_0)p(z^n|y^n, s_0)$ holds for every initial state $s_0 \in \mathcal{S}$. Note that (8) shows how to obtain $p(z^n|y^n, x^n, s_0)$ in a causal manner. Also note that degradedness does not eliminate the memory.

A special case of the physically degraded FSBC occurs when in (6b) it holds that $p(z_i|x^i, y^i, z^{i-1}, s_0) = p(z_i|y_i)$. When this is the case then

$$p(z^n|y^n, x^n, s_0) = p(z^n|y^n, s_0) = p(z^n|y^n) = \prod_{i=1}^n p(z_i|y_i). \quad (9)$$

(9) is similar to the definition of degradedness for the broadcast channel with random parameters used in [14]. We next define stochastic degradedness.

Definition 8: The FSBC is called *stochastically degraded* if there exists a p.m.f. $\tilde{p}(z|y)$ such that for every blocklength n and initial state $s_0 \in \mathcal{S}$

$$\begin{aligned}
p(z^n|x^n, s_0) &= \sum_{\mathcal{Y}^n} p(y^n, z^n|x^n, s_0)\\
&= \sum_{\mathcal{Y}^n} p(y^n|x^n, s_0) \prod_{i=1}^n \tilde{p}(z_i|y_i). \qquad (10)
\end{aligned}$$

We now make the following comments.

Comment 1: There are situations in which one can arrive at (10) from the per-symbol p.m.f. For example, assume that there exists a p.m.f. $\tilde{p}(z|y)$ such that we can write the p.m.f. $p(z, s|x, s')$ as

$$\begin{aligned}
p(z, s|x, s') &= \sum_{\mathcal{Y}} p(y, s|x, s')p(z|y, s, x, s')\\
&= \sum_{\mathcal{Y}} p(y, s|x, s')\tilde{p}(z|y). \qquad (11)
\end{aligned}$$

In fact, when (11) holds then also (6b) holds (with $\tilde{p}(z_i|y_i)$ on the right-hand side). Using (11) we can write the p.m.f. $p(z^n|x^n, s_0)$ as

$$\begin{aligned}
p(z^n|x^n, s_0) &= \sum_{\mathcal{S}^n} p(z^n, s^n|x^n, s_0)\\
&\overset{(a)}{=} \sum_{\mathcal{S}^n} \prod_{i=1}^n p(z_i, s_i|s_{i-1}, x_i)\\
&\overset{(b)}{=} \sum_{\mathcal{S}^n} \prod_{i=1}^n \sum_{y_i \in \mathcal{Y}} p(y_i, s_i|s_{i-1}, x_i)\tilde{p}(z_i|y_i)\\
&= \sum_{\mathcal{S}^n} \sum_{\mathcal{Y}^n} \prod_{i=1}^n p(y_i, s_i|s_{i-1}, x_i)\tilde{p}(z_i|y_i)\\
&= \sum_{\mathcal{S}^n} \sum_{\mathcal{Y}^n} p(y^n, s^n|x^n, s_0) \prod_{i=1}^n \tilde{p}(z_i|y_i)\\
&= \sum_{\mathcal{Y}^n} p(y^n|x^n, s_0) \prod_{i=1}^n \tilde{p}(z_i|y_i) \qquad (12)
\end{aligned}$$

where (a) follows from (2) after summing over $\mathcal{Y}^n$, and (b) follows from (11). Since (12) is the same as (10), then when (11) holds the channel is stochastically degraded.

Comment 2 (An Alternative Definition of Physical and Stochastic Degradedness): We note that the definitions of physical and stochastic degradedness do not explicitly involve the state sequence. The main reason is that in our work the transmitter and receivers are oblivious of the state sequence, thus the rate expressions do not involve the state evolution during the transmission, only the initial state. In order to apply degradedness under these assumptions we need only to verify (7). However, (7) does not characterize the *actual causal* operation of channel, and therefore we cannot conclude from (7) the relationship between the outputs at time i and the previous outputs and the ith input. In order to characterize FSBCs through the causal relationship between their inputs and outputs, which represents how they *physically operate*, we decided to define degradedness via (6). We also note that definitions 7 and 8 also apply to channels with memory beyond the framework of FSBCs (subject to properly handling the initial state) and that for memoryless channels these definitions specialize to the standard definition (5).

There are ways to define degradedness which explicitly involve the state sequence. One way is to say that the FSBC is called physically degraded if its p.m.f. satisfies

$$p(z_i, s_i|y_i^n, x_i^n, s_{i-1}) = p(z_i, s_i|y_i^n, s_{i-1}). \qquad (13)$$

Thus, (Y_i^n, S_{i-1}) makes (Z_i, S_i) independent of X^n. A possible interpretation of (13) is that conditioning on S_{i-1} makes (Z_i, S_i) independent of X^{i-1} due to the finite-state property of the channel, and further conditioning on Y_i^n makes (Z_i, S_i) independent of X_i^n due to degradedness. We see that contrary to Definition 7 which defines degradedness in a causal manner, here degradedness is defined using the noncausal signal Y_i^n. This is necessary in order to obtain (7) which is required for exploiting degradedness in scenarios without feedback.

Using (13) we obtain

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

where (a) holds since (X_i, S_{i-1}) makes (Z_i, S_i) independent of the past and (b) follows from (13). Hence, (7) is verified.

Comment 3: 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, i.e., $\tilde{p}(z^n|y^n, x^n, s_0) = \prod_{i=1}^n \tilde{p}(z_i|y_i)$. It is indeed possible to provide an n-letter definition for stochastic degradedness, i.e., define a channel to be stochastically degraded if for *every* blocklength n we can find a $\tilde{p}(z^n|y^n, s_0)$

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

However, as in the previous comment, this definition does not give insight into the causal operation of the degrading channel. It should be noted that the same type of degrading channel was considered for the degraded BC with random parameters studied in [14]. We note that for stochastic degradedness we also presented a condition on the per-symbol p.m.f. that guarantees degradedness [see (11)]. In the next subsection, we provide an actual communication scenario example that corresponds to Definition 8.

D. The Stochastically Degraded FSBC: An Example

Consider the scenario depicted in Fig. 3 of a base station that transmits to two mobile units, located approximately on the same line-of-sight (LoS) from the base station (BS). The LoS is indicated by the dashed line in Fig. 3. Let the BS transmit a BPSK signal and let the received signals be subject to additive white Gaussian noise. Assume that each receiver first quantizes its channel outputs into binary symbols with a threshold at zero. The quantized signal is used for decoding the messages. The resulting scenario is the binary symmetric broadcast channel (BSBC, [1]). Denote the situation in which there is no traffic on the road between the BS and the mobiles as state A. Let the channel BS–Rx_1 have a crossover probability $\epsilon_1(A) = 0.1$ and the channel BS–Rx_2 have a crossover probability $\epsilon_2(A) = 0.15$. This can be represented as a stochastically degraded BC with a degrading channel $p(z|y, s = A)$ whose crossover probability is

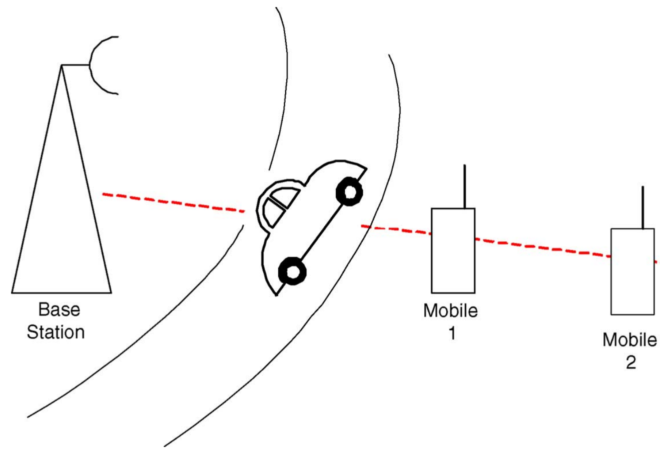

Fig. 3. A stochastically degraded FSBC example: the mobile units are located approximately on the same line-of-sight from the base-station (indicated by the dashed line). Passing cars affect the channels to both mobile units simultaneously.

$$p\big(z \neq y|y, S = A\big) = \frac{\epsilon_2 - \epsilon_1}{1 - 2\epsilon_1} = 0.0625.$$

Assume that occasionally a car passes on the road between the BS and the mobiles. This causes attenuation in both channels simultaneously. Call this state B and let $\epsilon_1(B) = 0.18$ and $\epsilon_2(B) = 0.22$. Again we have $p\big(z \neq y|y, S = B\big) = 0.0625$[1]. Hence, the degrading channel is the same for both states, irrespective of the state sequence. This satisfies (10).

In this example, the state sequence represents the traffic pattern, which depends on the lengths and speeds of the cars as well as on the distances between cars. Hence, the states are not independent. This channel is actually a Gilbert–Elliott channel [28], where a passing car corresponds to a "bad" state and without a car the channel is in a "good" state.

III. Capacity Theorems for Degraded FSBCs

In this section, we present the capacity regions for the physically degraded and the stochastically degraded FSBCs.

[1]The scenario parameters assumed in this example are: two-ray propagation model, Base station Tx power = 30 dBm, Base station antenna gain = 10 dBi, Rx antenna gain = 0 dBi, Rx noise floor = -90 dBm, Base station antenna height = 10 m, Rx antenna height = 1.5 m, BS-Rx_1 distance = 7.2 Km and BS-Rx_2 distance = 8 Km. We also assume a passing car increases the path attenuation by 3 dB.

A. Physically Degraded FSBCs

Let $\mathcal{Q}_n$ be the set of all joint distributions $p(u^n, x^n)$ on $(\times_{i=1}^n \mathcal{U}_i, \mathcal{X}^n)$ where the cardinality of the auxiliary RV U^n is bounded by

$$\| \times_{i=1}^n \mathcal{U}_i \| \leq \min\left\{ \|\mathcal{X}\|^n, \|\mathcal{Y}\|^n \cdot \|\mathcal{S}\|, \|\mathcal{Z}\|^n \cdot \|\mathcal{S}\| \right\} + 2\|\mathcal{S}\| + 1. \quad (14)$$

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

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

The main result is stated in the following theorem.

Theorem 1: The capacity region of the physically degraded FSBC is given by

$$\mathcal{C}_{pd} = \lim_{n \to \infty} \mathcal{R}_n \quad (15)$$

and the limit exists[2].

Proof Outline: The converse proof is based on manipulating Fano's inequality. The details are given in Appendix I. The detailed achievability proof is deferred to Appendix II; we provide here a basic sketch of the proof to highlight its key elements:

1) Begin by considering the situation where there is no common message. Fix n and a joint probability distribution $p(u^n, x^n)$. Generate a superposition codebook with u^n being the cloud centers, indexed by the messages in $\mathcal{M}_2$, and x^n being the cloud elements, indexed by the message pairs in $\mathcal{M}_1 \times \mathcal{M}_2$.
2) Using a decoder at Rx_2 according to
$$\hat{m}_2 = \arg\max_{\tilde{m}_2} \sum_{s_0 \in \mathcal{S}} \frac{1}{\|\mathcal{S}\|} p\big(z^n \big| u^n(\tilde{m}_2), s_0\big)$$
we show that a positive error exponent[3] for decoding M_2 can be obtained as long as $R_2 \leq \min_{s_0'' \in \mathcal{S}} \frac{1}{n} I(U^n; Z^n | s_0'')_p - \frac{\log_2 \|\mathcal{S}\|}{n} \triangleq R_{2,n}(p)$. As degradedness implies that for every s_0, $\frac{1}{n} I(U^n; Z^n | s_0) \leq \frac{1}{n} I(U^n; Y^n | s_0)$, then a positive error exponent can be achieved also for decoding M_2 at Rx_1.
3) Using maximum-likelihood decoding at Rx_1 according to
$$\hat{m}_1 = \arg\max_{\tilde{m}_1} \sum_{s_0 \in \mathcal{S}} \frac{1}{\|\mathcal{S}\|} p\big(y^n \big| x^n(\tilde{m}_1, m_2), s_0\big)$$
we show that, given M_2 was correctly decoded at Rx_1, a positive error exponent for decoding M_1 at Rx_1 can be achieved as long as $R_1 \leq \min_{s_0' \in \mathcal{S}} \frac{1}{n} I(X^n; Y^n | U^n, s_0')_p - \frac{\log_2 \|\mathcal{S}\|}{n} \triangleq R_{1,n}(p)$.
4) Next, for a fixed n and some integer K, we let $b = Kn$. As in [15], we construct a distribution for (U^b, X^b) by taking the product of the basic distribution for a block of n symbols b times:
$$p(u^b, x^b) = \prod_{k=1}^{K} p_{U^n, X^n}\left(u_{(k-1)n+1}^{kn}, x_{(k-1)n+1}^{kn} \right).$$
For blocklength b we generate a superposition codebook according to $p(u^b, x^b)$. Now, we show that taking K large enough results in a maximum average probability of error that is arbitrarily small, hence $(R_{1,n}(p), R_{2,n}(p))$ is achievable.
5) Finally, for $\lambda > 0$ define $C^n(\lambda)$ and $F_n(\lambda)$:
$$C^n(\lambda) = \max_{p(u^n, x^n)} \left\{ \min_{s_0'' \in \mathcal{S}} \frac{1}{n} I(U^n; Z^n | s_0'')_p + \lambda \min_{s_0' \in \mathcal{S}} \frac{1}{n} I(X^n; Y^n | U^n, s_0')_p \right\} \quad (16)$$
$$F_n(\lambda) = C^n(\lambda) - (1 + \lambda) \frac{\log_2 \|\mathcal{S}\|}{n}. \quad (17)$$
6) We derive a bound on the cardinality of the auxiliary RV U^n. This bound is given in (14). The cardinality bound asserts that the optimization problem for finding the maximum rate pairs has a finite-dimension domain for any given n. The maximization in (16) is carried out subject to the cardinality bound on $\times_{i=1}^n \mathcal{U}_i$.
7) We show that $F_n(\lambda)$ is sup-additive[4] and bounded, hence
$$C^{\infty}(\lambda) \triangleq \lim_{n \to \infty} C^n(\lambda) = \lim_{n \to \infty} F_n(\lambda) = \sup_n F_n(\lambda). \quad (18)$$
Therefore, the boundary of the largest achievable region can be written as
$$R_2(R_1) = \inf_{0 \leq \lambda \leq 1} \left(C^{\infty}(\lambda) - \lambda R_1 \right). \quad (19)$$
8) The common message is incorporated by splitting the rate to the weak receiver Rx_2 into common and private parts, as in [22, Theorem 14.6.4]. Recall that in step 2 we showed that for the physically degraded FSBC, if the weak receiver Rx_2 can decode a message with an arbitrarily small probability of error, the strong receiver Rx_1 can achieve an arbitrarily small probability of error when decoding the same message as well. Therefore, successful decoding of the common message at Rx_2 grantees its successful decoding Rx_1. Therefore, only the rate R_2 is reduced to accommodate the common information.

[2]The limit of sets is defined as follows [24]: Let $\{\mathcal{A}_i\}_{i=1}^{\infty}$ be a sequence of sets. We now define the following limits: $\liminf_{n \to \infty} \mathcal{A}_n = \{\mathbf{a} : \mathbf{a} = \lim_{n \to \infty} \mathbf{a}_n, \mathbf{a}_n \in \mathcal{A}_n\}$ ($\mathbf{a}$ belongs to all but finitely many $\mathcal{A}_n$'s), and $\limsup_{n \to \infty} \mathcal{A}_n = \{\mathbf{a} : \mathbf{a} = \lim_{k \to \infty} \mathbf{a}_k, \mathbf{a}_k \in \mathcal{A}_{n_k}$, for n_k an increasing subsequence of integers$\}$ ($\mathbf{a}$ belongs to infinitively many $\mathcal{A}_n$'s). A sequence of sets $\{\mathcal{A}_i\}_{i=1}^{\infty}$ is said to converge to a limit $\mathcal{A}$ if $\limsup_{n \to \infty} \mathcal{A}_n = \liminf_{n \to \infty} \mathcal{A}_n = \mathcal{A} \triangleq \lim_{n \to \infty} \mathcal{A}_n$. In this paper, we examine the limit of a sequence of convex sets in the first quadrant whose boundary includes the positive axes. We show that the boundary of the sets converges to a unique boundary as the blocklength n grows to infinity. This implies that the sets converge to a limiting set as well.

[3]We develop an exponential upper bound on the average probability of error. At Rx_2 the bound is of the form $P_{e2}^{(n)} \leq 2^{-nF_2(R_2, p(\mathbf{u}, \mathbf{x}), n)}$. We refer to $F_2(R_2, p(\mathbf{u}, \mathbf{x}), n)$ as the *error exponent* for Rx_2.

[4]Here we use [18, Pg. 165]: a sequence $\{a_N\}_{N \in \mathbb{N}}$ is called sup-additive if for all $n \geq 1$ and all $N > n$, $Na_N \geq na_n + (N - n)a_{N-n}$.

Comment 4: Note that we use $\times_{i=1}^{n}\mathcal{U}_i$ and not $\mathcal{U}^n$ since we cannot guarantee that all the RVs $\{U_i\}_{i=1}^{n}$ have the same cardinality. Thus, the sample space over which the auxiliary RV U^n is defined is not the n-fold product of a single set $\mathcal{U}$, but the product $\mathcal{U}_1 \times \mathcal{U}_2 \times \cdots \times \mathcal{U}_n$ (clearly one may set all $\mathcal{U}_i$ to the set with the maximal cardinality but that may exceed the cardinality bound on the vector).

B. Stochastically Degraded FSBCs

Since the capacity region of the broadcast channel depends only on the conditional marginals $p(y^n|x^n, s_0)$ and $p(z^n|x^n, s_0)$ (see [22, Ch. 14.6]) then the capacity region of the stochastically degraded FSBC is the same as that of the corresponding physically degraded FSBC:

Corollary 1: For the stochastically degraded FSBC of Definition 8, the capacity region is given by Theorem 1.

C. Indecomposable FSBCs

When the FSBC is indecomposable, the effect of the initial state on the state transitions becomes negligible as n increases. Therefore, the maximum over all $s_0 \in \mathcal{S}$ of each of the mutual information expressions used in the definition of the region $\mathcal{R}_n$, asymptotically equals the minimum over all $s_0 \in \mathcal{S}$. Let us define

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

The capacity regions for indecomposable FSBCs are now characterized in the following theorem.

Theorem 2:

1) For indecomposable physically degraded FSBCs the capacity region is characterized by

$$\mathcal{C}_{pd}^{\text{indecomp.}} \stackrel{(a)}{=} \lim_{n\to\infty} \mathcal{R}_n \stackrel{(b)}{=} \lim_{n\to\infty} \tilde{\mathcal{R}}_n \qquad (20)$$

and the limits exist.

2) For indecomposable stochastically degraded FSBCs, the capacity region is characterized by (20).

Proof: Equality (a) follows from Theorem 1. Equality (b) is proved in Appendix IV.

D. Discussion

We note the following facts about the capacity region of FSBCs and the capacity-achieving codes.

1) Let $\mathcal{R}_n(s_0)$ be defined as

$$\mathcal{R}_n(s_0) = co \bigcup_{q_n \in \mathcal{Q}_n} \Big\{ (R_0, R_1, R_2) : R_0 \geq 0, R_1 \geq 0, R_2 \geq 0, \\ R_1 \leq \frac{1}{n} I(X^n; Y^n | U^n, s_0)_{q_n}, \\ R_0 + R_2 \leq \frac{1}{n} I(U^n; Z^n | s_0)_{q_n} \Big\}. \qquad (21)$$

Then, from the derivation of the achievable region for a fixed n it follows that

$$\mathcal{R}_n \subseteq \bigcap_{s_0 \in \mathcal{S}} \mathcal{R}_n(s_0). \qquad (22)$$

The region $\mathcal{R}_n(s_0)$ can be viewed as the rate region with blocklength n obtained when the receivers are given the initial channel state s_0. The inclusion follows because achieving a rate point inside $\bigcap_{s_0 \in \mathcal{S}} \mathcal{R}_n(s_0)$ may require a different input distribution $p(u^n, x^n)$ for each initial channel state. Thus, not all rate points in the intersection are necessarily achievable without knowledge of the initial state at the transmitter. This inclusion illustrated in Fig. 4. It then follows that $\mathcal{C}_{pd} \subseteq \lim_{n\to\infty} \bigcap_{s_0 \in \mathcal{S}} \mathcal{R}_n(s_0)$, if the limit exists.

2) The codebook structure that achieves capacity is a superposition codebook. This introduces a structural constraint when optimizing the codebook for achieving the maximum rate triplets.

3) For indecomposable FSBCs (such as discrete finite-memory BCs) (20) implies that it is not necessary, in the limit of large blocklength, to use knowledge of the initial channel state in the design of the encoder and decoders. We can therefore fix an initial state $s_0 \in \mathcal{S}$ and the capacity $\mathcal{C}_{pd}^{indecomp.}$ can be approached arbitrarily close by finding the input distribution $p(u^n, x^n)$ that maximizes $\mathcal{R}_n(s_0)$. This happens because for indecomposable channels the effect of the initial channel state becomes negligible over time. This also implies that for indecomposable FSBCs the inclusion relationship (22) holds with equality, and that both the compound approach (Definition 4) and the average $P_e^{(n)}$ approach ((4)) lead to the same capacity region.

4) The auxiliary RV U^n introduces difficulties mainly in places where we need to rely on its cardinality. This is because we cannot translate the bound on the cardinality of U^n into a bound on the cardinality of a subset of U^n. In particular, we cannot use the cardinality of U^n when deriving the capacity region for the indecomposable FSBC in Appendix IV. Moreover, letting $n = m_1 + m_2$, from (2) we have that

$$p(z^{m_1}, y^{m_1}, s^{m_1} | x^n, s_0) = p(z^{m_1}, y^{m_1}, s^{m_1} | x^{m_1}, s_0).$$

But because $p(x^{m_1}|u^n) \neq p(x^{m_1}|u^{m_1})$ it follows that

$$p(z^{m_1}, y^{m_1}, s^{m_1} | u^n, s_0) \neq p(z^{m_1}, y^{m_1}, s^{m_1} | u^{m_1}, s_0).$$

This is a major difference between FSBCs and point-to-point and MAC channels. Consider, for example,

$$\max_{p(u^n,x^n)} \Big\{ \max_{s_0'' \in \mathcal{S}} \frac{1}{n} I(U^n; Z^n | s_0'')_p \\ + \lambda \max_{s_0' \in \mathcal{S}} \frac{1}{n} I(X^n; Y^n | U^n, s_0')_p \Big\} \qquad (23)$$

which serves as an upper bound on the achievable region. While in point-to-point channels the corresponding upper bound converges for all channels (see [5, Theorem

4.6.1]), for FSBCs (23) can be shown to converge only in the indecomposable case. Therefore, using superposition coding, the dependence between U^n and (Y^n, Z^n) is fundamentally different from the dependence between X^n and (Y^n, Z^n). This is also in contrast to DMBCs. For this reason in the following proofs, we pay special attention to derivations that involve manipulation of the distribution chain.

5) As mentioned in Comment 2, Definition 7 which defines physical degradedness without using the channel states is a general definition and applies to classes of channels beyond FSBCs. It should be noted that even if the state sequence is available at the transmitter, Definition 7 still captures the essence of degradedness. If the state sequence is available at both decoders then one may use a definition for degradedness that explicitly involves the states (represent the memory through the states). For example, one may define physical degradedness using the per-symbol p.m.f. of the FSBC:

$$p(z|y,x,s,s') = p(z|y,s,s'). \tag{24}$$

Then we obtain

$$\begin{aligned} p(z^n, s^n|y^n, s^n, x^n, s_0) &= \frac{p(z^n, y^n, s^n, x^n|s_0)}{p(y^n, s^n, x^n|s_0)} \\ &= \frac{\prod_{i=1}^n p(z_i, y_i, s_i, x_i|z^{i-1}, y^{i-1}, s^{i-1}, x^{i-1}, s_0)}{\prod_{i=1}^n p(y_i, s_i, x_i|y^{i-1}, s^{i-1}, x^{i-1}, s_0)} \\ &= \frac{\prod_{i=1}^n p(z_i, y_i, s_i|s_{i-1}, x_i)}{\prod_{i=1}^n p(y_i, s_i|s_{i-1}, x_i)} \\ &= \prod_{i=1}^n p(z_i|y_i, x_i, s_i, s_{i-1}) \\ &= \prod_{i=1}^n p(z_i|y_i, s_i, s_{i-1}). \end{aligned}$$

Thus

$$p(z^n, s^n|y^n, s^n, x^n, s_0) = p(z^n, s^n|y^n, s^n, s_0) \tag{25}$$

which implies the Markov chain $U^n|s_0 \leftrightarrow X^n|s_0 \leftrightarrow Y^n, S^n|s_0 \leftrightarrow Z^n, S^n|s_0$ for every $s_0 \in \mathcal{S}$. Then, using the data processing inequality we obtain $\min_{s_0\in\mathcal{S}} I(U^n; Y^n, S^n|s_0) \geq \min_{s_0\in\mathcal{S}} I(U^n; Z^n, S^n|s_0)$, which implies that small probability of error in decoding M_2 at Rx_2 guarantees a small probability of error in decoding M_2 in Rx_1. This definition is more specialized than definition 7 since it does not lead to (7). This is because knowledge of the state sequence s^n is needed in (25) but not in (7).

6) We note that for finite-state channels, the rate expressions usually involve limits of mutual information expressions taken with blocklength increasing to infinity. Efficient computation of these limits is not considered in this paper. The computation of such limits has been considered by other authors, see for example [19]. It is also possible to bound the capacity of FSCs using a finite blocklength as was done by Gallager in [5, eq. (4.6.11)].

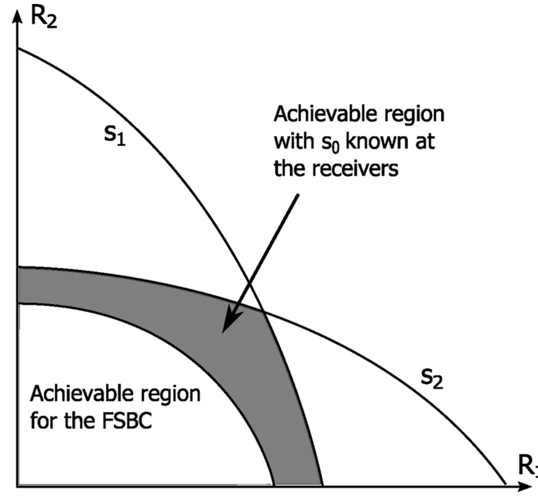

Fig. 4. The achievable region is a subset of the intersection of the achievable regions obtained with initial state known only at the receivers.

7) In order to demonstrate that multiletter expressions, i.e., coding with memory, provide benefit over memoryless codebooks (generated i.i.d.) for FSBCs, consider the following example: let $\tilde{\mathcal{S}} = \{1, 2, 3, \Phi\}$, $\mathcal{S} = \tilde{\mathcal{S}} \times \mathcal{X}$, $\|\mathcal{X}\| = 2$, and $S_i = (\tilde{S}_i, X_i)$. This can model a channel whose state is affected by the previous inputs (ISI) as well as by some "internal" process $\tilde{S}$. Assume that we are interested in sending only a common message to both receivers. This simplifies the exposition as we do not need to consider the auxiliary RV U^n.
Let $\tilde{S} = \Phi$ be a "short circuit" state: once the channel arrives to this state the outputs become zero and the internal state $\tilde{S}$ remains "stuck" at Φ (this can model a failure of a hard disk in a magnetic recording system). In terms of the per-symbol p.m.f., $p(y, z, s|x, s') = p(s|x, s')p(y, z|x, s, s')$, this can be written as

$$\begin{aligned} p\big(S_i = (\Phi, X_i)\big|X_i, S_{i-1} = (\Phi, X_{i-1})\big) &= 1 \\ p\big(Y_i = 0, Z_i = 0\big|X_i, S_i, S_{i-1} = (\Phi, X_{i-1})\big) &= 1. \end{aligned}$$

Next, let

$$p\big(S_i = (\Phi, 1)\big|X_i = 1, X_{i-1} = 1, \tilde{S}_{i-1}\big) = 1$$

thus, once $X_i = X_{i-1} = 1$ the channel switches into the "short circuit" mode (e.g., the average current is too strong). It is easy to see the following

- For codebooks generated i.i.d: if a codebook is generated i.i.d such that $p(x_i = 1) > 0$ then as $n \to \infty$ the probability that every codeword will contain the pair $(1, 1)$ approaches 1. This means that as $n \to \infty$ the probability that the achievable rate is zero approaches 1. Alternatively, if we try to avoid such situations by setting $p(x_i = 1) = 0$ the achievable rate is again zero. Thus, the rate with codebooks generated i.i.d. is zero.
- For codebooks generated with memory: in this case, all sequences that contain $(1, 1)$ can be *a priori* eliminated by assigning them zero probability. This allows to transmit information at positive rates.

IV. Conclusion

We defined the degraded finite-state broadcast channel and derived its capacity region for both indecomposable and nonindecomposable channels. Specifically, we first defined the notion of degradedness for broadcast channels with memory. Then we presented two possible definitions for the error probability, one based on the worst-case initial state and the other based on the averaged probability of error with the average taken w.r.t the initial state distribution. Averaged error probability is best suited for indecomposable channels, whereas the worst case error probability is more appropriate for nonindecomposable broadcast channels. When the channel is indecomposable, in fact, both definitions lead to the same capacity region. This is because for indecomposable channels, as time evolves the state distribution converges to a single distribution, and hence, for asymptotically large blocklengths the initial state does not affect the achievable rates.

Next, we discussed different possible ways to define degradedness for channels with memory. We derived the capacity region of general degraded FSBCs (the channel is general in the sense that there is no constraint on the state transitions) and of indecomposable degraded FSBCs under the worst-case criteria. As in the discrete, memoryless case, the capacity-achieving strategy uses a superposition codebook. However, as the channel has memory, we have to consider a correlated distribution over the entire codeword (i.e., a superposition codebook with memory), instead of generating codewords by independently selecting code symbols. A key element affected by this structure is the proof of convergence of the achievable region. Convergence is important as otherwise it may be that increasing the blocklength will decrease the achievable region, which means that this channel does not support reliable communication in the standard sense.

In future work, we intend to study the effect of feedback on the capacity region of the FSBC. It is well known that in the discrete, memoryless, degraded BC feedback does not increase capacity [27]. However, since for point-to-point FSCs feedback helps [8], then also for degraded FSBCs feedback can increase the capacity region.

Appendix A
Converse for *Theorem* 1

In the derivation of the converse we consider only the two private messages case since the common message can be incorporated by splitting the rate to Rx_2 into private and common rates, as in [22, Theorem 14.6.4]. The converse is stated in the following lemma.

Lemma A.1: If for some $\lambda > 0$ and some $\epsilon > 0$,

$$R_2 + \lambda R_1 \geq C^{\infty}(\lambda) + \epsilon \tag{A.1}$$

where $C^{\infty}(\lambda)$ is defined in (18), then there exist initial states $\tilde{s}_{0,n}, \tilde{s}'_{0,n} \in \mathcal{S}$ for which

$$P_{e2}^{(n)}(\tilde{s}_{0,n})R_2 + \lambda P_{e1}^{(n)}(\tilde{s}'_{0,n})R_1 \geq \epsilon - \frac{1}{n}(1+\lambda)(1+\log_2 \|\mathcal{S}\|). \tag{A.2}$$

The implication of (A.2), as explained in [16] for the DMBC, is that for large values of n at least one of the states $\tilde{s}_{0,n}$, $\tilde{s}'_{0,n}$ results in a probability of error (at the respective receiver) that is bounded away from zero. Hence, for large values of n, $\max_{s_0\in\mathcal{S}} P_e^{(n)}(s_0)$ cannot be made arbitrarily small[5], outside the region whose boundary is given in (19).

Proof: Recall that $P_{e2}^{(n)}(s_0)$ and $P_{e1}^{(n)}(s_0)$ denote the probabilities of error at Rx_2 and Rx_1, respectively, when the initial state s_0 is not available at the receivers and transmitter. From Fano's inequality (see [5, eq. 4.6.16]) we have that for any given initial state s_0

$$H(M_2|Z^n, s_0) \leq P_{e2}^{(n)}(s_0)nR_2 + 1 \tag{A.3a}$$
$$H(M_1|Y^n, s'_0) \leq P_{e1}^{(n)}(s'_0)nR_1 + 1. \tag{A.3b}$$

Denote with $\tilde{s}_{0,n}$ the initial channel state that maximizes $H(M_2|Z^n, s_0)$ and with $\tilde{s}'_{0,n}$ the initial channel state that maximizes $H(M_1|Y^n, s'_0)$. Now, note that

$$\begin{aligned}\min_{s_0\in\mathcal{S}} & I(M_2; Z^n|s_0) \\ &= \min_{s_0\in\mathcal{S}} \{H(M_2|s_0) - H(M_2|Z^n, s_0)\} \\ &= nR_2 - \max_{s_0\in\mathcal{S}} H(M_2|Z^n, s_0)\end{aligned} \tag{A.4}$$

$$\begin{aligned}\min_{s'_0\in\mathcal{S}} & I(M_1; Y^n|M_2, s'_0) \\ &= \min_{s'_0\in\mathcal{S}} \{H(M_1|M_2, s'_0) - H(M_1|Y^n, M_2, s'_0)\} \\ &= nR_1 - \max_{s'_0\in\mathcal{S}} H(M_1|Y^n, M_2, s'_0) \\ &\geq nR_1 - \max_{s'_0\in\mathcal{S}} H(M_1|Y^n, s'_0).\end{aligned} \tag{A.5}$$

Next, we show that

$$\min_{s_0\in\mathcal{S}} I(M_2; Z^n|s_0) + \lambda \min_{s'_0\in\mathcal{S}} I(M_1; Y^n|M_2, s'_0) \leq nC^n(\lambda). \tag{A.6}$$

First, by letting $U_i = M_2, i = 1, 2, \ldots, n$ we can write

$$I(M_2; Z^n|s_0) = I(U^n; Z^n|s_0).$$

We also have that

$$\begin{aligned}I(M_1; Y^n|M_2, s'_0) &= H(Y^n|M_2, s'_0) - H(Y^n|M_1, M_2, s'_0) \\ &\leq H(Y^n|M_2, s'_0) - H(Y^n|X^n, M_1, M_2, s'_0) \\ &= H(Y^n|U^n, s'_0) - H(Y^n|X^n, U^n, s'_0) \\ &= I(X^n; Y^n|U^n, s'_0)\end{aligned}$$

[5]Let s_0^* achieve the maximum in $\max_{s_0\in\mathcal{S}} P_e^{(n)}(s_0)$. Then $\max_{s_0\in\mathcal{S}} P_e^{(n)}(s_0) \geq \max\left\{P_{e1}^{(n)}(s_0^*), P_{e2}^{(n)}(s_0^*)\right\} \geq \delta > 0.$

where our definition of U^n satisfies the Markov relationship $U^n|s_0' \leftrightarrow X^n|s_0' \leftrightarrow Y^n|s_0'$.[6] Combining both derivations we obtain that for our choice of U^n:

$$\begin{aligned}\min_{s_0\in\mathcal{S}} I(M_2;Z^n|s_0) &+ \lambda \min_{s_0'\in\mathcal{S}} I(M_1;Y^n|M_2,s_0') \\ &\le \min_{s_0\in\mathcal{S}} I(U^n;Z^n|s_0) + \lambda \min_{s_0'\in\mathcal{S}} I(X^n;Y^n|U^n,s_0') \\ &\stackrel{(a)}{\le} nC^n(\lambda)\end{aligned}$$

where (a) is because $C^n(\lambda)$ is obtained by maximizing over all joint distributions $p(u^n,x^n)$ subject to the cardinality constraint (14) (we show in Appendix B-G that this is enough to achieve the maximum), hence, (A.6) is verified. Therefore

$$\begin{aligned}\min_{s_0\in\mathcal{S}} I(M_2;Z^n|s_0) &+ \lambda \min_{s_0'\in\mathcal{S}} I(M_1;Y^n|M_2,s_0') \\ &\qquad - (1+\lambda)\log_2\|\mathcal{S}\| \\ &\le n\left(C^n(\lambda) - (1+\lambda)\frac{\log_2\|\mathcal{S}\|}{n}\right) \\ &\le nC^\infty(\lambda) \qquad (A.7)\end{aligned}$$

where (A.7) is shown in Appendix C [see (B.26)].

Plugging (A.4) and (A.5) into (A.7) yields

$$\begin{aligned}&nR_2 - H(M_2|Z^n,\tilde{s}_{0,n}) + \lambda(nR_1 - H(M_1|Y^n,\tilde{s}_{0,n}')) \\ &\qquad - (1+\lambda)\log_2\|\mathcal{S}\| \le nC^\infty(\lambda) \\ &\Rightarrow H(M_2|Z^n,\tilde{s}_{0,n}) + \lambda H(M_1|Y^n,\tilde{s}_{0,n}') \\ &\qquad + (1+\lambda)\log_2\|\mathcal{S}\| \ge n\left(R_2 + \lambda R_1 - C^\infty(\lambda)\right) \\ &\Rightarrow H(M_2|Z^n,\tilde{s}_{0,n}) + \lambda H(M_1|Y^n,\tilde{s}_{0,n}') \\ &\ge n\left(\epsilon - (1+\lambda)\frac{\log_2\|\mathcal{S}\|}{n}\right),\end{aligned}$$

where the last line follows from (A.1). Combined with Fano's inequalities (A.3), we have

$$\begin{aligned}P_{e2}^{(n)}(\tilde{s}_{0,n})nR_2 + 1 + \lambda\left(P_{e1}^{(n)}(\tilde{s}_{0,n}')nR_1 + 1\right) \\ \ge n\epsilon - (1+\lambda)\log_2\|\mathcal{S}\|\end{aligned}$$

which completes the proof. ■

Appendix B
Achievability of the Rates of *Theorem* 1

We prove that all the rates of Theorem 1 are achievable for the physically degraded FSBC. Recall that for physically degraded FSBCs it was shown in Section II-C that condition (7) holds. In the derivation we shall consider transmission of only two private messages. We also recall that the transmitter and receivers do not know the channel states.

In order to show existence of an achievable rate we bound the probability of error for initial state s_0, averaged over all codebooks, denoted by $\bar{P}_e^{(n)}(s_0)$

$$\begin{aligned}\bar{P}_e^{(n)}(s_0) &= \bar{P}_{e1}^{(n)}(M_1|s_0) + \bar{P}_{e2}^{(n)}(M_2|s_0) \\ &\le \bar{P}_{e1}^{(n)}(M_1,M_2|s_0) + \bar{P}_{e2}^{(n)}(M_2|s_0) \\ &\le \bar{P}_{e12}^{(n)}(M_2|s_0) + \bar{P}_{e11}^{(n)}(M_1|M_2,s_0) \\ &\qquad + \bar{P}_{e2}^{(n)}(M_2|s_0) \qquad (B.1)\end{aligned}$$

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

In the following sections we provide bounds on the three error probabilities in (B.1).

A. Code Construction and Encoding

- Fix a blocklength n and a joint distribution $p(u^n,x^n)$. For each message $m_2\in\mathcal{M}_2$, independently generate a codeword according to $\Pr(\mathbf{u}(m_2)=u^n) = p_{U^n}(u^n)$.
- For each message $m_2\in\mathcal{M}_2$ generate a codebook with 2^{nR_1} codewords $\mathbf{x}(m_1,m_2)$, $m_1\in\mathcal{M}_1$ where each codeword is generated according to the probability $\Pr(\mathbf{x}(m_1,m_2)=x^n) = p_{X^n|U^n}(x^n|u^n(m_2))$, and is selected independently of the codewords $\{\mathbf{x}(m_1',m_2)\}_{m_1'\ne m_1}$.
- For transmitting the message pair (m_1,m_2) the transmitter outputs $\mathbf{x}(m_1,m_2)$.

B. Decoding at the Weak Receiver Rx_2

A natural way to construct a decoding rule for the BC with memory is to extend the decoding rule used for the DMBC to the scenario with memory. As in [5, Sec. 5.9], the decoder handles its ignorance of the initial state by averaging over all initial channel states $s_0\in\mathcal{S}$. The derivation follows the essential steps of [5, Sec. 5.9].

First, define

$$\begin{aligned}\tilde{p}(z^n|u^n) &\triangleq \sum_{s_0\in\mathcal{S}} \frac{1}{\|\mathcal{S}\|} p(z^n|u^n,s_0) \\ &= \sum_{s_0\in\mathcal{S}} \frac{1}{\|\mathcal{S}\|} \sum_{\mathcal{Y}^n,\mathcal{X}^n,\mathcal{S}^n} p(z^n,y^n,x^n,s^n|u^n,s_0) \\ &\triangleq \sum_{s_0\in\mathcal{S}} \frac{1}{\|\mathcal{S}\|} \sum_{\mathcal{Y}^n,\mathcal{X}^n,\mathcal{S}^n} p_{X^n|U^n}(x^n|u^n) \\ &\qquad \times p(z^n,y^n,s^n|x^n,s_0). \qquad (B.2)\end{aligned}$$

In (B.2), we used the $\triangleq$ notation since the decoder ignores its knowledge of the codebook $\mathbf{x}$ and averages over all $\mathbf{x}$ vectors according to the conditional distribution $p_{X^n|U^n}(x^n|u^n)$. Let the decoding rule be the approximate ML decoder according to $\tilde{p}(z^n|u^n)$. Then, for a received sequence $\mathbf{z}$,

$$\begin{aligned}\text{if } \forall m_2'\in\mathcal{M}_2, m_2'\ne\hat{m}_2 \quad \tilde{p}(\mathbf{z}|\mathbf{u}(\hat{m}_2)) \ge \tilde{p}(\mathbf{z}|\mathbf{u}(m_2')) \\ \Rightarrow g_{\mathbf{z}}(\mathbf{z}) = \hat{m}_2 \qquad (B.3)\end{aligned}$$

with ties broken arbitrarily. Note that the probability $\tilde{p}(\mathbf{z}|\mathbf{u}(m_2))$, for a *given* $\mathbf{u}(m_2)$ is obtained by averaging over all possible *random* codewords $\mathbf{x}$, state sequences $\mathbf{s}$ and received sequences at the strong receiver, $\mathbf{y}$. Therefore, this decoder is not the exact ML decoder as it ignores the knowledge

[6]The conditional Markov relationship means that the Markov chain holds only subject to fixing the value in the conditioning. In this specific case the Markov relationship implies $p(y^n|x^n,u^n,s_0') = p(y^n|x^n,s_0')$ for every initial state $s_0'\in\mathcal{S}$.

of the codebook $\left\{\mathbf{x}(m_1,m_2)\right\}_{m_1\in\mathcal{M}_1}$ and uses its generating statistics $p_{X^n|U^n}(x^n|u^n(m_2))$ instead, see also [16, Sec. 4] and [33, Sec. VII-B]. The decoder in (B.3) differs from the one in [16, Sec. 4] in that it does not use an i.i.d. approximation for the p.m.f. of a block of n transmissions.

Proceeding as in [5], we consider the error event

$$E_{m_2'} \triangleq \Big\{m_2' \text{ is decoded when } m_2 \text{ is transmitted using codeword } \mathbf{u}(m_2), \text{ and } \mathbf{z} \text{ is received at } \mathrm{Rx}_2\Big\}.$$

The probability of $E_{m_2'}$ averaged over all possible selections of $\mathbf{u}(m_2')$ is[7]

$$\begin{aligned}\Pr(E_{m_2'}) &= \Pr\left(m_2' \text{ decoded } |m_2,\mathbf{u}(m_2),\mathbf{z}\right)\\ &\overset{(a)}{=} \sum_{\mathbf{u}(m_2'):\tilde{p}(\mathbf{z}|\mathbf{u}(m_2'))\geq\tilde{p}(\mathbf{z}|\mathbf{u}(m_2))} p(\mathbf{u}(m_2'))\\ &\leq \sum_{\mathbf{u}(m_2')\in\times_{i=1}^n\mathcal{U}_i} p(\mathbf{u}(m_2'))\frac{\tilde{p}(\mathbf{z}|\mathbf{u}(m_2'))^s}{\tilde{p}(\mathbf{z}|\mathbf{u}(m_2))^s} \end{aligned}\tag{B.4}$$

with $s > 0$. Here, (a) holds since $g_{\mathbf{z}}(\mathbf{z}) = m_2'$ implies $\tilde{p}(\mathbf{z}|\mathbf{u}(m_2')) \geq \tilde{p}(\mathbf{z}|\mathbf{u}(m_2))$.

Following similar steps to [5, Ch. 5.6] we can use (B.4) to bound the probability of error when m_2 is transmitted using $\mathbf{u}(m_2)$ and $\mathbf{z}$ is received by

$$\begin{aligned}&\Pr\left(\text{error } |m_2,\mathbf{u}(m_2),\mathbf{z}\right)\\ &= \Pr\left(\bigcup_{m_2'\neq m_2} E_{m_2'}\right)\\ &\leq \left[||\mathcal{M}_2|| \sum_{\tilde{\mathbf{u}}\in\times_{i=1}^n\mathcal{U}_i} p(\tilde{\mathbf{u}})\frac{\left(\sum_{s_0\in\mathcal{S}}\frac{1}{||\mathcal{S}||}p(\mathbf{z}|\tilde{\mathbf{u}},s_0)\right)^s}{\left(\sum_{s_0\in\mathcal{S}}\frac{1}{||\mathcal{S}||}p(\mathbf{z}|\mathbf{u}(m_2),s_0)\right)^s}\right]^\rho\end{aligned}\tag{B.5}$$

where $0<\rho\leq 1$.

The average probability of error for a given s_0, averaged over all codewords $\mathbf{u}(m_2)$ and received sequences $\mathbf{z}$ can be written as

$$\begin{aligned}&\bar{P}_{e2}^{(n)}(m_2|s_0)\\ &\quad= \sum_{\mathbf{u}(m_2)\in\times_{i=1}^n\mathcal{U}_i} p(\mathbf{u}(m_2))\sum_{\mathcal{Z}^n} p(\mathbf{z}|\mathbf{u}(m_2),s_0)\\ &\qquad\qquad\times \Pr\left(\text{ error } |m_2,\mathbf{u}(m_2),\mathbf{z}\right)\\ &\quad\leq ||\mathcal{S}||\sum_{\mathcal{Z}^n}\sum_{\mathbf{u}(m_2)\in\times_{i=1}^n\mathcal{U}_i} p(\mathbf{u}(m_2))\\ &\quad\times\left(\sum_{s_0\in\mathcal{S}}\frac{1}{||\mathcal{S}||}p(\mathbf{z}|\mathbf{u}(m_2),s_0)\right)\Pr\left(\text{error}|m_2,\mathbf{u}(m_2),\mathbf{z}\right)\end{aligned}$$

as summing over all initial states s_0 results in an upper bound. Using the bound (B.5) in this expression, we can upper bound the average probability of error for the message m_2 for *any* s_0 by

$$\begin{aligned}&\bar{P}_{e2}^{(n)}(m_2|s_0)\\ &\quad\leq ||\mathcal{S}||\sum_{\mathcal{Z}^n}\sum_{\mathbf{u}(m_2)\in\times_{i=1}^n\mathcal{U}_i} p(\mathbf{u}(m_2))\tilde{p}(\mathbf{z}|\mathbf{u}(m_2))\times\\ &\qquad\left[(||\mathcal{M}_2||-1)\sum_{\tilde{\mathbf{u}}\in\times_{i=1}^n\mathcal{U}_i} p(\tilde{\mathbf{u}})\frac{\tilde{p}(\mathbf{z}|\tilde{\mathbf{u}})^s}{\tilde{p}(\mathbf{z}|\mathbf{u}(m_2))^s}\right]^\rho\\ &\quad\overset{(a)}{\leq} ||\mathcal{S}||\cdot||\mathcal{M}_2||^\rho||\mathcal{S}||^\rho\\ &\qquad\times\max_{s_0\in\mathcal{S}}\sum_{\mathcal{Z}^n}\left[\sum_{\times_{i=1}^n\mathcal{U}_i} p(\mathbf{u})p(\mathbf{z}|\mathbf{u},s_0)^{\frac{1}{1+\rho}}\right]^{1+\rho}\end{aligned}\tag{B.6}$$

where in (a) we set $s=\frac{1}{1+\rho}$ (see [5, Sec. 5.6]). Note that the bound on $\bar{P}_{e2}^{(n)}(m_2|s_0)$ is independent of the particular s_0, hence it bounds the average probability of error for all messages and initial states.

Next, define

$$\begin{aligned}&E_{n,2}(\rho,p(\mathbf{u},\mathbf{x}),s_0)\\ &\qquad\triangleq -\frac{1}{n}\log_2\sum_{\mathcal{Z}^n}\left[\sum_{\times_{i=1}^n\mathcal{U}_i} p(\mathbf{u})p(\mathbf{z}|\mathbf{u},s_0)^{\frac{1}{1+\rho}}\right]^{1+\rho}\\ &F_{n,2}(\rho,p(\mathbf{u},\mathbf{x}))\\ &\qquad\triangleq \min_{s_0\in\mathcal{S}} E_{n,2}(\rho,p(\mathbf{u},\mathbf{x}),s_0)-\rho\frac{\log_2||\mathcal{S}||}{n}\end{aligned}$$

$0\leq\rho\leq 1$. Then, from (B.6), the probability of error averaged over all codebooks generated according to the distribution $p(\mathbf{u},\mathbf{x})$ is given by

$$\bar{P}_{e2}^{(n)}(m_2|s_0)\leq||\mathcal{S}||2^{-n(F_{n,2}(\rho,p(\mathbf{u},\mathbf{x}))-\rho R_2)}.\tag{B.7}$$

The next step is to show that for a fixed initial state s_0, $\exists\rho$, $0<\rho\leq 1$ for which $E_{n,2}(\rho,p(\mathbf{u},\mathbf{x}),s_0)-\rho\frac{\log_2||\mathcal{S}||}{n}-\rho R_2$ is positive as long as

$$R_2<\frac{1}{n}I(U^n;Z^n|s_0)-\frac{\log_2||\mathcal{S}||}{n}.$$

First, we note that the maximum of $E_{n,2}(\rho,p(\mathbf{u},\mathbf{x}),s_0)-\rho\frac{\log_2||\mathcal{S}||}{n}-\rho R_2$ versus ρ can be found by equating the first derivative vs. ρ to zero, as long as the second derivative is negative. The first derivative is

$$\begin{aligned}&\frac{\partial}{\partial\rho}\left(E_{n,2}(\rho,p(\mathbf{u},\mathbf{x}),s_0)-\rho\frac{\log_2||\mathcal{S}||}{n}-\rho R_2\right)\\ &\qquad=\frac{\partial}{\partial\rho}E_{n,2}(\rho,p(\mathbf{u},\mathbf{x}),s_0)-\frac{\log_2||\mathcal{S}||}{n}-R_2.\end{aligned}$$

Since $\mathbf{u}$ is selected independently of the initial state, then $p(\mathbf{u})=p(\mathbf{u}|s_0)$ and we arrive at

$$\frac{\partial}{\partial\rho}E_{n,2}(\rho,p(\mathbf{u},\mathbf{x}),s_0)\bigg|_{\rho=0}=\frac{1}{n}I(U^n;Z^n|s_0).$$

Lastly, extending the technique for discrete, memoryless, point-to-point channels [5, Theorem 5.6.3] to the case with

[7] As $\mathbf{z}$ is given and decoding is done using $\tilde{p}(\mathbf{z}|\mathbf{u})$, the initial state does not affect $\Pr(E_{m_2'})$.

memory following the argument in the proof of [5, Lemma 5.9.2], it can be shown that for every initial state s_0, if

$$R_2 < \frac{1}{n} I(U^n; Z^n | s_0) - \frac{\log_2 ||\mathcal{S}||}{n}$$

we can find a $\rho^* > 0$ such that

$$E_{n,2}(\rho^*, p(\mathbf{u}, \mathbf{x}), s_0) - \rho^* \frac{\log_2 ||\mathcal{S}||}{n} - \rho^* R_2 > 0.$$

We provide more details on this argument when considering the rate bound on R_1 in Appendix B-D.

Thus we obtain that for a given n, if

$$R_2 \le \min_{s_0 \in \mathcal{S}} \frac{1}{n} I(U^n; Z^n | s_0) - \frac{\log_2 ||\mathcal{S}||}{n} \triangleq R_{2,n}(p) \qquad \text{(B.8)}$$

there exists $\rho^* > 0$ for which $F_{n,2}(\rho^*, p(\mathbf{u}, \mathbf{x})) - \rho^* R_2$ is positive, hence $\bar{P}_{e2}^{(n)}(m_2 | s_0)$ in (B.7) has a positive error exponent. As a last comment, note that the expression in (B.8) can be negative. However, in Appendix C we show that $\max_{p(u^n, x^n)} R_{2,n}(p)$ is sup-additive. Thus, as long as the limit $\lim_{n \to \infty} \max_{p(u^n, x^n)} R_{2,n}(p)$ is nonzero, then there exist some $p(u^n, x^n)$ and $N_0 \in \mathbb{N}$ such that for all $n > N_0$, $R_{2,n}(p) > 0$, and we can restrict our attention only to positive rates.

C. Decoding the Message m_2 at the Strong Receiver Rx_1

Define

$$\begin{aligned}
&\tilde{p}(y^n | u^n) \\
&\quad \triangleq \sum_{s_0 \in \mathcal{S}} \frac{1}{||\mathcal{S}||} p(y^n | u^n, s_0) \\
&\quad \triangleq \sum_{s_0 \in \mathcal{S}} \frac{1}{||\mathcal{S}||} \sum_{\mathcal{Z}^n, \mathcal{X}^n, \mathcal{S}^n} p_{X^n|U^n}(x^n | u^n) p(z^n, y^n, s^n | x^n, s_0)
\end{aligned}$$

where again, by definition the detector averages over all codewords $\mathbf{x}$ according to $p_{X^n|U^n}(x^n | u^n)$. Let the decoding rule be the approximate maximum-likelihood decoder according to $\tilde{p}(y^n | u^n)$

$$\begin{aligned}
&\text{if } \forall m_2' \in \mathcal{M}_2, m_2' \neq \hat{m}_2 \quad \tilde{p}(\mathbf{y} | \mathbf{u}(\hat{m}_2)) \ge \tilde{p}(\mathbf{y} | \mathbf{u}(m_2')) \\
&\qquad \Rightarrow \mathrm{Rx}_1 \text{ decides on } \hat{m}_2
\end{aligned}$$

with ties broken arbitrarily. The average probability of error when m_2 is transmitted using $\mathbf{u}(m_2)$, $\{m_2'\}_{m_2' \in \mathcal{M}_2, m_2' \neq m_2}$ is transmitted using $\{\mathbf{u}(m_2')\}_{m_2' \in \mathcal{M}_2, m_2' \neq m_2}$, respectively, and s_0 is the initial state is given by

$$\begin{aligned}
&\bar{P}_{e12}^{(n)}\left(m_2 | \mathbf{u}(m_2), \{\mathbf{u}(m_2')\}_{m_2' \neq m_2}, s_0\right) \\
&\quad = \sum_{\substack{\mathbf{y}: \exists m_2' \neq m_2, \\ \tilde{p}(\mathbf{y} | \mathbf{u}(m_2')) \ge \tilde{p}(\mathbf{y} | \mathbf{u}(m_2))}} \sum_{\mathcal{Z}^n, \mathcal{X}^n, \mathcal{S}^n} p(\mathbf{z}, \mathbf{y}, \mathbf{x}, \mathbf{s} | \mathbf{u}(m_2), s_0).
\end{aligned}$$

Applying the same steps used in the derivation of the bound on the probability of error at Rx_2 to bound $\bar{P}_{e12}^{(n)}(m_2 | s_0)$, we conclude that the error exponent for decoding the message m_2 at Rx_1 can be made positive for every $s_0 \in \mathcal{S}$ as long as

$$R_2 \le \min_{s_0 \in \mathcal{S}} \frac{1}{n} I(U^n; Y^n | s_0) - \frac{\log_2 ||\mathcal{S}||}{n}. \qquad \text{(B.9)}$$

However, using the physical degradedness (7) is follows that

$$\begin{aligned}
p(y^n, z^n | u^n, s_0) &= \sum_{\mathcal{X}^n} p(y^n, z^n, x^n | u^n, s_0) \\
&= p(y^n | u^n, s_0) p(z^n | y^n, s_0), \quad \forall s_0 \in \mathcal{S}
\end{aligned}$$

and by the data processing inequality we obtain $I(U^n; Y^n | s_0') \ge I(U^n; Z^n | s_0')$, where s_0' minimizes (B.9). Thus

$$\begin{aligned}
\min_{s_0 \in \mathcal{S}} \frac{1}{n} I(U^n; Z^n | s_0) &\le \frac{1}{n} I(U^n; Z^n | s_0') \\
&\le \frac{1}{n} I(U^n; Y^n | s_0') = \min_{s_0 \in \mathcal{S}} \frac{1}{n} I(U^n; Y^n | s_0).
\end{aligned}$$

Recall that if $R_2 < R_{2,n}(p)$ then for decoding m_2 at Rx_2 it is possible to find a $\rho^* > 0$ that results in a positive error exponent. The above implies that in these circumstances it is possible to find a (possibly different) $\rho^* > 0$ that results in a positive error exponent for decoding m_2 at Rx_1, hence $\bar{P}_{e12}^{(n)}(M_2 | s_0)$ can be made arbitrarily small by increasing n.

D. Decoding the Message m_1 at the Strong Receiver Rx_1

Decoding m_1 at Rx_1 takes place after it decoded m_2. Denote with $\hat{\hat{m}}_2$ the message for Rx_2 that was decoded at Rx_1. Next, define

$$\tilde{p}(y^n | x^n) \triangleq \sum_{\mathcal{Z}^n, \mathcal{S}^n, s_0 \in \mathcal{S}} \frac{1}{||\mathcal{S}||} p(z^n, y^n, s^n | x^n, s_0).$$

Let the decoding rule be the maximum-likelihood decoder according to $\tilde{p}(y^n | x^n)$:

$$\begin{aligned}
&\text{if } \forall m_1' \in \mathcal{M}_1, m_1' \neq \hat{m}_1 \\
&\qquad \tilde{p}(\mathbf{y} | \mathbf{x}(\hat{m}_1, \hat{\hat{m}}_2)) \ge \tilde{p}(\mathbf{y} | \mathbf{x}(m_1', \hat{\hat{m}}_2)) \Rightarrow g_{\mathbf{y}}(\mathbf{y}) = \hat{m}_1
\end{aligned}$$

with ties broken arbitrarily. The probability of error, for a fixed codebook, when m_1 is transmitted, $\hat{\hat{m}}_2$ is decoded at Rx_1 and the initial state is s_0, is given by

$$\begin{aligned}
&P_{e11}^{(n)}(m_1 | \hat{\hat{m}}_2, s_0) = \\
&\qquad \sum_{\substack{\mathbf{y}: \exists m_1' \neq m_1, \\ \tilde{p}(\mathbf{y} | \mathbf{x}(m_1', \hat{\hat{m}}_2)) \ge \tilde{p}(\mathbf{y} | \mathbf{x}(m_1, \hat{\hat{m}}_2))}} \sum_{\mathcal{Z}^n, \mathcal{S}^n} p(\mathbf{z}, \mathbf{y}, \mathbf{s} | \mathbf{x}(m_1, \hat{\hat{m}}_2), s_0).
\end{aligned}$$

We proceed assuming that $\hat{\hat{m}}_2 = m_2$. The bound on the probability of error averaged over all selections of the codebook $\left\{\mathbf{x}(m_1, m_2)\right\}_{m_1 \in \mathcal{M}_1}$ for a fixed $\mathbf{u}(m_2)$ (i.e., the codebooks are generated according to $p_{X^n|U^n}(x^n | u^n = \mathbf{u}(m_2))$) is given by [cf. (B.6)]

$$\begin{aligned}
&\bar{P}_{e11}^{(n)}(m_1 | \mathbf{u}(m_2), s_0) \le ||\mathcal{S}|| \left(||\mathcal{M}_1|| - 1\right)^{\rho} \\
&\quad \times \sum_{\mathcal{Y}^n} \left[\sum_{\mathcal{X}^n} p(\mathbf{x} | \mathbf{u}(m_2)) \left(\sum_{s_0 \in \mathcal{S}} \frac{1}{||\mathcal{S}||} p(\mathbf{y} | \mathbf{x}, s_0) \right)^{\frac{1}{1+\rho}} \right]^{1+\rho}.
\end{aligned}$$

Averaging over all possible selections of $\mathbf{u}(m_2)$ we finally obtain

$$\bar{P}_{e11}^{(n)}(m_1 | m_2, s_0) = \sum_{\times_{i=1}^{n} \mathcal{U}_i} p(\mathbf{u}(m_2)) \bar{P}_{e11}^{(n)}(m_1 | \mathbf{u}(m_2), s_0)$$

which is the probability of error averaged over all codebooks. Bounding the expression using similar steps to those leading to (B.6) we obtain

$$\begin{aligned}
&\bar{P}_{e11}^{(n)}(m_1|m_2,s_0)\\
&\leq ||\mathcal{S}||\cdot||\mathcal{M}_1||^{\rho}\sum_{\times_{i=1}^n\mathcal{U}_i}p(\mathbf{u}(m_2))\\
&\quad\times\sum_{\mathcal{Y}^n}\left[\sum_{\mathcal{X}^n}p(\mathbf{x}|\mathbf{u}(m_2))\left(\sum_{s_0\in\mathcal{S}}\frac{1}{||\mathcal{S}||}p(\mathbf{y}|\mathbf{x},s_0)\right)^{\frac{1}{1+\rho}}\right]^{1+\rho}\\
&\leq ||\mathcal{S}||\cdot||\mathcal{M}_1||^{\rho}||\mathcal{S}||^{\rho}\\
&\quad\times\max_{s_0\in\mathcal{S}}\sum_{\times_{i=1}^n\mathcal{U}_i}p(\mathbf{u})\sum_{\mathcal{Y}^n}\left[\sum_{\mathcal{X}^n}p(\mathbf{x}|\mathbf{u})p(\mathbf{y}|\mathbf{x},s_0)^{\frac{1}{1+\rho}}\right]^{1+\rho}.
\end{aligned}\tag{B.10}$$

As in the analysis for Rx_2, for every initial state s_0 we find the maximum $R_1(s_0)$ for which a positive error exponent exists. Then, minimizing $R_1(s_0)$ over all initial states guarantees a positive error exponent for any $s_0\in\mathcal{S}$. Using

$$E_{n,11}(\rho,p(\mathbf{u},\mathbf{x}),s_0)\triangleq -\frac{1}{n}\log_2\sum_{\times_{i=1}^n\mathcal{U}_i}p(\mathbf{u})\sum_{\mathcal{Y}^n}\left[\sum_{\mathcal{X}^n}p(\mathbf{x}|\mathbf{u})p(\mathbf{y}|\mathbf{x},s_0)^{\frac{1}{1+\rho}}\right]^{1+\rho}\tag{B.11}$$

$$F_{n,11}(\rho,p(\mathbf{u},\mathbf{x}))\triangleq\min_{s_0\in\mathcal{S}}E_{n,11}(\rho,p(\mathbf{u},\mathbf{x}),s_0)-\rho\frac{\log_2||\mathcal{S}||}{n}\tag{B.12}$$

we rewrite (B.10) as

$$\bar{P}_{e11}^{(n)}(m_1|m_2,s_0)\leq ||\mathcal{S}||2^{-n(F_{n,11}(\rho,p(\mathbf{u},\mathbf{x}))-\rho R_1)}.\tag{B.13}$$

To find a $\rho>0$ that results in $F_{n,11}(\rho,p(\mathbf{u},\mathbf{x}))-\rho R_1>0$ we first equate the first derivative of $E_{n,11}(\rho,p(\mathbf{u},\mathbf{x}),s_0)$ w.r.t. ρ to zero. The resulting derivative, evaluated at $\rho=0$ is

$$\frac{\partial}{\partial\rho}\log_2\sum_{\mathcal{Y}^n}\sum_{\times_{i=1}^n\mathcal{U}_i}p(\mathbf{u})\left[\sum_{\mathcal{X}^n}p(\mathbf{x}|\mathbf{u})p(\mathbf{y}|\mathbf{x},s_0)^{\frac{1}{1+\rho}}\right]^{1+\rho}\Bigg|_{\rho=0} = -\frac{1}{n}I(X^n;Y^n|U^n,s_0).$$

Next we show that if $R_1\leq\frac{1}{n}I(X^n;Y^n|U^n,s_0)-\frac{\log_2||\mathcal{S}||}{n}$, a positive error exponent can be found assuming that the initial state is s_0. The result is summarized in the following lemma.

Lemma B.1: For any initial state $s_0\in\mathcal{S}$ and superposition codebook satisfying the Markov relationship $U^n|s_0\leftrightarrow X^n|s_0\leftrightarrow Y^n|s_0$, as long as

$$R_1<\frac{1}{n}I(X^n;Y^n|U^n,s_0)-\frac{\log_2||\mathcal{S}||}{n}\tag{B.14}$$

then there exists $0<\rho^*(s_0)\leq 1$ such that

$$E_{n,11}(\rho^*(s_0),p(\mathbf{u},\mathbf{x}),s_0)-\rho^*(s_0)\frac{\log_2||\mathcal{S}||}{n}-\rho^*(s_0)R_1>0.$$

Then, for any $R_1\leq\min_{s_0'\in\mathcal{S}}\frac{1}{n}I(X^n;Y^n|U^n,s_0')-\frac{\log_2||\mathcal{S}||}{n}$ taking $\rho^*=\min_{s_0\in\mathcal{S}}\rho^*(s_0)$ we obtain

$$F_{n,11}(\rho^*,p(\mathbf{u},\mathbf{x}))-\rho^*R_1>0$$

i.e., the error exponent in (B.13) is positive.

Proof Outline: We highlight the main elements of the proof. The details of the proof are similar to [5, Appendix 5B]

- First, note that for any $s_0\in\mathcal{S}$, $\left[E_{n,11}(\rho,p(\mathbf{u},\mathbf{x}),s_0)-\rho\frac{\log_2||\mathcal{S}||}{n}-\rho R_1\right]\Big|_{\rho=0}=0.$
- We show that $E_{n,11}(\rho,p(\mathbf{u},\mathbf{x}),s_0)-\rho\frac{\log_2||\mathcal{S}||}{n}-\rho R_1$ is concave in ρ for any $s_0\in\mathcal{S}$. Thus, if it decreases with ρ at $\rho=0$, it will keep decreasing and the error exponent will be negative [i.e., the bound in (B.13) is greater than 1]. Therefore, for the error bound to be useful, this expression must increase with ρ at $\rho=0$.
- Next we note that for a given $s_0\in\mathcal{S}$ the derivative at zero is $\left[\frac{\partial E_{n,11}(\rho,p(\mathbf{u},\mathbf{x}),s_0)}{\partial\rho}-\frac{\log_2||\mathcal{S}||}{n}-R_1\right]\Big|_{\rho=0}$. Making this derivative positive yields an upper bound on R_1.
- Furthermore, $\frac{\partial E_{n,11}(\rho,p(\mathbf{u},\mathbf{x}),s_0)}{\partial\rho}$ is analytic in ρ for $\rho\geq 0$. Therefore, $E_{n,11}(\rho,p(\mathbf{u},\mathbf{x}),s_0)-\rho\frac{\log_2||\mathcal{S}||}{n}-\rho R_1$ is continuous.
- A positive derivative at zero, combined with the continuity of the expression and its first derivative, implies that there is a region of positive ρ for which the error exponent is positive. Taking any $\rho>0$ in this region yields a positive error exponent.

Since $R_1\leq\frac{1}{n}I(X^n;Y^n|U^n,s_0)-\frac{\log_2||\mathcal{S}||}{n}$ results in a positive error exponent for initial state s_0, thus letting

$$R_1\leq\min_{s_0'\in\mathcal{S}}\frac{1}{n}I(X^n;Y^n|U^n,s_0')-\frac{\log_2||\mathcal{S}||}{n}\triangleq R_{1,n}(p)$$

guarantees the existence of a positive error exponent for every initial state, and therefore the error exponent with $F_{n,11}(\rho,p(\mathbf{u},\mathbf{x}))$ in (B.12) is positive. As in Appendix II-B, we deal with negative $R_{1,n}(p)$ by taking n large enough.

Combining with Appendix II-C it follows that if also $R_2\leq R_{2,n}(p)$ then it is possible to achieve positive error exponents for both $\bar{P}_{e11}^{(n)}(m_1|m_2,s_0)$ and $\bar{P}_{e12}^{(n)}(m_2|s_0)$, simultaneously.

E. Proving That the Rate Pair $(R_{1,n}(p),R_{2,n}(p))$ is Achievable

The results of Appendices II-B – II-D imply that for physically degraded FSBCs, any rate pair (R_1,R_2) that belongs to the convex hull of the region

$$\bigcup_{p(u^n,x^n)}\Bigg\{(R_1,R_2)\in\mathbb{R}_+^2:$$

$$R_1\leq\min_{s_0'\in\mathcal{S}}\frac{1}{n}I(X^n;Y^n|U^n,s_0')-\frac{\log_2||\mathcal{S}||}{n}\tag{B.15}$$

$$R_2\leq\min_{s_0\in\mathcal{S}}\frac{1}{n}I(U^n;Z^n|s_0)-\frac{\log_2||\mathcal{S}||}{n}\Bigg\}\tag{B.16}$$

results in positive error exponents.

We now show that the rates of (B.15) – (B.16) are achievable. It is enough to show that given a maximum average probability of error $\epsilon>0$, then for the positive rate-pair

$$(R_{1,n}(p),R_{2,n}(p))=\Bigg(\min_{s_0'\in\mathcal{S}}\frac{1}{n}I(X^n;Y^n|U^n,s_0')-\frac{\log_2||\mathcal{S}||}{n}\quad\min_{s_0\in\mathcal{S}}\frac{1}{n}I(U^n;Z^n|s_0)-\frac{\log_2||\mathcal{S}||}{n}\Bigg)$$

we can find a blocklength b_0 such that for all $b>b_0$ a code $(R_{1,n}(p),R_{2,n}(p),b)$ with $\max_{s_0\in\mathcal{S}}\bar{P}_e^{(b)}(s_0)\leq\epsilon$ can be constructed.

Therefore, the bound on the probability of error satisfies the achievability criteria of Definition 5. Let $b=Kn$ for some

$K \in \mathbb{N}$, and let $q_n(u^n, x^n)$ be the probability distribution used to determine $R_{1,n}(q_n)$ and $R_{2,n}(q_n)$. Set the probability for generating the length b codewords to

$$p(u^b, x^b) = \prod_{k=1}^{K} q_n(u_{(k-1)n+1}^{kn}, x_{(k-1)n+1}^{kn}) \triangleq \prod_{k=1}^{K} q_n(\mathbf{u}_k, \mathbf{x}_k). \tag{B.17}$$

We now prove the following lemma:

Lemma B.2: For $F_{n,11}(\rho, p(u^b, x^b))$ defined in (B.12) under the distribution law (B.17) it holds that

$$F_{b,11}\left(\rho, \prod_{k=1}^{K} q_n(\mathbf{u}_k, \mathbf{x}_k)\right) = F_{b,11}(\rho, p(u^b, x^b)) \geq F_{n,11}(\rho, q_n(u^n, x^n)).$$

Proof: The proof uses the same essential steps as in Gallager's derivation [5, Sec. 5.9] and [15, Lemma 19]. However we have to make sure that the main steps still hold also for a superposition codebook (this is not trivial, see Appendix IV).

Let $N = M + L$, and consider $F_M(\rho, p_M(u^M, x^M))$ and $F_L(\rho, p_L(u^L, x^L))$. Let $p_N(u^N, x^N) = p_M(u^M, x^M)p_L(u_{M+1}^N, x_{M+1}^N)$. We denote $(\mathbf{u}_1, \mathbf{x}_1) \triangleq (u^M, x^M)$ and $(\mathbf{u}_2, \mathbf{x}_2) \triangleq (u_{M+1}^N, x_{M+1}^N)$. Then,

$$F_{N,11}(\rho, p_N(u^N, x^N)) = \min_{s_0 \in \mathcal{S}} E_{N,11}(\rho, p_N(u^N, x^N), s_0) - \rho \frac{\log_2 ||\mathcal{S}||}{N}.$$

Let $s_{0,N}$ be the minimizing state. Raising both sides to the power of two we obtain from (B.11) that

$$2^{-NF_{N,11}(\rho, p_N(u^N, x^N))} = ||\mathcal{S}||^\rho \sum_{\times_{i=1}^N \mathcal{U}_i} p(u^N) \times \sum_{\mathcal{Y}^N} \left[\sum_{\mathcal{X}^N} p(x^N|u^N) p(y^N|x^N, s_{0,N})^{\frac{1}{1+\rho}}\right]^{1+\rho}. \tag{B.18}$$

Now consider the derivation from (B.19) shown at the bottom of the page to (B.20) shown at the top of the next page.

$$\sum_{\times_{i=1}^N \mathcal{U}_i} p(u^N) \sum_{\mathcal{Y}^N} \left[\sum_{\mathcal{X}^N} p(x^N|u^N) p(y^N|x^N, s_{0,N})^{\frac{1}{1+\rho}}\right]^{1+\rho} \tag{B.19}$$

$$\begin{aligned}
&\stackrel{(a)}{=} \sum_{\times_{i=1}^M \mathcal{U}_i} \sum_{\times_{i=M+1}^N \mathcal{U}_i} p_M(\mathbf{u}_1) p_L(\mathbf{u}_2) \sum_{\mathcal{Y}^M} \sum_{\mathcal{Y}^L} \left[\sum_{\mathcal{X}^M} \sum_{\mathcal{X}^L} p_M(\mathbf{x}_1|\mathbf{u}_1) p_L(\mathbf{x}_2|\mathbf{u}_2) p(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}_1, \mathbf{x}_2, s_{0,N})^{\frac{1}{1+\rho}}\right]^{1+\rho} \\
&= \sum_{\times_{i=1}^M \mathcal{U}_i} \sum_{\times_{i=M+1}^N \mathcal{U}_i} p_M(\mathbf{u}_1) p_L(\mathbf{u}_2) \sum_{\mathcal{Y}^M} \sum_{\mathcal{Y}^L} \left[\sum_{\mathcal{X}^M} \sum_{\mathcal{X}^L} p_M(\mathbf{x}_1|\mathbf{u}_1) p_L(\mathbf{x}_2|\mathbf{u}_2) \right.\\
&\qquad \left. \times \left[\sum_{s_M \in \mathcal{S}} p(\mathbf{y}_1, s_M|\mathbf{x}_1, \mathbf{x}_2, s_{0,N}) p(\mathbf{y}_2|\mathbf{y}_1, s_M, \mathbf{x}_1, \mathbf{x}_2, s_{0,N})\right]^{\frac{1}{1+\rho}}\right]^{1+\rho} \\
&\stackrel{(b)}{=} \sum_{\times_{i=1}^M \mathcal{U}_i} \sum_{\times_{i=M+1}^N \mathcal{U}_i} p_M(\mathbf{u}_1) p_L(\mathbf{u}_2) \sum_{\mathcal{Y}^M} \sum_{\mathcal{Y}^L} \left[\sum_{\mathcal{X}^M} \sum_{\mathcal{X}^L} p_M(\mathbf{x}_1|\mathbf{u}_1) p_L(\mathbf{x}_2|\mathbf{u}_2) \right.\\
&\qquad \left. \times \left[\sum_{s_M \in \mathcal{S}} p(\mathbf{y}_1, s_M|\mathbf{x}_1, s_{0,N}) p(\mathbf{y}_2|s_M, \mathbf{x}_2)\right]^{\frac{1}{1+\rho}}\right]^{1+\rho} \\
&\stackrel{(c)}{\leq} \sum_{\times_{i=1}^M \mathcal{U}_i} \sum_{\times_{i=M+1}^N \mathcal{U}_i} p_M(\mathbf{u}_1) p_L(\mathbf{u}_2) \\
&\qquad \times \sum_{\mathcal{Y}^M} \sum_{\mathcal{Y}^L} \left[\sum_{\mathcal{X}^M} \sum_{\mathcal{X}^L} \sum_{s_M \in \mathcal{S}} p_M(\mathbf{x}_1|\mathbf{u}_1) p_L(\mathbf{x}_2|\mathbf{u}_2) p(\mathbf{y}_1, s_M|\mathbf{x}_1, s_{0,N})^{\frac{1}{1+\rho}} p(\mathbf{y}_2|s_M, \mathbf{x}_2)^{\frac{1}{1+\rho}}\right]^{1+\rho} \\
&\stackrel{(d)}{\leq} ||\mathcal{S}||^\rho \sum_{s_M \in \mathcal{S}} \sum_{\times_{i=1}^M \mathcal{U}_i} \sum_{\times_{i=M+1}^N \mathcal{U}_i} p_M(\mathbf{u}_1) p_L(\mathbf{u}_2) \sum_{\mathcal{Y}^M} \sum_{\mathcal{Y}^L} \left[\sum_{\mathcal{X}^M} \sum_{\mathcal{X}^L} \right.\\
&\qquad \left. p_M(\mathbf{x}_1|\mathbf{u}_1) p_L(\mathbf{x}_2|\mathbf{u}_2) p(\mathbf{y}_1, s_M|\mathbf{x}_1, s_{0,N})^{\frac{1}{1+\rho}} p(\mathbf{y}_2|s_M, \mathbf{x}_2)^{\frac{1}{1+\rho}}\right]^{1+\rho} \\
&= \sum_{s_M \in \mathcal{S}} \sum_{\times_{i=1}^M \mathcal{U}_i} p_M(\mathbf{u}_1) \sum_{\mathcal{Y}^M} \left(\sum_{\mathcal{X}^M} p_M(\mathbf{x}_1|\mathbf{u}_1) p(\mathbf{y}_1, s_M|\mathbf{x}_1, s_{0,N})^{\frac{1}{1+\rho}}\right)^{1+\rho} 2^{-L\left(E_{L,11}(\rho, p_L(\mathbf{u}_2, \mathbf{x}_2), s_M) - \rho \frac{\log_2 ||\mathcal{S}||}{L}\right)}
\end{aligned}$$

$$\begin{aligned}
&\overset{(e)}{\leq} \sum_{s_M\in\mathcal{S}} \sum_{\times_{i=1}^M \mathcal{U}_i} p_M(\mathbf{u}_1) \sum_{\mathcal{Y}^M} \left(\sum_{\mathcal{X}^M} p_M(\mathbf{x}_1|\mathbf{u}_1) p(\mathbf{y}_1, s_M|\mathbf{x}_1, s_{0,N})^{\frac{1}{1+\rho}} \right)^{1+\rho} 2^{-LF_{L,11}(\rho, p_L(\mathbf{u}_2,\mathbf{x}_2))} \\
&\overset{(f)}{\leq} \sum_{\times_{i=1}^M \mathcal{U}_i} p_M(\mathbf{u}_1) \sum_{\mathcal{Y}^M} \left(\sum_{\mathcal{X}^M} p_M(\mathbf{x}_1|\mathbf{u}_1) \left(\sum_{s_M\in\mathcal{S}} p(\mathbf{y}_1, s_M|\mathbf{x}_1, s_{0,N}) \right)^{\frac{1}{1+\rho}} \right)^{1+\rho} 2^{-LF_{L,11}(\rho, p_L(\mathbf{u}_2,\mathbf{x}_2))} \\
&= \sum_{\times_{i=1}^M \mathcal{U}_i} p_M(\mathbf{u}_1) \sum_{\mathcal{Y}^M} \left(\sum_{\mathcal{X}^M} p_M(\mathbf{x}_1|\mathbf{u}_1) p(\mathbf{y}_1|\mathbf{x}_1, s_{0,N})^{\frac{1}{1+\rho}} \right)^{1+\rho} 2^{-LF_{L,11}(\rho, p_L(\mathbf{u}_2,\mathbf{x}_2))},
\end{aligned} \tag{B.20}$$

(a) is because

$$\begin{aligned}
p_N(\mathbf{x}_1,\mathbf{x}_2,\mathbf{u}_1,\mathbf{u}_2) &= p_M(\mathbf{u}_1,\mathbf{x}_1)p_L(\mathbf{u}_2,\mathbf{x}_2) \\
\Rightarrow p_N(\mathbf{u}_1,\mathbf{u}_2) &= p_M(\mathbf{u}_1)p_L(\mathbf{u}_2) \\
\Rightarrow p_N(\mathbf{x}_1,\mathbf{x}_2|\mathbf{u}_1,\mathbf{u}_2) &= p_M(\mathbf{x}_1|\mathbf{u}_1)p_L(\mathbf{x}_2|\mathbf{u}_2)
\end{aligned}$$

(b) is because $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent and therefore $\mathbf{X}_2 \perp\!\!\!\perp (\mathbf{X}_1, \mathbf{Y}_1, S_M, S_0)$ when $\mathbf{Y}_2$, $\mathbf{Z}_2$, and $\mathbf{S}_2$ are not given, and also because $\mathbf{Y}_2$ is independent of S_0, $\mathbf{X}_1$ and $\mathbf{Y}_1$ when S_M is given. In (c) we used $\left(\sum_i a_i\right)^r \leq \sum_i a_i^r$, $a_i \geq 0$, $0 < r \leq 1$ ([25, Sec. 2.10, Theorem 19]), and in (d) we used $\left(\sum_i P_i a_i\right)^r \leq \sum_i P_i a_i^r$, $a_i \geq 0$, $r \geq 1$ and $\{P_i\}$ is a p.m.f. ([25, Section 2.9, Theorem 16]). For (e) we used the fact that $F_{L,11}(\rho, p_L(\mathbf{u}_2,\mathbf{x}_2))$ is evaluated with the initial state $s_0 \in \mathcal{S}$ that minimizes $E_{L,11}(\rho, p(\mathbf{u},\mathbf{x}), s_0)$ and (f) follows from Minkowski's inequality: $\sum_k \left(\sum_j Q_j a_{jk}^{\frac{1}{r}}\right)^r \leq \left(\sum_j Q_j \left(\sum_k a_{jk}\right)^{\frac{1}{r}}\right)^r$, $r \geq 1$, $a_{jk} \geq 0$, $\{Q_j\}$ is a p.m.f. ([25, Sec. 2.11, Theorem 24]). Plugging this back into (B.18) yields

$$\begin{aligned}
&2^{-NF_{N,11}(\rho, p_N(u^N, x^N))} \\
&\leq \|\mathcal{S}\|^\rho \sum_{\times_{i=1}^M \mathcal{U}_i} p_M(\mathbf{u}_1) \\
&\qquad \times \sum_{\mathcal{Y}^M} \left(\sum_{\mathcal{X}^M} p_M(\mathbf{x}_1|\mathbf{u}_1) p(\mathbf{y}_1|\mathbf{x}_1, s_{0,N})^{\frac{1}{1+\rho}} \right)^{1+\rho} \\
&\qquad\qquad \times 2^{-LF_{L,11}(\rho, p_L(\mathbf{u}_2,\mathbf{x}_2))} \\
&= 2^{-M\left(E_{M,11}(\rho, p_M(\mathbf{u}_1,\mathbf{x}_1), s_{0,N}) - \rho\frac{\log_2\|\mathcal{S}\|}{M}\right)} \\
&\qquad\qquad \times 2^{-LF_{L,11}(\rho, p_L(\mathbf{u}_2,\mathbf{x}_2))} \\
&\leq 2^{-MF_{M,11}(\rho, p_M(\mathbf{u}_1,\mathbf{x}_1))} 2^{-LF_{L,11}(\rho, p_L(\mathbf{u}_2,\mathbf{x}_2))}
\end{aligned}$$

hence

$$\begin{aligned}
&NF_{N,11}(\rho, p_N(u^N, x^N)) \\
&\qquad \geq MF_{M,11}(\rho, p_M(\mathbf{u}_1,\mathbf{x}_1)) + LF_{L,11}(\rho, p_L(u^L, x^L)).
\end{aligned}$$

Therefore

$$\begin{aligned}
&bF_{b,11}\left(\rho, \prod_{k=1}^K q_n(\mathbf{u}_k,\mathbf{x}_k)\right) \\
&\quad \geq nF_{n,11}(\rho, q_n(\mathbf{u},\mathbf{x})) \\
&\qquad\qquad + (K-1)nF_{(K-1)n,11}(\rho, \prod_{k=1}^{K-1} q_n(\mathbf{u}_k,\mathbf{x}_k)) \\
&\quad \geq 2nF_{n,11}(\rho, q_n(\mathbf{u},\mathbf{x})) \\
&\qquad\qquad + (K-2)nF_{(K-2)n,11}(\rho, \prod_{k=1}^{K-2} q_n(\mathbf{u}_k,\mathbf{x}_k)) \\
&\quad \overset{(a)}{\geq} KnF_{n,11}(\rho, q_n(\mathbf{u},\mathbf{x}))
\end{aligned}$$

where (a) follows by repeating the expansion in the first step. ■

Using Lemma B.2 in (B.13) we can write the bound on $\bar{P}_{e11}^{(b)}(m_1|m_2, s_0)$ as

$$\begin{aligned}
&\bar{P}_{e11}^{(Kn)}(m_1|m_2, s_0) \\
&\qquad \leq \|\mathcal{S}\| 2^{-Kn\left(F_{Kn,11}(\rho, \prod_{k=1}^K q_n(\mathbf{u}_k,\mathbf{x}_k)) - \rho R_1(q_n)\right)} \\
&\qquad \leq \|\mathcal{S}\| 2^{-Kn(F_{n,11}(\rho, q_n(\mathbf{u},\mathbf{x})) - \rho R_1(q_n))}.
\end{aligned}$$

From Lemma B.1 we conclude that there exists a positive ρ_{11} such that $F_{n,11}(\rho_{11}, q_n(\mathbf{u},\mathbf{x})) - \rho_{11} R_{1,n}(q_n) = \epsilon_{11}$ for some $\epsilon_{11} > 0$. Hence, the probability $\bar{P}_{e11}^{(Kn)}(m_1|m_2, s_0)$ can be made arbitrarily small by taking K large enough. The same considerations can be repeated for $\bar{P}_{e12}^{(Kn)}(m_2|s_0)$ and $\bar{P}_{e2}^{(Kn)}(m_2|s_0)$. Thus, taking K large enough we have that $\bar{P}_{e}^{(Kn)}(s_0) \leq \bar{P}_{e11}^{(Kn)}(m_1|m_2, s_0) + \bar{P}_{e12}^{(Kn)}(m_2|s_0) + \bar{P}_{e2}^{(Kn)}(m_2|s_0)$ can be made arbitrarily small, hence $(R_{1,n}(q_n), R_{2,n}(q_n))$ is achievable for every q_n: finding the minimal K that results in $\bar{P}_{e}^{(Kn)}(s_0) \leq \epsilon, \forall s_0 \in \mathcal{S}$ allows us to generate codes for any blocklength $b > Kn$ with $\bar{P}_{e}^{(b)}(s_0) \leq \epsilon, \forall s_0 \in \mathcal{S}$. This is done by designing codes for $(R_{1,n}(q_n) - \frac{\delta}{2}, R_{2,n}(q_n) - \frac{\delta}{2})$, taking K large enough such that $\frac{\max\{R_{1,n}(q_n), R_{2,n}(q_n)\}}{K} < \frac{\delta}{2}$, and adding zeros when the blocklength is not an integer multiple of K.

We omit the details of the derivation that shows

$$F_{Kn,2}\left(\rho, \prod_{k=1}^K q_n(\mathbf{u}_k,\mathbf{x}_k)\right) \geq F_{n,2}(\rho, q_n(\mathbf{u},\mathbf{x})) \tag{B.21}$$

except the critical step of introducing s_M into the channel that connects U^n and Z^n, which is characterized by $p(\mathbf{z}|\mathbf{u}, s_0)$. This p.m.f. is used in the expressions for $F_{Kn,2}$ and $F_{n,2}$

$$\begin{aligned}
&p(\mathbf{z}|\mathbf{u}, s_0) \\
&= p(\mathbf{z}_1,\mathbf{z}_2|\mathbf{u}_1,\mathbf{u}_2, s_0) \\
&= \sum_{s_M} p(\mathbf{z}_1,\mathbf{z}_2, s_M|\mathbf{u}_1,\mathbf{u}_2, s_0) \\
&= \sum_{\mathcal{X}^n,\mathcal{Y}^n,\mathcal{S}^n} p(\mathbf{z}_1,\mathbf{z}_2,\mathbf{y}_1,\mathbf{y}_2,\mathbf{s}_1,\mathbf{s}_2|\mathbf{x}_1,\mathbf{x}_2,\mathbf{u}_1,\mathbf{u}_2, s_0) \\
&\qquad\qquad \times p(\mathbf{x}_1,\mathbf{x}_2|\mathbf{u}_1,\mathbf{u}_2, s_0) \\
&= \sum_{\mathcal{X}^n,\mathcal{Y}^n,\mathcal{S}^n} p_M(\mathbf{x}_1|\mathbf{u}_1)p_L(\mathbf{x}_2|\mathbf{u}_2)p(\mathbf{z}_1,\mathbf{y}_1,\mathbf{s}_1|\mathbf{x}_1,\mathbf{x}_2, s_0) \\
&\qquad\qquad \times p(\mathbf{z}_2,\mathbf{y}_2,\mathbf{s}_2|\mathbf{z}_1,\mathbf{y}_1,\mathbf{s}_1,\mathbf{x}_1,\mathbf{x}_2, s_0) \\
&= \sum_{\mathcal{X}^n,\mathcal{Y}^M,\mathcal{S}^M} p(\mathbf{z}_1,\mathbf{y}_1,\mathbf{s}_1|\mathbf{x}_1,\mathbf{x}_2, s_0)p(\mathbf{z}_2|s_M,\mathbf{x}_2,\mathbf{u}_2) \\
&\qquad\qquad \times p_M(\mathbf{x}_1|\mathbf{u}_1)p_L(\mathbf{x}_2|\mathbf{u}_2, s_M)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\mathcal{X}^n,\mathcal{Y}^M,\mathcal{S}^M} p(\mathbf{z}_1,\mathbf{y}_1,\mathbf{s}_1|\mathbf{x}_1,s_0)p(\mathbf{z}_2,\mathbf{x}_2|s_M,\mathbf{u}_2) \\
&\qquad\qquad \times p_M(\mathbf{x}_1|\mathbf{u}_1) \\
&= \sum_{\mathcal{X}^M,\mathcal{Y}^M,\mathcal{S}^M} p(\mathbf{z}_1,\mathbf{y}_1,\mathbf{s}_1,\mathbf{x}_1|\mathbf{u}_1,s_0)p(\mathbf{z}_2|s_M,\mathbf{u}_2) \\
&= \sum_{s_M\in\mathcal{S}} p(\mathbf{z}_1,s_M|\mathbf{u}_1,s_0)p(\mathbf{z}_2|s_M,\mathbf{u}_2). \qquad \text{(B.22)}
\end{aligned}
$$

The expansion (B.22) is the key showing that (B.21) holds.

F. The Boundary of the Achievable Region

The achievable region for a given n is given by

$$
\begin{aligned}
\underline{\mathcal{R}}_n &= co \bigcup_{q_n\in\mathcal{Q}_n} \underline{\mathcal{R}}_n(q_n), \\
\underline{\mathcal{R}}_n(q_n) &\triangleq \Big\{(R_1,R_2): 0\le R_1\le R_{1,n}(q_n) \\
&\qquad\qquad 0\le R_2\le R_{2,n}(q_n)\Big\}, \\
R_{1,n}(q_n) &\triangleq \min_{s_0'\in\mathcal{S}} \frac{1}{n} I(X^n;Y^n|U^n,s_0')_{q_n} - \frac{\log_2\|\mathcal{S}\|}{n} \\
R_{2,n}(q_n) &\triangleq \min_{s_0''\in\mathcal{S}} \frac{1}{n} I(U^n;Z^n|s_0'')_{q_n} - \frac{\log_2\|\mathcal{S}\|}{n}
\end{aligned}
$$

where $\mathcal{Q}_n$ is defined in Section III-A to be the set of all joint distributions $p_{U^n,X^n}(u^n,x^n)$ such that the cardinality of U^n satisfies (14). To study the asymptotic properties of the achievable region, we first characterize its boundary. We next show that for a fixed blocklength n, the boundary of the region $\underline{\mathcal{R}}_n$ is given by

$$R_2(R_1) = \inf_{0\le\lambda\le 1} \left\{C^n(\lambda) - \lambda R_1\right\} - (1+\lambda)\frac{\log_2|\mathcal{S}|}{n}.$$

Since the achievable region is completely characterized by $C^n(\lambda)$, this characterization simplifies the proof of convergence of the achievable region: instead of considering the entire region, it is enough to study only the properties of the scalar quantity $C^n(\lambda)$. Convergence of $C^n(\lambda)$ implies convergence of the region. Additionally this characterization helps in bounding the cardinality of the auxiliary random vector U^n.

Fix $\lambda\ge 0$ and $p(u^n,x^n)$ and consider the line

$$\tilde{R}_2^\lambda(R_1) = R_{2,n}(p) + \lambda R_{1,n}(p) - \lambda R_1.$$

This line is either tangent, upper bounds or intersects with the region of positive error exponents. Hence, for a fixed $\lambda\ge 0$, the line

$$R_2^\lambda(R_1) = \max_{p(u^n,x^n)} \left\{R_{2,n}(p) + \lambda R_{1,n}(p)\right\} - \lambda R_1$$

upper bounds the region of positive error exponents, otherwise for the same λ there exists $\tilde{p}(u^n,x^n)$ for which $R_{2,n}(\tilde{p}) + \lambda R_{1,n}(\tilde{p}) > \max_{p(u^n,x^n)}\left\{R_{2,n}(p)+\lambda R_{1,n}(p)\right\}$, thus contradicting the maximization. Following [16] and [26, Lemma 3], we can write the boundary of $\underline{\mathcal{R}}_n$ as the least upper bound

$$
\begin{aligned}
R_2(R_1) = \inf_{\lambda\ge 0}\Bigg[\max_{p(u^n,x^n)}\Bigg\{ &\min_{s_0''\in\mathcal{S}} \frac{1}{n} I(U^n;Z^n|s_0'')_p - \frac{\log_2\|\mathcal{S}\|}{n} \\
&+ \lambda\left(\min_{s_0'\in\mathcal{S}} \frac{1}{n} I(X^n;Y^n|U^n,s_0')_p - \frac{\log_2\|\mathcal{S}\|}{n}\right)\Bigg\} \\
&- \lambda R_1\Bigg]. \qquad \text{(B.23)}
\end{aligned}
$$

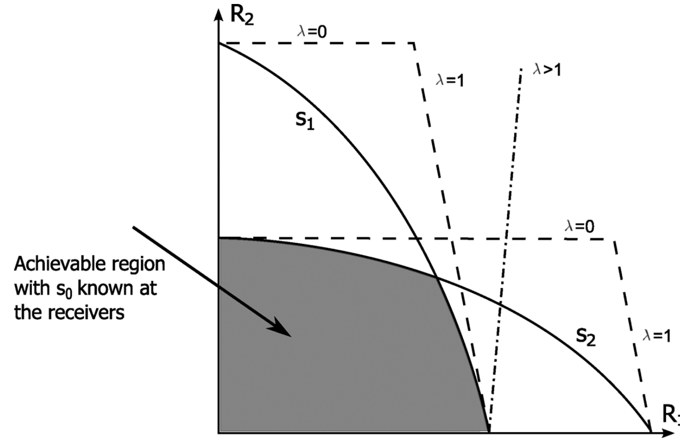

Fig. 5. Lines bounding the achievable region for the FSBC, and the resulting outer bound on the achievable region. Note that the line with $\lambda > 1$ constitutes an outer bound looser than the line with $\lambda = 1$.

This situation is illustrated in Fig. 5.

We now show that in order to obtain a complete description of the region $\underline{\mathcal{R}}_n$ via (B.23) it is enough to consider only $0\le\lambda\le 1$. This is stated in the following Lemma.

Lemma B.3: A complete description of the region $\underline{\mathcal{R}}_n$ via (B.23) is achieved by restricting the range of λ to $0\le\lambda\le 1$.

Proof: Note that when $\lambda = 0$, $R_2^0(R_1) = \max_{p(u^n,x^n)}\min_{s_0''\in\mathcal{S}}\frac{1}{n}I(U^n;Z^n|s_0'') - \frac{\log_2\|\mathcal{S}\|}{n}$ is the resulting upper bound, irrespective of R_1. Next, consider $\lambda = 1$ and denote $C_0(1) \triangleq \max_{p(x^n)}\min_{s_0\in\mathcal{S}}\frac{1}{n}I(X^n;Y^n|s_0) - \frac{\log_2\|\mathcal{S}\|}{n}$. Then

$$
\begin{aligned}
&R_2^1(R_1) \\
&= \max_{p(u^n,x^n)}\Bigg\{\min_{s_0''\in\mathcal{S}}\frac{1}{n}I(U^n;Z^n|s_0'')_p \\
&\qquad + \min_{s_0'\in\mathcal{S}}\frac{1}{n}I(X^n;Y^n|U^n,s_0')_p - 2\frac{\log_2\|\mathcal{S}\|}{n}\Bigg\} - R_1 \\
&\le \max_{p(u^n,x^n)}\Bigg\{\min_{s_0'\in\mathcal{S}}\Bigg[\frac{1}{n}I(U^n;Z^n|s_0')_p \\
&\qquad + \frac{1}{n}I(X^n;Y^n|U^n,s_0')_p\Bigg] - 2\frac{\log_2\|\mathcal{S}\|}{n}\Bigg\} - R_1 \\
&\le \max_{p(u^n,x^n)}\Bigg\{\min_{s_0'\in\mathcal{S}}\Bigg[\frac{1}{n}I(U^n;Y^n|s_0')_p \\
&\qquad + \frac{1}{n}I(X^n;Y^n|U^n,s_0')_p\Bigg] - 2\frac{\log_2\|\mathcal{S}\|}{n}\Bigg\} - R_1 \\
&= \max_{p(u^n,x^n)}\min_{s_0'\in\mathcal{S}}\frac{1}{n}I(X^n;Y^n|s_0')_{\tilde{p}} - 2\frac{\log_2\|\mathcal{S}\|}{n} - R_1 \\
&= (C_0(1) - R_1) - \frac{\log_2\|\mathcal{S}\|}{n}
\end{aligned}
$$

irrespective of $p(u^n)$. Hence, the distribution that achieves the bound is atomic in U^n. If $\lambda > 1$, then again, denoting with $(\tilde{p},\tilde{s}_0'',\tilde{s}_0')$ the triplet that achieves the max-min solution we obtain

$$R_2^\lambda(R_1) = \max_{p(u^n,x^n)}\Bigg\{\min_{s_0''\in\mathcal{S}}\Bigg[\lambda\frac{1}{n}I(U^n;Z^n|s_0'')_p$$

$$
\begin{aligned}
&\qquad\qquad - (\lambda-1)\frac{1}{n}I(U^n;Z^n|s_0'')_p\Big]\\
&\qquad\qquad + \lambda \min_{s_0''\in\mathcal{S}} \frac{1}{n}I(X^n;Y^n|U^n,s_0')_p\Big\}\\
&\qquad\qquad - (1+\lambda)\frac{\log_2\|\mathcal{S}\|}{n} - \lambda R_1\\
&\le \max_{p(u^n,x^n)}\Big\{\lambda \min_{s_0''\in\mathcal{S}}\frac{1}{n}I(U^n;Z^n|s_0'')_p\\
&\qquad\qquad + \lambda \min_{s_0''\in\mathcal{S}} \frac{1}{n}I(X^n;Y^n|U^n,s_0')_p\Big\}\\
&\qquad\qquad - (1+\lambda)\frac{\log_2\|\mathcal{S}\|}{n} - \lambda R_1\\
&\le \max_{p(u^n,x^n)}\Big\{\lambda \min_{s_0'\in\mathcal{S}}\Big[\frac{1}{n}I(U^n;Z^n|s_0')_p\\
&\qquad\qquad + \frac{1}{n}I(X^n;Y^n|U^n,s_0')_p\Big]\Big\}\\
&\qquad\qquad - (1+\lambda)\frac{\log_2\|\mathcal{S}\|}{n} - \lambda R_1\\
&\le \lambda \max_{p(u^n,x^n)} \min_{s_0'\in\mathcal{S}} \frac{1}{n}I(X^n;Y^n|s_0')_p\\
&\qquad\qquad - (1+\lambda)\frac{\log_2\|\mathcal{S}\|}{n} - \lambda R_1\\
&= \lambda(C_0(1)-R_1) - \frac{\log_2\|\mathcal{S}\|}{n} \qquad \text{(B.24)}
\end{aligned}
$$

$R_2^\lambda(R_1) \le \lambda(C_0(1)-R_1) - \frac{\log_2\|\mathcal{S}\|}{n}$. Clearly when $\lambda > 1$, $R_2^1(C_0(1)) = R_2^\lambda(C_0(1)) = -\frac{\log_2\|\mathcal{S}\|}{n}$. However, when R_1 decreases, then the lines for $\lambda > 1$ pass to the right of $R_2^1(R_1)$, $R_1 < C_0(1)$, see Fig. 5, and thus they are not tight outer bounds. Taking the infimum over $\lambda \ge 0$ in (B.23) we conclude that it is enough to consider only $0 \le \lambda \le 1$, to obtain a complete characterization of the region. ■

The fact that λ has a finite range will be used in the proof of Lemma B.4 to show that the achievable region converges as $n \to \infty$. Define next

$$
\begin{aligned}
C^n(\lambda) \triangleq \max_{p(u^n,x^n)}\Big\{&\min_{s_0\in\mathcal{S}}\frac{1}{n}I(U^n;Z^n|s_0)\\
&+\lambda\min_{s_0'\in\mathcal{S}}\frac{1}{n}I(X^n;Y^n|U^n,s_0')\Big\}. \qquad \text{(B.25)}
\end{aligned}
$$

We have the following lemma.

Lemma B.4: $C^n(\lambda)$ defined in (B.25) converges as $n \to \infty$ to a finite limit given by

$$
\begin{aligned}
C^\infty(\lambda) &\triangleq \lim_{n\to\infty} C^n(\lambda)\\
&= \sup_n\left[C^n(\lambda) - (1+\lambda)\frac{\log_2\|\mathcal{S}\|}{n}\right]. \qquad \text{(B.26)}
\end{aligned}
$$

Furthermore, for a given n the achievable region is completely characterized by $C^n(\lambda)-(1+\lambda)\frac{\log_2\|\mathcal{S}\|}{n}$, and $C^\infty(\lambda)$ provides the largest achievable region.

Proof: The convergence to the limit is shown in Appendix III. The fact that $C^n(\lambda) - (1+\lambda)\frac{\log_2\|\mathcal{S}\|}{n}$ gives a complete characterization of the achievable region follows from [26, Corollary on p. 7]. The fact that $C^\infty(\lambda)$ provides the largest achievable region follows from its sup-additivity, also shown in Appendix III. ■

Lemma B.4 implies that the boundary of the largest achievable region, $R_2(R_1)$, can be written as

$$R_2(R_1) = \inf_{0\le\lambda\le1}\{C^\infty(\lambda) - \lambda R_1\}. \qquad \text{(B.27)}$$

We conclude that $C^\infty(\lambda)$ completely characterizes this region. Hence, when transmitting at the *positive* rate pair $(R_1-\delta, R_2(R_1)-\delta)$, $\delta > 0$, and given an arbitrary $\epsilon > 0$, then there exists some $n(\epsilon,\delta)$ such that for all $n \ge n(\epsilon,\delta)$, an $(R_1-\delta, R_2(R_1)-\delta, n)$ code with an $\max_{s_0\in\mathcal{S}} P_e^{(n)}(s_0) \le \epsilon$.

G. Cardinality Bounds

Finally, we discuss the cardinality of the auxiliary RV U^n, as the feasibility of the optimization problem for maximizing the rate pairs depends on the existence of such bounds. These bounds are summarized in the following pair of lemmas.

Lemma B.5: $C^n(\lambda)$ can be completely characterized by a RV U^n whose cardinality is upper bounded by $\|\times_{i=1}^n \mathcal{U}_i\| \le \|\mathcal{X}\|^n + 2\|\mathcal{S}\| + 1$.

Proof: Fix n and let $\mathbf{p}(u^n) \triangleq \Big\{p\big(X^n = (a_1,a_2,\ldots,a_n)\big|u^n\big)\Big\}_{(a_1,a_2,\ldots,a_n)\in\mathcal{X}^n}$ be a probability distribution on $\mathcal{X}^n$, indexed by u^n. We index with a variable U^n all the probability distributions on $\mathcal{X}^n$. Let $\mathcal{P}$ be the collection of all distributions $\{\mathbf{p}(u^n)\}_{u^n\in\times_{i=1}^n\mathcal{U}_i}$. The set $\mathcal{P}$ is compact [35, proof of Lemma 3][8]. We now define a set of functions on $\mathbf{p}(u^n)$ and $\mathcal{X}^n$

$$g(a^n;\mathbf{p}(u^n)) \triangleq p(X^n = a^n|U^n = u^n). \qquad \text{(B.28)}$$

From (2) and the superposition structure we can write the joint distribution for a block of n transmissions for some fixed u^n and s_0 as

$$
\begin{aligned}
&p(z^n,y^n,x^n|u^n,s_0)\\
&\quad= \sum_{\mathcal{S}^n} p(x^n|u^n)\prod_{i=1}^n p(y_i,z_i,s_i|x_i,s_{i-1})\\
&\quad\forall (z^n,y^n,x^n)\in\mathcal{Z}^n\times\mathcal{Y}^n\times\mathcal{X}^n. \qquad \text{(B.29)}
\end{aligned}
$$

We also define for every $s_0 \in \mathcal{S}$ the pair of functions

$$g_1(\mathbf{p},s_0) \triangleq I(X^n;Y^n|\mathbf{p},s_0) \qquad \text{(B.30a)}$$
$$g_2(\mathbf{p},s_0) \triangleq H(Z^n|\mathbf{p},s_0). \qquad \text{(B.30b)}$$

These functions are computable from the distribution $p(z^n,y^n,x^n|u^n,s_0)$ on $\mathcal{Z}^n\times\mathcal{Y}^n\times\mathcal{X}^n$ given in (B.29). As in [35, proof of Theorem 2] we let $f(u^n)$ be a Borel measure on $\times_{i=1}^n\mathcal{U}_i$. We denote the measure $f(u^n)$ on $\times_{i=1}^n\mathcal{U}_i$ that

[8] $\mathcal{P}$ is compact by the Krein–Milman Theorem [36] since every element in $\mathcal{P}$ can be written as a convex combination of the extreme points of $\mathcal{P}$. To understand the meaning of an extreme point of $\mathcal{P}$ consider as an example the case of $n=2$ and $\mathcal{X} = \{A,B\}$. Each element of $\mathcal{P}$ is a vector $\mathbf{p}$ of length 4 whose elements add to 1. The set $\mathcal{P}$ of probability distributions on $\mathcal{X}^2$ is characterized by vectors of the form

$$
\begin{aligned}
\mathbf{p} = \Big\{&p(X^2=(A,A)), p(X^2=(A,B)),\\
&\qquad p(X^2=(B,A)), p(X^2=(B,B))\Big\}\\
\equiv \Big\{&p_1(u^2), p_2(u^2), p_3(u^2), p_4(u^2)\Big\}.
\end{aligned}
$$

The extreme points of $\mathcal{P}$ are

$$\Big\{(1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1)\Big\}.$$

achieves the optimal solution for $C^n(\lambda)$ by $f^*(u^n)$. This measure induces a measure on $\mathcal{P}$ which we denote with $\mu^*(d\mathbf{p})$. Therefore, we can write the optimal joint distribution as

$$\begin{aligned} p^*(y^n, z^n, s^n, x^n, u^n|s_0) &= p^*(u^n)p(x^n|u^n)p(y^n, z^n, s^n|x^n, s_0) \\ &= \mu^*(d\mathbf{p})p(x^n;\mathbf{p})p(y^n, z^n, s^n|x^n, s_0) \end{aligned}$$

$p(x^n;\mathbf{p})$ is the probability that $X^n = x^n$ when the p.m.f. is given by the vector $\mathbf{p}$ (this is equivalent to $p(x^n|u^n)$ and $g(x^n;\mathbf{p}(u^n))$). This is simply element in $\mathbf{p}$ which corresponds to x^n. Now it follows that

$$\begin{aligned} p^*(y^n, z^n, x^n|s_0) &= \int_{\mathcal{P}} \sum_{\mathcal{S}^n} \mu^*(d\mathbf{p})p(x^n;\mathbf{p})p(y^n, z^n, s^n|x^n, s_0) \\ &= \left(\int_{\mathcal{P}} \mu^*(d\mathbf{p})p(x^n;\mathbf{p})\right) \\ &\quad \times \sum_{\mathcal{S}^n} \prod_{i=1}^n p(y_i, z_i, s_i|x_i, s_{i-1}) \end{aligned} \tag{B.31a}$$

$$p^*(z^n|s_0) = \sum_{\mathcal{X}^n, \mathcal{Y}^n} p^*(y^n, z^n, x^n|s_0). \tag{B.31b}$$

Therefore, from $\left\{\int_{\mathcal{P}} \mu^*(d\mathbf{p})p(x^n;\mathbf{p})\right\}_{x^n\in\mathcal{X}^n}$ we can obtain $\left\{p^*(z^n|s_0)\right\}_{\mathcal{Z}^n\times\mathcal{S}}$, from which we can obtain $\left\{H(Z^n|s_0)\right\}_{s_0\in\mathcal{S}}$. Using the set of functions defined in (B.28) and (B.30) we have

$$\begin{aligned} \int_{\mathcal{P}} \mu^*(d\mathbf{p})\, g(x^n;\mathbf{p}) &= \int_{\mathcal{P}} \mu^*(d\mathbf{p})p(x^n;\mathbf{p}) \\ &= p^*(X^n = x^n) \quad \forall x^n \in \mathcal{X}^n \end{aligned} \tag{B.32a}$$

$$\begin{aligned} \int_{\mathcal{P}} \mu^*(d\mathbf{p})\, g_1(\mathbf{p}, s_0) &= \int_{\mathcal{P}} \mu^*(d\mathbf{p})I(X^n;Y^n|\mathbf{p}, s_0) \\ &= I(X^n;Y^n|U^n, s_0),\ \forall s_0 \in \mathcal{S} \end{aligned} \tag{B.32b}$$

$$\begin{aligned} \int_{\mathcal{P}} \mu^*(d\mathbf{p})\, g_2(\mathbf{p}, s_0) &= \int_{\mathcal{P}} \mu^*(d\mathbf{p})H(Z^n|\mathbf{p}, s_0) \\ &= H(Z^n|U^n, s_0), \quad \forall s_0 \in \mathcal{S}. \end{aligned} \tag{B.32c}$$

Using (B.32) in (B.31), we can evaluate $H(Z^n|s_0)$, from which is possible to evaluate $C^n(\lambda)$.

Now we invoke the [35, Lemma 3][9] and conclude that there exists a discrete random variable U^n whose cardinality $\| \times_{i=1}^n \mathcal{U}_i\| \leq \|\mathcal{X}\|^n + 2\|\mathcal{S}\| + 1$ that achieves the maximum of $C^n(\lambda)$. This follows from noting that there are $\|\mathcal{X}\|^n$ equations in (B.32a), $\|\mathcal{S}\|$ equations in (B.32b) and $\|\mathcal{S}\|$ equations in (B.32c).

[9] Statement of [35, Lemma 3]: Let $\mathcal{P}_n$ be the set of all probability vectors of length n, $\mathbf{p} = (p_1, p_2, \ldots, p_n)$, 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

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

Lemma B.6: $C^n(\lambda)$ can be completely characterized by a RV U^n whose cardinality is upper bounded by $\min\left\{\|\mathcal{Y}\|^n, \|\mathcal{Z}\|^n\right\} \cdot \|\mathcal{S}\| + 2\|\mathcal{S}\| + 1$.

Proof: To show the bound $\|\mathcal{Z}\|^n \cdot \|\mathcal{S}\| + 2\|\mathcal{S}\| + 1$ we recall that in order to obtain a complete characterization of $C^n(\lambda)$ then we need the $2\|\mathcal{S}\|$ expressions for $\left\{g_1(\mathbf{p}, s_0)\right\}_{s_0\in\mathcal{S}}$ and $\left\{g_2(\mathbf{p}, s_0)\right\}_{s_0\in\mathcal{S}}$ defined in (B.30) as well as $\left\{H(Z^n|s_0)\right\}_{s_0\in\mathcal{S}}$. Alternatively to Lemma B.5 we can obtain $\left\{p^*(z^n|s_0)\right\}_{\mathcal{Z}^n\times\mathcal{S}}$ directly using $\|\mathcal{Z}\|^n \cdot \|\mathcal{S}\|$ expressions of the form

$$\begin{aligned} \tilde{g}(z^n;\mathbf{p}) &= p(Z^n = z^n|\mathbf{p}, s_0) \\ &= \sum_{\mathcal{X}^n, \mathcal{Y}^n, \mathcal{S}^n} p(x^n;\mathbf{p}) \prod_{i=1}^n p(y_i, z_i, s_i|x_i, s_{i-1}), \\ &\qquad \forall z^n \in \mathcal{Z}^n, s_0 \in \mathcal{S}. \end{aligned}$$

When using the optimal measure $\mu^*(\mathbf{p})$ then

$$\begin{aligned} \int_{\mathcal{P}} \mu^*(d\mathbf{p})\tilde{g}(z^n;\mathbf{p}) &= \int_{\mathcal{P}} \mu^*(d\mathbf{p}) \sum_{\mathcal{X}^n, \mathcal{Y}^n, \mathcal{S}^n} p(x^n;\mathbf{p}) \\ &\quad \times \prod_{i=1}^n p(y_i, z_i, s_i|x_i, s_{i-1}) \\ &= p^*(z^n|s_0). \end{aligned}$$

Again, from [35, Lemma 3] it follows that the cardinality of U^n can be bounded by $\|\mathcal{Z}\|^n \cdot \|\mathcal{S}\| + 2\|\mathcal{S}\| + 1$.

To show the second bound recall that the channel is physically degraded. Thus $p(z^n|x^n, y^n, s_0) = p(z^n|y^n, s_0) = \prod_{i=1}^n p(z_i|y^i, z^{i-1}, s_0)$, when the terms in the product are defined in (6b). Therefore, $p(z^n|s_0)$ can be obtained from $p(y^n|s_0)$ via $p(z^n|s_0) = \sum_{\mathcal{Y}^n} p(y^n|s_0) \prod_{i=1}^n p(z_i|y^i, z^{i-1}, s_0)$. Repeating arguments similar to the above, we conclude that we can bound the cardinality of U^n by $\|\mathcal{Y}\|^n \cdot \|\mathcal{S}\| + 2\|\mathcal{S}\| + 1$, completing the proof of the lemma. ∎

Combining Lemma B.5 and Lemma B.6, we conclude that the cardinality of U^n can be bounded by

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

Appendix C
Proof of Lemma B.4 (Convergence of $\lim_{n\to\infty} C^n(\lambda)$)

In this section we prove Lemma B.4. This lemma states that

$$\begin{aligned} &\lim_{n\to\infty} \max_{p(u^n, x^n)} \left\{ \min_{s_0''\in\mathcal{S}} \frac{1}{n} I(U^n;Z^n|s_0'') + \lambda \min_{s_0'\in\mathcal{S}} \frac{1}{n} I(X^n;Y^n|U^n, s_0') - (1+\lambda)\frac{\log_2\|\mathcal{S}\|}{n} \right\} \\ &= \lim_{n\to\infty} \max_{p(u^n, x^n)} \left\{ \min_{s_0''\in\mathcal{S}} \frac{1}{n} I(U^n;Z^n|s_0'') + \lambda \min_{s_0'\in\mathcal{S}} \frac{1}{n} I(X^n;Y^n|U^n, s_0') \right\} \\ &\triangleq \lim_{n\to\infty} C^n(\lambda) \end{aligned}$$

exists and is finite for the physically degraded FSBC.

Details of the Proof: Recall the definition of $F_n(\lambda)$ in (17):

$$F_n(\lambda) = C^n(\lambda) - (1+\lambda)\frac{\log_2\|\mathcal{S}\|}{n}. \tag{C.1}$$

Note that

$$\lim_{n\to\infty} F_n(\lambda) = \lim_{n\to\infty} C^n(\lambda)$$

if the limit exists. We show that the limit exists by demonstrating that $F_n(\lambda)$ is sup-additive. Let $s_0'' = s_0^z(l)$ minimize $\frac{1}{l}I(U^l;Z^l|s_0'')_{q_1}$ and $s_0' = s_0^y(l)$ minimize $\frac{1}{l}I(X^l;Y^l|U^l,s_0')_{q_1}$ for the triplet $(q_1(u^l,x^l), s_0^z(l), s_0^y(l))$ that achieves the max-min solution for $F_l(\lambda)$, and let $(q_2(u^m,x^m), s_0^z(m), s_0^y(m))$ achieve the max-min solution for $F_m(\lambda)$. To show that $F_n(\lambda)$ is sup-additive, we show that for $n=m+l$ it satisfies

$$nF_n(\lambda) \ge lF_l(\lambda) + mF_m(\lambda).$$

Let $q(u^n,x^n) = q_1(u^l,x^l)q_2(u_{l+1}^n,x_{l+1}^n)$. Also let $s_0^z(n)$ and $s_0^y(n)$ be the states that minimize $F_n(\lambda)$ subject to the input distribution $q(u^n,x^n)$. Then, by definition

$$\begin{aligned}
nF_n(\lambda) &+ (1+\lambda)\log_2\|\mathcal{S}\| \\
&\ge \Big[I(U^n;Z^n|s_0^z(n)) + \lambda I(X^n;Y^n|U^n,s_0^y(n))\Big]_{q_1q_2} \\
&\ge I(U_1^l;Z_1^l|s_0^z(n))_{q_1} + I(U_1^l,U_{l+1}^n;Z_{l+1}^n|Z_1^l,s_0^z(n))_{q_1q_2} \\
&\qquad + \lambda\Bigg(I(X_1^l;Y_1^l|U_1^l,U_{l+1}^n,s_0^y(n))_{q_1q_2} \\
&\qquad + I(X_1^l,X_{l+1}^n;Y_{l+1}^n|U_1^l,U_{l+1}^n,Y_1^l,s_0^y(n))_{q_1q_2}\Bigg).
\end{aligned}$$

By construction, (U^n,X^n) is independent of the initial state. Thus

$$\begin{aligned}
q(u_1^l,x_1^l,u_{l+1}^n,x_{l+1}^n|s_0^y(n)) &= q_1(u_1^l,x_1^l)q_2(u_{l+1}^n,x_{l+1}^n) \\
\Rightarrow q(u_1^l,x_1^l,u_{l+1}^n) &= q_1(u_1^l,x_1^l)q_2(u_{l+1}^n) \\
\Rightarrow q(x_1^l|u_1^l,u_{l+1}^n) &= q_1(x_1^l|u_1^l) \\
\Rightarrow q(x_1^l|u_1^l,u_{l+1}^n,s_0^y(n)) &= q_1(x_1^l|u_1^l,s_0^y(n))
\end{aligned}$$

hence

$$\begin{aligned}
I(X_1^l;&Y_1^l|U_1^l,U_{l+1}^n,s_0^y(n))_{q_1q_2} \\
&= H(X_1^l|U_1^l,s_0^y(n)) - H(X_1^l|Y_1^l,U_1^l,U_{l+1}^n,s_0^y(n)) \\
&\ge H(X_1^l|U_1^l,s_0^y(n)) - H(X_1^l|Y_1^l,U_1^l,s_0^y(n)) \\
&= I(X_1^l;Y_1^l|U_1^l,s_0^y(n))_{q_1} \\
&\ge I(X_1^l;Y_1^l|U_1^l,s_0^y(l))_{q_1}.
\end{aligned}$$

In conclusion

$$\begin{aligned}
nF_n(\lambda) &\ge I(U_1^l;Z_1^l|s_0^z(l))_{q_1} \\
&\qquad + I(U_1^l,U_{l+1}^n;Z_{l+1}^n|Z_1^l,s_0^z(n))_{q_1q_2} \\
&\qquad + \lambda\Bigg(I(X_1^l;Y_1^l|U_1^l,s_0^y(l))_{q_1} \\
&\qquad + I(X_1^l,X_{l+1}^n;Y_{l+1}^n|U_1^l,U_{l+1}^n,Y_1^l,s_0^y(n))_{q_1q_2}\Bigg) \\
&\qquad - (1+\lambda)\log_2\|\mathcal{S}\| \\
&\ge lF_l(\lambda) + I(U_{l+1}^n;Z_{l+1}^n|Z_1^l,s_0^z(n))_{q_1q_2} \\
&\qquad + \lambda I(X_{l+1}^n;Y_{l+1}^n|U_1^l,U_{l+1}^n,Y_1^l,s_0^y(n))_{q_1q_2}.
\end{aligned}$$

Now, note that

$$\begin{aligned}
p(u_{l+1}^n,z_1^l|s_l,s_0^z(n)) &= q_2(u_{l+1}^n)p(z_1^l|s_l,s_0^z(n)) \\
\Rightarrow p(u_{l+1}^n|z_1^l,s_l,s_0^z(n)) &= q_2(u_{l+1}^n).
\end{aligned}$$

Hence

$$\begin{aligned}
I(U_{l+1}^n;&Z_{l+1}^n|Z_1^l,s_0^z(n))_{q_1q_2} + \log_2\|\mathcal{S}\| \\
&\overset{(a)}{\ge} I(U_{l+1}^n;Z_{l+1}^n|S_l,Z_1^l,s_0^z(n))_{q_1q_2} \\
&= H(U_{l+1}^n|Z_1^l,S_l,s_0^z(n))_{q_1q_2} \\
&\qquad - H(U_{l+1}^n|Z_1^l,Z_{l+1}^n,S_l,s_0^z(n))_{q_1q_2} \\
&= H(U_{l+1}^n)_{q_2} - H(U_{l+1}^n|Z_1^l,Z_{l+1}^n,S_l,s_0^z(n))_{q_1q_2} \\
&\ge H(U_{l+1}^n|S_l,s_0^z(n))_{q_1q_2} - H(U_{l+1}^n|Z_{l+1}^n,S_l,s_0^z(n))_{q_1q_2} \\
&= I(U_{l+1}^n;Z_{l+1}^n|S_l,s_0^z(n))_{q_1q_2} \\
&\overset{(b)}{\ge} I(U_{l+1}^n;Z_{l+1}^n|s_l=s_0^z(m))_{q_1q_2} \\
&= I(U_{l+1}^n;Z_{l+1}^n|s_l=s_0^z(m))_{q_2}
\end{aligned}$$

where (a) follows from [5, Lemma 1 in Appendix 4A][10] and (b) is because $s_0^z(m)$ is the minimizing state. We also have

$$\begin{aligned}
p(x_{l+1}^n,&u_1^l,u_{l+1}^n,y_1^l|s_l,s_0^y(n)) \\
&= q_2(x_{l+1}^n,u_{l+1}^n|s_l,s_0^y(n))p(y_1^l,u_1^l|s_l,s_0^y(n)) \\
&= q_2(x_{l+1}^n|u_{l+1}^n,s_l,s_0^y(n)) \\
&\qquad \times q_2(u_{l+1}^n|s_l,s_0^y(n))p(y_1^l,u_1^l|s_l,s_0^y(n)) \\
&= q_2(x_{l+1}^n|u_{l+1}^n,s_l,s_0^y(n))p(y_1^l,u_1^l,u_{l+1}^n|s_l,s_0^y(n)),
\end{aligned}$$

hence

$$p(x_{l+1}^n|u_1^l,u_{l+1}^n,y_1^l,s_l,s_0^y(n)) = q_2(x_{l+1}^n|u_{l+1}^n,s_l,s_0^y(n)).$$

Therefore

$$\begin{aligned}
I(X_{l+1}^n;&Y_{l+1}^n|U_1^l,U_{l+1}^n,Y_1^l,s_0^y(n))_{q_1q_2} + \log_2\|\mathcal{S}\| \\
&\ge I(X_{l+1}^n;Y_{l+1}^n|S_l,U_1^l,U_{l+1}^n,Y_1^l,s_0^y(n))_{q_1q_2} \\
&= H(X_{l+1}^n|U_1^l,U_{l+1}^n,Y_1^l,S_l,s_0^y(n))_{q_1q_2} \\
&\qquad - H(X_{l+1}^n|Y_{l+1}^n,U_1^l,U_{l+1}^n,Y_1^l,S_l,s_0^y(n))_{q_1q_2} \\
&= H(X_{l+1}^n|U_{l+1}^n,S_l,s_0^y(n))_{q_1q_2} \\
&\qquad - H(X_{l+1}^n|Y_{l+1}^n,U_1^l,U_{l+1}^n,Y_1^l,S_l,s_0^y(n))_{q_1q_2} \\
&\ge H(X_{l+1}^n|U_{l+1}^n,S_l,s_0^y(n))_{q_1q_2} \\
&\qquad - H(X_{l+1}^n|Y_{l+1}^n,U_{l+1}^n,S_l,s_0^y(n))_{q_1q_2} \\
&= I(X_{l+1}^n;Y_{l+1}^n|U_{l+1}^n,S_l,s_0^y(n))_{q_1q_2} \\
&\ge I(X_{l+1}^n;Y_{l+1}^n|U_{l+1}^n,s_l=s_0^y(m))_{q_2}.
\end{aligned}$$

So, finally

$$\begin{aligned}
nF_n(\lambda) &\ge lF_l(\lambda) + I(U_{l+1}^n;Z_{l+1}^n|Z_1^l,s_0^z(n))_{q_1q_2} \\
&\qquad + \lambda I(X_{l+1}^n;Y_{l+1}^n|U_1^l,U_{l+1}^n,Y_1^l,s_0^y(n))_{q_1q_2} \\
&\ge lF_l(\lambda) + I(U_{l+1}^n;Z_{l+1}^n|s_l=s_0^z(m))_{q_2} \\
&\qquad + \lambda I(X_{l+1}^n;Y_{l+1}^n|U_{l+1}^n,s_l=s_0^y(m))_{q_2} \\
&\qquad - (1+\lambda)\log_2\|\mathcal{S}\| \\
&= lF_l(\lambda) + mF_m(\lambda).
\end{aligned}$$

Here we used the stationarity of the channel when conditioned on the initial state, which gives

$$\begin{aligned}
mF_m(\lambda) &= I(U_{l+1}^n;Z_{l+1}^n|s_l=s_0^z(m))_{q_2} \\
&\qquad + \lambda I(X_{l+1}^n;Y_{l+1}^n|U_{l+1}^n,s_l=s_0^y(m))_{q_2} \\
&\qquad - (1+\lambda)\log_2\|\mathcal{S}\|.
\end{aligned}$$

[10]**Lemma.** *Let* X,Y,Z,S be a joint ensemble. If S has a finite cardinality then $|I(X;Y|Z,S) - I(X;Y|Z)| \le \log_2\|\mathcal{S}\|$. This lemma implies that the RVs A^n,B^n,C^n,S satisfy

$$I(A^n;B^n|C^n) \ge I(A^n;B^n|C^n,S) - \log_2\|\mathcal{S}\|$$
$$I(A^n;B^n|C^n,S) \ge I(A^n;B^n|C^n) - \log_2\|\mathcal{S}\|.$$

In summary, the way we showed sup-additivity is by breaking the expressions for length n into expressions of length l and expressions of length m. The critical part here is to consider the length m sequence from $l+1$ to n. Here we used the fact that the channel is stationary, thus $p(z_{l+1}^n, y_{l+1}^n | x_{l+1}^n, s_l = s_0^*) = p(z_1^m, y_1^m | x_1^m = x_{l+1}^n, s_0^*)$ [follows from (2)]. This, combined with the fact that the cardinality bound depends on the length of the sequence and not on its starting point, leads to the conclusion that the same joint distribution on (U_1^m, X_1^m) that maximizes $F_m(\lambda)$ will maximize the segment from $l+1$ to n (i.e., be the maximizing distribution on (U_{l+1}^n, X_{l+1}^n), with the same initial state $S_l = s_0$).

We also have that

$$\begin{aligned} F_n(\lambda) = \max_{p(u^n,x^n)} &\left\{ \min_{s_0'' \in \mathcal{S}} \frac{1}{n} I(U^n; Z^n | s_0'') \right. \\ &\left. + \lambda \min_{s_0' \in \mathcal{S}} \frac{1}{n} I(X^n; Y^n | U^n, s_0') \right\} \\ &- (1+\lambda) \frac{\log_2 \|\mathcal{S}\|}{n} \\ \leq{}& \log_2 \|\mathcal{Z}\| + \lambda \log_2 \|\mathcal{X}\| \\ \leq{}& \log_2 \|\mathcal{Z}\| + \log_2 \|\mathcal{X}\| < \infty \end{aligned} \tag{C.2}$$

since $\lambda \leq 1$. The fact that $F_n(\lambda)$ is bounded, independent of n and is also sup-additive implies that $\lim_{n\to\infty} F_n(\lambda)$ converges and is equal to the supremum over all n[11]:

$$\lim_{n\to\infty} F_n(\lambda) = \sup_n F_n(\lambda) < \infty.$$

Appendix D
Proof of *Theorem* 2 (Indecomposable FSBCs)

A. Preliminaries

First, let us define

$$\begin{aligned} \overline{F_n(\lambda)} &= \max_{p(u^n,x^n)} \frac{1}{n} \left[\max_{s_0'' \in \mathcal{S}} I(U^n; Z^n | s_0'') \right. \\ &\qquad \left. + \lambda \max_{s_0' \in \mathcal{S}} I(X^n; Y^n | U^n, s_0') \right] \\ \underline{F_n(\lambda)} &= \max_{p(u^n,x^n)} \frac{1}{n} \left[\min_{s_0'' \in \mathcal{S}} I(U^n; Z^n | s_0'') \right. \\ &\qquad \left. + \lambda \min_{s_0' \in \mathcal{S}} I(X^n; Y^n | U^n, s_0') \right] \triangleq C^n(\lambda). \end{aligned} \tag{D.1}$$

Clearly $\overline{F_n(\lambda)} \geq \underline{F_n(\lambda)}$. Thus, if the limits $\lim_{n\to\infty} \underline{F_n(\lambda)}$ and $\lim_{n\to\infty} \overline{F_n(\lambda)}$ exist then

$$\lim_{n\to\infty} \underline{F_n(\lambda)} \leq \lim_{n\to\infty} \overline{F_n(\lambda)}.$$

In Appendix III, we showed that the limit $\lim_{n\to\infty} \underline{F_n(\lambda)}$ exists. The fact that for the indecomposable channel $\lim_{n\to\infty} \overline{F_n(\lambda)}$ exists will follow from the proof of Theorem 2.

[11]Here we use [5, Lemma 2 in Appendix 4A] which states that if a sequence $\{a_N\}_{N \in \mathbb{N}}$ satisfies that $\bar{a} = \sup_N a_N < \infty$, and for all $n \geq 1$ and all $N > n$, $N a_N \geq n a_n + (N-n) a_{N-n}$, then $\lim_{N\to\infty} a_N = \bar{a}$.

B. Proof of Theorem 2

Let $\tilde{p}(u^n, x^n)$, $\tilde{s}_0''$ and $\tilde{s}_0'$ be the distribution and the pair of initial states that maximize $\overline{F_n(\lambda)}$. Let $\underline{F}_n(\lambda, p(u^n, x^n))$ be the expression of $\underline{F_n(\lambda)}$ when the distribution is specified (i.e., the maximization over $p(u^n, x^n)$ in (D.1) is dropped) and let $\underset{\sim}{s_0''}$ and $\underset{\sim}{s_0'}$ be the states that minimize, respectively, the first and second mutual information expressions in $\underline{F}_n(\lambda, \tilde{p}(u^n, x^n))$. Then,

$$\underline{F_n(\lambda)} \geq \underline{F}_n(\lambda, \tilde{p}(u^n, x^n)).$$

Let m_1 be fixed, and in addition denote

$$\begin{aligned} &m_2 = n - m_1 \\ &\mathbf{u}_1 \triangleq u^{m_1}, \quad \mathbf{x}_1 \triangleq x^{m_1}, \quad \mathbf{y}_1 \triangleq y^{m_1}, \quad \mathbf{z}_1 \triangleq z^{m_1} \\ &\mathbf{u}_2 \triangleq u_{m_1+1}^n, \quad \mathbf{x}_2 \triangleq x_{m_1+1}^n, \quad \mathbf{y}_2 \triangleq y_{m_1+1}^n, \quad \mathbf{z}_2 \triangleq z_{m_1+1}^n. \end{aligned}$$

First consider

$$\begin{aligned} &p(\mathbf{z}_1, s_{m_1} | \mathbf{u}_1, \mathbf{u}_2, \tilde{s}_0'') \\ &= \sum_{\mathcal{X}^{m_1}, \mathcal{Y}^{m_1}, \mathcal{S}^{m_1-1}} p(\mathbf{x}_1, \mathbf{y}_1, \mathbf{s}_1, \mathbf{z}_1 | \mathbf{u}_1, \mathbf{u}_2, \tilde{s}_0'') \\ &= \sum_{\mathcal{X}^{m_1}, \mathcal{Y}^{m_1}, \mathcal{S}^{m_1-1}} p(\mathbf{y}_1, \mathbf{s}_1, \mathbf{z}_1 | \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2, \tilde{s}_0'') p(\mathbf{x}_1 | \mathbf{u}_1, \mathbf{u}_2, \tilde{s}_0'') \\ &\overset{(a)}{=} \sum_{\mathcal{X}^{m_1}, \mathcal{Y}^{m_1}, \mathcal{S}^{m_1-1}} p(\mathbf{y}_1, \mathbf{s}_1, \mathbf{z}_1 | \mathbf{x}_1, \tilde{s}_0'') p(\mathbf{x}_1 | \mathbf{u}_1, \mathbf{u}_2, \tilde{s}_0'') \\ &= \sum_{\mathcal{X}^{m_1}} p(\mathbf{z}_1, s_{m_1} | \mathbf{x}_1, \tilde{s}_0'') p(\mathbf{x}_1 | \mathbf{u}_1, \mathbf{u}_2, \tilde{s}_0'') \end{aligned}$$

where (a) follows from (2). Therefore, $p(\mathbf{z}_1, s_{m_1} | \mathbf{u}_1, \mathbf{u}_2, \tilde{s}_0'')$ is not independent of $\mathbf{U}_2$, since $\mathbf{U}_2$ can be correlated with $\mathbf{X}_1$. This is in contrast with the situation in [5, Theorem 4.6.4], where the conditional distribution for the first m_1 symbols is independent of the remaining symbols. The joint distribution can now be written as

$$\begin{aligned} &p(\mathbf{z}_2 | \mathbf{z}_1, \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2, s_{m_1}, \tilde{s}_0'') \\ &\quad = \sum_{\mathbf{x}_2} p(\mathbf{z}_2, \mathbf{x}_2 | \mathbf{z}_1, \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2, s_{m_1}, \tilde{s}_0'') \\ &\quad = \sum_{\mathbf{x}_2} \Big[p(\mathbf{x}_2 | \mathbf{z}_1, \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2, s_{m_1}, \tilde{s}_0'') \\ &\qquad\qquad \times p(\mathbf{z}_2 | \mathbf{z}_1, \mathbf{x}_1, \mathbf{x}_2, \mathbf{u}_1, \mathbf{u}_2, s_{m_1}, \tilde{s}_0'') \Big] \\ &\quad \overset{(a)}{=} \sum_{\mathbf{x}_2} p(\mathbf{x}_2 | \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2) p(\mathbf{z}_2 | \mathbf{z}_1, \mathbf{x}_1, \mathbf{x}_2, s_{m_1}, \tilde{s}_0'') \\ &\quad \overset{(b)}{=} \sum_{\mathbf{x}_2} p(\mathbf{x}_2 | \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2) p(\mathbf{z}_2 | \mathbf{x}_2, s_{m_1}) \\ &\quad = \sum_{\mathbf{x}_2} p(\mathbf{x}_2 | \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2, s_{m_1}) p(\mathbf{z}_2 | \mathbf{x}_1, \mathbf{x}_2, s_{m_1}, \mathbf{u}_1, \mathbf{u}_2) \\ &\quad = p(\mathbf{z}_2 | \mathbf{x}_1, s_{m_1}, \mathbf{u}_2, \mathbf{u}_1) \end{aligned} \tag{D.2}$$

where (a) is because there is no feedback and (b) is because given $(S_{m_1}, \mathbf{X}_2)$, $\mathbf{Z}_2$ is independent of the past.

Now expand $I(\mathbf{U}_1, \mathbf{U}_2; \mathbf{Z}_1, \mathbf{Z}_2 | \tilde{s}_0'')_{\tilde{p}}$ in $\overline{F_n(\lambda)}$ as

$$\begin{aligned} &I(\mathbf{U}_1, \mathbf{U}_2; \mathbf{Z}_1, \mathbf{Z}_2 | \tilde{s}_0'')_{\tilde{p}} \\ &\quad = I(\mathbf{U}_1, \mathbf{U}_2; \mathbf{Z}_1 | \tilde{s}_0'')_{\tilde{p}} + I(\mathbf{U}_1, \mathbf{U}_2; \mathbf{Z}_2 | \mathbf{Z}_1, \tilde{s}_0'')_{\tilde{p}} \\ &\quad \leq m_1 \log_2 \|\mathcal{Z}\| + I(\mathbf{U}_1, \mathbf{U}_2; \mathbf{Z}_2 | \mathbf{Z}_1, \tilde{s}_0'')_{\tilde{p}} \end{aligned}$$

$$\overset{(a)}{\leq} m_1 \log_2 ||\mathcal{Z}|| + \log_2 ||\mathcal{S}|| + m_1 \log_2 ||\mathcal{X}|| + I(\mathbf{U}_1, \mathbf{U}_2; \mathbf{Z}_2 | S_{m_1}, \mathbf{X}_1, \mathbf{Z}_1, \tilde{s}_0'')_{\tilde{p}}$$

where (a) is due to [5, Lemma 1 in Appendix 4A]. Similarly we obtain for $I(U^n; Z^n | \underset{\sim}{s_0''})_{\tilde{p}}$ in $\underline{\mathrm{F}}_n(\lambda, \tilde{p}(u^n, x^n))$

$$\begin{aligned} I(U^n; Z^n | \underset{\sim}{s_0''})_{\tilde{p}} &= I(\mathbf{U}_1, \mathbf{U}_2; \mathbf{Z}_1, \mathbf{Z}_2 | \underset{\sim}{s_0''})_{\tilde{p}} \\ &\geq -\log_2 ||\mathcal{S}|| - m_1 \log_2 ||\mathcal{X}|| \\ &\quad + I(\mathbf{U}_1, \mathbf{U}_2; \mathbf{Z}_2 | S_{m_1}, \mathbf{X}_1, \mathbf{Z}_1, \underset{\sim}{s_0''})_{\tilde{p}}. \end{aligned}$$

Therefore

$$\begin{aligned} &\frac{1}{n}\left(I(U^n; Z^n | \tilde{s}_0'')_{\tilde{p}} - I(U^n; Z^n | \underset{\sim}{s_0''})_{\tilde{p}} \right) \\ &\leq \frac{1}{n}\Big(m_1 \log_2 ||\mathcal{Z}|| + 2\log_2 ||\mathcal{S}|| + 2m_1 \log_2 ||\mathcal{X}|| \\ &\qquad + I(\mathbf{U}_1, \mathbf{U}_2; \mathbf{Z}_2 | S_{m_1}, \mathbf{X}_1, \mathbf{Z}_1, \tilde{s}_0'')_{\tilde{p}} \\ &\qquad - I(\mathbf{U}_1, \mathbf{U}_2; \mathbf{Z}_2 | S_{m_1}, \mathbf{X}_1, \mathbf{Z}_1, \underset{\sim}{s_0''})_{\tilde{p}} \Big) \\ &\overset{(a)}{=} \frac{1}{n}\Big(m_1 \log_2 ||\mathcal{Z}|| + 2\log_2 ||\mathcal{S}|| + 2m_1 \log_2 ||\mathcal{X}|| \\ &\qquad + \sum_{\mathcal{X}^{m_1}, s_{m_1} \in \mathcal{S}} \left(p(s_{m_1}, x^{m_1} | \tilde{s}_0'') - p(s_{m_1}, x^{m_1} | \underset{\sim}{s_0''}) \right) \\ &\qquad \times I(\mathbf{U}_1, \mathbf{U}_2; \mathbf{Z}_2 | x^{m_1}, s_{m_1}) \Big) \\ &\overset{(b)}{\leq} \frac{1}{n}\Big(m_1 \log_2 ||\mathcal{Z}|| + 2\log_2 ||\mathcal{S}|| + 2m_1 \log_2 ||\mathcal{X}|| \\ &\qquad + \sum_{\mathcal{X}^{m_1}} p(x^{m_1}) \epsilon ||\mathcal{S}|| (n - m_1) \log_2 ||\mathcal{Z}|| \Big) \\ &= \frac{1}{n}\Big(m_1 \log_2 ||\mathcal{Z}|| + 2\log_2 ||\mathcal{S}|| \\ &\qquad + 2m_1 \log_2 ||\mathcal{X}|| + \epsilon ||\mathcal{S}|| (n - m_1) \log_2 ||\mathcal{Z}|| \Big) \\ &\overset{n\to\infty}{\longrightarrow} \epsilon ||\mathcal{S}|| \log_2 ||\mathcal{Z}|| \end{aligned} \tag{D.3}$$

where (a) follows from (D.2) and (b) is due to Definition 2. Now, for $I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_1, \mathbf{Y}_2 | \mathbf{U}_1, \mathbf{U}_2, \tilde{s}_0')$ in $\overline{F_n(\lambda)}$ we have

$$\begin{aligned} I(\mathbf{X}_1, &\mathbf{X}_2; \mathbf{Y}_1, \mathbf{Y}_2 | \mathbf{U}_1, \mathbf{U}_2, \tilde{s}_0')_{\tilde{p}} \\ &= I(\mathbf{X}_1; \mathbf{Y}_1, \mathbf{Y}_2 | \mathbf{U}_1, \mathbf{U}_2, \tilde{s}_0')_{\tilde{p}} \\ &\qquad + I(\mathbf{X}_2; \mathbf{Y}_1, \mathbf{Y}_2 | \mathbf{X}_1, \mathbf{U}_1, \mathbf{U}_2, \tilde{s}_0')_{\tilde{p}} \\ &\leq \log_2 ||\mathcal{X}^{m_1}|| + I(\mathbf{X}_2; \mathbf{Y}_1 | \mathbf{X}_1, \mathbf{U}_1, \mathbf{U}_2, \tilde{s}_0')_{\tilde{p}} \\ &\qquad + I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{Y}_1, \mathbf{X}_1, \mathbf{U}_1, \mathbf{U}_2, \tilde{s}_0')_{\tilde{p}} \\ &\leq m_1 \log_2 ||\mathcal{X}|| + m_1 \log_2 ||\mathcal{Y}|| + \log_2 ||\mathcal{S}|| \\ &\qquad + I(\mathbf{X}_2; \mathbf{Y}_2 | S_{m_1}, \mathbf{Y}_1, \mathbf{X}_1, \mathbf{U}_1, \mathbf{U}_2, \tilde{s}_0')_{\tilde{p}}. \end{aligned}$$

Similarly for $I(X^n; y^n | U^n, \underset{\sim}{s_0'})_{\tilde{p}}$ in $\underline{\mathrm{F}}_n(\lambda, \tilde{p}(u^n, x^n))$

$$\begin{aligned} &I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}_1, \mathbf{Y}_2 | \mathbf{U}_1, \mathbf{U}_2, \underset{\sim}{s_0'})_{\tilde{p}} \\ &\qquad \geq -\log_2 ||\mathcal{S}|| + I(\mathbf{X}_2; \mathbf{Y}_2 | S_{m_1}, \mathbf{Y}_1, \mathbf{X}_1, \mathbf{U}_1, \mathbf{U}_2, \underset{\sim}{s_0'})_{\tilde{p}}. \end{aligned}$$

Thus,

$$\begin{aligned} &\frac{1}{n}\left(I(X^n; Y^n | U^n, \tilde{s}_0')_{\tilde{p}} - I(X^n; Y^n | U^n, \underset{\sim}{s_0'})_{\tilde{p}} \right) \\ &\leq \frac{1}{n}\Big(m_1 \log_2 ||\mathcal{X}|| + m_1 \log_2 ||\mathcal{Y}|| + 2\log_2 ||\mathcal{S}|| \\ &\qquad + I(\mathbf{X}_2; \mathbf{Y}_2 | S_{m_1}, \mathbf{Y}_1, \mathbf{X}_1, \mathbf{U}_1, \mathbf{U}_2, \tilde{s}_0')_{\tilde{p}} \\ &\qquad - I(\mathbf{X}_2; \mathbf{Y}_2 | S_{m_1}, \mathbf{Y}_1, \mathbf{X}_1, \mathbf{U}_1, \mathbf{U}_2, \underset{\sim}{s_0'}) \Big)_{\tilde{p}} \\ &\overset{(a)}{=} \frac{1}{n}\Big(m_1 \log_2 ||\mathcal{X}|| + m_1 \log_2 ||\mathcal{Y}|| + 2\log_2 ||\mathcal{S}|| \\ &\qquad + \sum_{\mathcal{X}^{m_1}} p(x^{m_1}) \sum_{s_{m_1} \in \mathcal{S}} \left(p(s_{m_1} | x^{m_1}, \tilde{s}_0') - p(s_{m_1} | x^{m_1}, \underset{\sim}{s_0'}) \right) \\ &\qquad \times I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{U}_2, \mathbf{U}_1, s_{m_1}, x^{m_1})_{\tilde{p}} \Big) \\ &\leq \frac{1}{n}\Big(m_1 \log_2 ||\mathcal{X}|| + m_1 \log_2 ||\mathcal{Y}|| + 2\log_2 ||\mathcal{S}|| \\ &\qquad + \sum_{\mathcal{X}^{m_1}} p(x^{m_1}) \sum_{s_{m_1} \in \mathcal{S}} \left| p(s_{m_1} | x^{m_1}, \tilde{s}_0') - p(s_{m_1} | x^{m_1}, \underset{\sim}{s_0'}) \right| \\ &\qquad \times \left| (n - m_1) \log_2 ||\mathcal{X}|| \right| \Big) \\ &\overset{(b)}{\leq} \frac{1}{n}\Big(m_1 \log_2 ||\mathcal{X}|| + m_1 \log_2 ||\mathcal{Y}|| + 2\log_2 ||\mathcal{S}|| \\ &\qquad + \sum_{\mathcal{X}^{m_1}} p(x^{m_1}) \epsilon ||\mathcal{S}|| (n - m_1) \log_2 ||\mathcal{X}|| \Big) \\ &\overset{n\to\infty}{\longrightarrow} \epsilon ||\mathcal{S}|| \log_2 ||\mathcal{X}|| \end{aligned} \tag{D.4}$$

where for (a) we write

$$\begin{aligned} &p(\mathbf{y}_2 | s_{m_1}, \mathbf{y}_1, \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2, s_0) \\ &\quad = \sum_{\mathcal{X}^n_{m_1+1}} p(\mathbf{y}_2, \mathbf{x}_2 | s_{m_1}, \mathbf{y}_1, \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2, s_0) \\ &\quad = \sum_{\mathcal{X}^n_{m_1+1}} \Big[p(\mathbf{x}_2 | s_{m_1}, \mathbf{y}_1, \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2, s_0) \\ &\qquad\qquad \times p(\mathbf{y}_2 | \mathbf{x}_2, s_{m_1}, \mathbf{y}_1, \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2, s_0) \Big] \\ &\quad \overset{(c)}{=} \sum_{\mathcal{X}^n_{m_1+1}} p(\mathbf{x}_2 | \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2) p(\mathbf{y}_2 | \mathbf{x}_2, s_{m_1}, \mathbf{y}_1, \mathbf{x}_1, s_0) \\ &\quad = \sum_{\mathcal{X}^n_{m_1+1}} p(\mathbf{x}_2 | \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2) p(\mathbf{y}_2 | \mathbf{x}_2, s_{m_1}) \\ &\quad = \sum_{\mathcal{X}^n_{m_1+1}} p(\mathbf{x}_2 | s_{m_1}, \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2) p(\mathbf{y}_2 | \mathbf{x}_2, s_{m_1}, \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2) \\ &\quad = \sum_{\mathcal{X}^n_{m_1+1}} p(\mathbf{y}_2, \mathbf{x}_2 | s_{m_1}, \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2) \\ &\quad = p(\mathbf{y}_2 | s_{m_1}, \mathbf{x}_1, \mathbf{u}_1, \mathbf{u}_2) \end{aligned}$$

where (c) is because there is no feedback. Finally, (b) follows from the definition of the indecomposable channel (see Definition 2).

Combining (D.3) with (D.4), we obtain that if the limits exist then

$$\lim_{n\to\infty} \left(\overline{F_n(\lambda)} - \underline{\mathrm{F}}_n(\lambda, \tilde{p}(u^n, x^n)) \right) \leq \epsilon ||\mathcal{S}|| (\log_2 ||\mathcal{Z}|| + \lambda \log_2 ||\mathcal{X}||)$$

as $\overline{F_n(\lambda)} \geq \underline{F_n(\lambda)} \geq \underline{\mathrm{F}}_n(\lambda, \tilde{p}(u^n, x^n))$ then, using $0 \leq \lambda \leq 1$ and the sandwich theorem also

$$\lim_{n\to\infty} \left(\overline{F_n(\lambda)} - \underline{F_n(\lambda)} \right) \leq \epsilon ||\mathcal{S}|| (\log_2 ||\mathcal{Z}|| + \log_2 ||\mathcal{X}||).$$

Finally, because this is true for any $\epsilon > 0$, then taking $\epsilon \to 0$ both bounds coincide, namely

$$\lim_{n\to\infty} \overline{F_n(\lambda)} = \lim_{n\to\infty} \underline{F_n(\lambda)}.$$

As the right-hand side exists and finite, then this also proves that $\lim_{n\to\infty} \overline{F_n(\lambda)}$ exists and finite. This means that when the channel is indecomposable the capacity region of Theorem 1 is the same for all initial states.

REFERENCES

[1] T. M. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.
[2] B. McMillan, "The basic theorems of information theory," *Ann. Math. Statist.*, vol. 24, no. 2, pp. 196–219, 1953.
[3] C. E. Shannon, "Certain results in coding theory for noisy channels," *Inf. Control*, vol. 1, pp. 6–25, Sep. 1957.
[4] D. Blackwell, L. Breiman, and A. J. Thomasian, "Proof of Shannon's transmission theorem for finite-state indecomposable channels," *Ann. Math. Statist.*, vol. 29, no. 4, pp. 1209–1220, 1958.
[5] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[6] A. J. Goldsmith and P. P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inf. Theory*, vol. IT-42, pp. 868–886, Mar. 1996.
[7] T. Holliday, A. Goldsmith, and P. Glynn, "Capacity of finite state channels based on Lyapunov exponents of random matrices," *IEEE Trans. Inf. Theory*, vol. IT-52, pp. 3509–3532, Aug. 2006.
[8] H. Permuter, P. Cuff, B. Van Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-54, pp. 3150–3165, Jul. 2008.
[9] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory*, vol. IT-55, pp. 323–349, Jan. 2009.
[10] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacities of certain channel classes under random coding," *Ann. Math. Statist.*, vol. 31, no. 3, pp. 558–567, 1960.
[11] A. Lapidoth and P. Narayan, "Relieble communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. IT-44, pp. 2148–2177, Jun. 1998.
[12] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Inf. Control*, vol. 9, no. 1, pp. 19–31, 1980.
[13] A. Winshtok and Y. Steinberg, "The arbitrarily varying degraded broadcast channel with states known at the encoder," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Seattle, WA, 2006, pp. 2156–2160.
[14] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information," *IEEE Trans. Inf. Theory*, vol. IT-51, pp. 2867–2877, Aug. 2005.
[15] 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. IT-55, pp. 2455–2477, Jun. 2009.
[16] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, vol. 10, no. 3, pp. 3–14, 1974.
[17] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inf. Theory*, vol. IT-19, pp. 197–207, Feb. 1973.
[18] O. Kallenberg, *Foundations of Modern Probability*. New York: Springer-Verlag, 1997.
[19] P. Sadeghi, P. O. Vontobel, and R. Shams, "Optimization of information rate upper and lower bounds for channels with memory," *IEEE Trans. Inf. Theory*, vol. IT-55, pp. 663–688, Feb. 2009.
[20] A. J. Goldsmith and M. Effros, "The capacity region of broadcast channels with intersymbol interference and colored Gaussian noise," *IEEE Trans. Inf. Theory*, vol. IT-47, pp. 219–240, Jan. 2001.
[21] A. Lapidoth and I. E. Telater, "The compound channel capacity of a class of finite-state channels," *IEEE Trans. Inf. Theory*, vol. IT-44, pp. 973–983, Mar. 1998.
[22] T. M. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[23] J. L. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory and its Appl.*, Waikiki, HI, Nov. 1990, pp. 303–305.
[24] G. Salinetti and R. J.-B. Wets, "On the convergence of sequences of convex sets in finite dimensions," *SIAM Rev.*, vol. 21, no. 1, pp. 18–33, Jan. 1979.
[25] G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*. Cambridge , U.K.: Cambridge Press, 1964.
[26] M. Salehi, Cardinality Bounds on Auxiliary Variables in Multiple-User Theory via the Method of Ahlswede and Körner. Stanford Univ., Dep. Statist., Stanford, CA, 1978, Technical Report No. 33.
[27] A. El-Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-24, pp. 379–381, Mar. 1978.
[28] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert–Elliott channel," *IEEE Trans. Inf. Theory*, vol. IT-35, pp. 1277–1290, Jun. 1989.
[29] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 2000.
[30] Y. Steinberg and S. Shamai, "Achievable rates for the broadcast channel with states known at the transmitter," in *Proc. Int. Symp. Inf. Theory (ISIT)*, Adelaide, Australia, 2005, p. 2184.
[31] S. Verdu, "Multiple-access channels with memory with and without frame synchronism," *IEEE Trans. Inf. Theory*, vol. IT-35, pp. 605–619, Mar. 1989.
[32] S. Verdu, "The capacity region of the symbols-asynchronous Gaussain multiple-access channel," *IEEE Trans. Inf. Theory*, vol. IT-35, pp. 733–751, Apr. 1989.
[33] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. IT-49, pp. 4–21, Jan. 2003.
[34] C. E. Shannon, J. Neyman, Ed., "Two-way communication channels," in *Proc. 4th Berkeley Symp. Math. Statist. Probabil.*, Berkeley, CA, 1961, vol. 1, pp. 611–644.
[35] R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Thoery*, vol. IT-21, pp. 629–637, Nov. 1975.
[36] M. Krein and D. Milman, "On extreme points of regular convex sets," *Studia Mathematica*, vol. 9, pp. 133–138, 1940.

Ron Dabora (M'03) received the B.Sc. and M.Sc. degrees in 1994 and 2000, respectively, from Tel-Aviv University, and the Ph.D. degree in 2007 from Cornell University, Ithaca, NY, all in electrical engineering.

From 1994 to 2000, he worked as an engineer with 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 scholar with the Department of Electrical Engineering, Stanford University, Stanford, CA. Since 2009, he has been an Assistant Professor with the Department of Electrical and Computer Engineering, Ben-Gurion University, Israel.

Andrea J. Goldsmith (S'90–M'93–SM'99–F'03) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of California, Berkeley.

She is a professor of Electrical Engineering with Stanford University, and was previously an Assistant Professor of Electrical Engineering at the California Institute of Technology, 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 author of the book "Wireless Communications" and coauthor of *MIMO Wireless Communications* (Cambridge, U.K.: Cambridge University Press).

Dr. Goldsmith is a Fellow of Stanford University. 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 Editor for the *Journal on Foundations and Trends in Communications* and *Information Theory and in Networks*. 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 Co-Chair for the 2007 IEEE International Symposium on Information Theory. She also founded the student committee of the IEEE Information Theory society. She is an inaugural recipient of Stanford's postdoctoral mentoring award, and serves as Stanford University's Faculty Senate Chair for the 2009/2010 academic year.