The Capacity Region of the Degraded Finite-State Broadcast Channel

Ron Dabora and Andrea Goldsmith

Wireless Systems Lab
Department of Electrical Engineering
Stanford University

Abstract

We consider the discrete, time-varying broadcast channel with memory, under the assumption that the channel states belong to a set of finite cardinality. We begin with the definition of the general finite-state broadcast channel and discuss the possible definitions of the probability of error for this model. Next, we define the physically degraded finite-state broadcast channel and show that a superposition codebook with memory achieves capacity for this scenario. We then define the stochastically degraded finite-state broadcast channel, and relate this model to a physical communication scenario. We use the capacity result for the physically degraded channel to characterize the capacity of the stochastically degraded one. In both scenarios we consider the general finite-state channel as well as the indecomposable one, in which the effect of the initial state on future states becomes negligible with time.

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 (Rx1) and \( z \) is the channel output at receiver 2 (Rx2). In the years following its introduction the study of the BC focused on the memoryless scenario, i.e., when the probability of a block of \( n \) transmissions is given by \( p(y^n, z^n|x^n) = \prod_{i=1}^{n} p(y_i, z_i|x_i) \). In recent years, capacity analysis of time-varying channels with memory has been the focus of considerable interest, especially in the

The authors are with the Wireless Systems Lab, Department of Electrical Engineering, Stanford University, Stanford, CA 94305. Email: \{ron, andrea\}@wsl.stanford.edu. Parts of this paper will be presented at the IEEE Information Theory Workshop (ITW), Porto, Portugal, May 2008.

April 30, 2008 DRAFT
Gaussian setup. This has been motivated by the proliferation of mobile communications for which the channel is subject to multipath and correlated fading. The correlation of the fading process introduces memory to the channel.

There are several approaches to model time-varying channels with memory. The approach we focus on in this paper is the finite-state channel (FSC) model, introduced as early as 1953 [2]. In this model the channel memory is captured by the state of the channel at the end of the previous symbol transmission. Letting \( S_{i-1} \) denote the state of the channel at the end of transmission interval \( i - 1 \), the transition function of the FSC at the \( i \)'th 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 previous state, \( S_{i-1} \), makes \( (Y_i, S_i) \) independent of the entire history \( (X^{i-1}, Y^{i-1}, S^{i-1}) \). The capacity of FSCs without feedback was originally studied by Shannon in 1957 [3]. In his work, Shannon derived the capacity of indecomposable FSCs in situations where the transmitter can calculate the state sequence (i.e. the current state is a function of the previous state and the current channel input). Later, Blackwell, Breiman and Thomasian derived the capacity of indecomposable FSCs without Shannon’s restrictions [4]. The general (i.e. non-indecomposable) FSC was studied by Gallager, who included an extensive treatment of the subject in his book [5]. The capacity of FSCs for specific classes has been studied by several other authors, see [6], [7] and references therein. Recently, the capacity of FSCs with feedback has been studied [8]. Specifically, it was shown in [9] that feedback can increase the capacity of some FSCs. 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 past 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. When the alphabets are continuous, and the channel output consists of a weighted, finite linear combination of the input symbols with additive colored Gaussian noise, this model represents a Gaussian scenario with finite intersymbol interference [5]. Note, however, that for modeling memory using channel states, the latter case requires an infinite state space. Finally, we note that for finite-state channels, the rate expressions usually involve limits of mutual information expressions taken with blocklength increasing to infinity. The computation of such limits has also been considered, see for example [18].

There are other models for time-varying channels. One model is the arbitrarily varying channel (AVC). 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], [12]). The state transitions in the AVC are independent of the channel input and output symbols. For a comprehensive discussion on models for time-varying
channels see the survey paper [11].

In the context of discrete multi-user channels with states, the capacity of degraded arbitrarily varying broadcast channels (DAVBCs) has been recently investigated in [13] and [14]. In [13] DAVBCs with causal and non-causal side information at the transmitter are considered, and in [14] the capacity of DAVBCs with causal side information at the transmitter and non-causal side information at the good receiver is derived. The finite-state multiple-access channel (MAC) was studied in [15]. This scenario is characterized by the channel transition function \( p(y, s|x_1, x_2, s') \), and the work in [15] also considered the effect of feedback on the achievable rates.

In this work we study the finite-state broadcast channel (FSBC), depicted in Figure 1. In the FSBC scenario, the channel from the transmitter to the receivers is governed by a state sequence that depends on the channel inputs, outputs and previous states. 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 non-indecomposable channels. We will provide a precise definition of indecomposable FSBCs in Definition 2. Loosely speaking, in indecomposable channels, the effect of the initial state on future states becomes negligible as time evolves. For non-indecomposable channels the initial state may affect the state sequence indefinitely. Thus, for example, there may be states that will never be revisited a second time. We consider both types of channels in this work.

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]. We also note that the finite-memory Gaussian BC was considered in the work of [19], in which the capacity region of Gaussian BCs with intersymbol interference and colored Gaussian noise was derived.

Main Contributions and Organization

We study the capacity region of the degraded FSBC assuming that the receivers and transmitter operate without knowledge of the channel states. First, we consider the question of defining the achievable rate based on the average probability of error going to zero with code blocklength. Here, there are two possible approaches. One approach is the compound channel approach (see [20]), where we require that the achievable rates have their average probability of error go to zero for every initial state. The second approach is to define the average probability of error using only the channel input and outputs in the definition, without explicitly considering the states (see [15]). For non-indecomposable channels, the compound approach seems the more suitable of the two approaches since we generally require the code associated with an achievable rate to guarantee a small probability of error for every possible initial state. When the channel is indecomposable it would appear that the definition of average probability of error should not depend on the initial state and, in fact, both error probability definitions proposed above lead to the same capacity result. We will expand on these ideas in the next section. We also define, for the first time, the notion of degradedness for a BC with memory. We then prove an achievability result and an upper bound, with the average probability of error defined as in the compound channel approach. These results give the capacity region of the non-indecomposable degraded FSBC and of the indecomposable degraded FSBC, both for the physically degraded and stochastically degraded channels.

The rest of this paper is organized as follows: in Section II we introduce the channel model and state the formal definitions for the scenario. In Section III we present the capacity results for the general degraded FSBC and the indecomposable FSBC. Finally, in Section IV we conclude the paper. The proof of the converse and the achievability theorem are given in Appendix B and Appendix C, respectively. In Appendix F we discuss the implication of the alternative definition for the probability of error.

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 obvious from the context. 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 $i$th element of a vector $\mathbf{x}$ is denoted with $x_i$ and we use $x_i^j$ where $i \leq j$ to denote the vector $(x_i, x_{i+1}, \ldots, x_j)$; $x^j$ is a short form notation for $x_1^j$, and $x \equiv x^n$. 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 $X_i^j \triangleq (X_i, X_{i+1}, \ldots, X_j)$ for $i \leq j$. 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 [21, Chapter 2]. $I(\cdot; \cdot)_q$ denotes the mutual information evaluated with a p.m.f. $q$ on the random variables. Finally, $\mathcal{co} \mathcal{R}$ denotes the convex hull of the set $\mathcal{R}$, $\mathbb{N}$ denotes the set of natural numbers, and we use $X \perp Y$ 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 a block of $n$ transmissions is

$$p(y^n, z^n, s^n, x^n|s_0)$$

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

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

$$\overset{(b)}{=} \prod_{i=1}^{n} p(x_i|x^{i-1}, y^{i-1}, z^{i-1}) \prod_{i=1}^{n} p(y_i, z_i, s_i|x_i, s_{i-1}), \tag{1}$$

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 $\{s_k\}_{k=0}^{i-1}$ when $x^{i-1}$, $y^{i-1}$ and $z^{i-1}$ are given, and (b) captures the fact that given $s_{i-1}$, 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 [22])

\[ p(x_i|x^{i-1}, y^{i-1}, z^{i-1}) = p(x_i|x^{i-1}), \]

\[ \Rightarrow \prod_{i=1}^{n} p(x_i|x^{i-1}, y^{i-1}, z^{i-1}) = p(x^n) = p(x^n|s_0), \]

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

(2)

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

\[ |p(s_n|x, s_0) - p(s_n|x, s'_0)| \leq \epsilon, \]

for all \( s_n, x, \) and initial states \( s_0 \) and \( s'_0 \).

Note that (3) is identical to the definition of the indecomposable point-to-point FSC [5, Equation 4.6.26]. This is because indecomposability characterizes the interaction of the states and the channel inputs, averaging out the channel output(s), [5, Section 4.6]. As in both the point-to-point channel and the broadcast channel there is a single channel input, the criteria for the channel to be indecomposable is the same, despite the fact that the models are different.

**Definition 3.** An \((R_0, R_1, R_2, n)\) *deterministic code* for the FSBC consists of three message sets, \( \mathcal{M}_0 = \{1, 2, \ldots, 2^{nR_0}\} \), \( \mathcal{M}_1 = \{1, 2, \ldots, 2^{nR_1}\} \) and \( \mathcal{M}_2 = \{1, 2, \ldots, 2^{nR_2}\} \), and three mappings \((f, g_y, g_z)\) such that

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

is the encoder, and

\[ g_y : \mathcal{Y}^n \mapsto \mathcal{M}_0 \times \mathcal{M}_1, \]

\[ g_z : \mathcal{Z}^n \mapsto \mathcal{M}_0 \times \mathcal{M}_2, \]

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 \( \text{Rx}_1 \) and \( \text{Rx}_2 \), 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 \( \max_{s_0 \in S} P_e^{(n)}(s_0) \), where

\[ P_e^{(n)}(s_0) = \Pr \left( (g_y(Y^n) \neq (M_0, M_1) \text{ or } g_z(Z^n) \neq (M_0, M_2)|s_0) \right) \]
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 over 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 \( \max_{s_0 \in S} P_e^{(n)}(s_0) \leq \epsilon \).

Trivially, this requires \( P_e^{(n)}(s_0) \leq \epsilon \) for every initial state \( s_0 \in S \).

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

As discussed in Section I, there are two possible definitions for the average probability of error. Each definition 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 \( s_0 \), as is done, for example, in [20]. Alternatively, we can define the average probability of error without explicitly considering the initial state, as was done, for example, in [15]. This leads to two possible approaches to defining the achievable rates, each associated with a different definition of average error probability. The first approach is used in Definition 5: this is the *compound channel* approach (see also [11], [20]), where it is required that the probability of error can be made arbitrarily small for every initial state. This definition is particularly useful for non-indecomposable channels, since for such channels the effect of \( s_0 \) may never fade and we must account for all initial states. 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 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 non-indecomposable, 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 lead to the same capacity region. In this paper we use the compound approach. Details on the alternative approach can be found in Appendix F.

We now proceed to the definition of the *degraded* FSBC. Before stating the definition we remark that in this work 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 the discrete, memoryless broadcast channel (DMBC), physical degradedness is characterized via a “single-letter” relationship

\[ p(z|y, x) = p(z|y). \]  \hspace{2cm} (4)

In the memoryless case this is sufficient 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 will be made mathematically precise shortly:

S1. The channel output at the weak receiver, \( Z^i \), does not “contain information” not already present at the channel output of the strong receiver, \( Y^i \).

S2. \( Y^i \) makes \( Z^i \) independent of \( X^i \).

The single letter characterization (4) suffices to capture S1 and S2 for memoryless BCs. However, when there is memory, a single-letter characterization is not possible. Thus, we have the following definition:

**Definition 7.** The FSBC is called physically degraded if its p.m.f. satisfies

\[ p(y_i|x^i, y^{i-1}, z^{i-1}, s_0) = p(y_i|x^i, y^{i-1}, s_0), \]  \hspace{2cm} (5a)

\[ p(z_i|x^i, y^i, z^{i-1}, s_0) = p(z_i|y^i, z^{i-1}, s_0). \]  \hspace{2cm} (5b)

Condition (5a) 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 (5b) follows from the standard notion of degradedness (statement S2).
Using conditions (5a) and (5b) we obtain (when \( p(y^n, x^n|s_0) > 0 \))

\[
p(z^n|y^n, x^n, s_0) = \frac{p(z^n, y^n, x^n|s_0)}{p(y^n, x^n|s_0)}
\]

\[
= \prod_{i=1}^{n} p(z_i, y_i, x_i|z^{i-1}, y^{i-1}, x^{i-1}, s_0)
\]

\[
= \prod_{i=1}^{n} p(x_i|z^{i-1}, y^{i-1}, x^{i-1}) \prod_{i=1}^{n} p(z_i|y^{i-1}, x^{i}, s_0)
\]

\[
= (a) \prod_{i=1}^{n} p(x_i|z^{i-1}) \prod_{i=1}^{n} p(y_i|y^{i-1}, x^{i}, s_0)
\]

\[
= (b) \prod_{i=1}^{n} p(y_i|y^{i-1}, x^{i}, s_0) \prod_{i=1}^{n} p(z_i|z^{i-1}, y^{i}, x^{i}, s_0)
\]

\[
= (c) \prod_{i=1}^{n} p(z_i|z^{i-1}, y^i, s_0),
\]

where (a) is because there is no feedback, (b) follows from (5a) and (c) follows from (5b). We therefore conclude that when (5) holds then

\[
p(z^n|y^n, x^n, s_0) = p(z^n|y^n, s_0),
\]

hence \( p(y^n, z^n|x^n, s_0) = p(y^n|x^n, s_0)p(z^n|y^n, s_0) \). Note that (6) shows how to obtain \( p(z^n|y^n, x^n, s_0) \) in a causal manner. Also note that \( Z^n \) is a degraded version of \( Y^n \) but still depends on the state sequence (i.e., degradedness does not eliminate the memory).

A special case of the physically degraded FSBC occurs when in (5b) it holds that \( p(z_i|x^i, y^i, z^{i-1}, s_0) = p(z_i|y_i) \). Hence,

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

Equation (8) is similar to the definition of degradedness for the DAVBC used in [13].

**Definition 8.** The FSBC is called *stochastically degraded* if there exists a p.m.f. \( \tilde{p}(z|y) \) such that for every blocklength \( n \)

\[
p(z^n|x^n, s_0) = \sum_{y^n} p(z^n, y^n|x^n, s_0)
\]

\[
= \sum_{y^n} p(y^n|x^n, s_0) \prod_{i=1}^{n} \tilde{p}(z_i|y_i).
\]
Note that (9) is satisfied if there exists a p.m.f. $\hat{p}(z|y)$ such that

$$p(z_i, s_i|x_i, s_{i-1}) = \sum_{y} p(y_i, s_i|x_i, s_{i-1}) p(z_i|y_i, s_i, x_i, s_{i-1})$$

$$= \sum_{y} p(y_i, s_i|x_i, s_{i-1}) \hat{p}(z_i|y_i).$$

(10)

In fact, when (10) holds then also (5b) holds (with $\hat{p}(z_i|y_i)$ on the right-hand side). Using (10) we can write the p.m.f. $p(z^n|x^n, s_0)$ as

$$p(z^n|x^n, s_0) = \sum_{S^n} p(z^n, s^n|x^n, s_0)$$

$$= (a) \sum_{S^n} \prod_{i=1}^{n} p(z_i, s_i|x_i, s_{i-1}, x_i)$$

$$= (b) \sum_{S^n} \prod_{i=1}^{n} \sum_{y \in Y} p(y, s_i|x_i, s_{i-1}, x_i) \hat{p}(z_i|y_i)$$

(c) $$= \sum_{S^n} \sum_{Y^n} \prod_{i=1}^{n} p(y, s_i|x_i, s_{i-1}, x_i) \hat{p}(z_i|y_i)$$

$$= \sum_{S^n} \sum_{Y^n} \prod_{i=1}^{n} p(y, s_i|x_i, s_{i-1}, x_i) \hat{p}(z_i|y_i)$$

$$= \sum_{S^n} \sum_{Y^n} \prod_{i=1}^{n} p(y, s_i|x_i, s_{i-1}, x_i) \hat{p}(z_i|y_i)$$

(11)

where (a) follows from (2) after summing over $Y^n$, (b) follows from (10), and to see (c) we write explicitly the case of $n = 3$ (the same steps can be repeated for an arbitrary $n$ since $y_i$ appears only in the $i$'th summation):

$$\prod_{i=1}^{3} \left( \sum_{y_i \in Y} p(y_i, s_i|x_i, s_{i-1}, x_i) \hat{p}(z_i|y_i) \right)$$

$$= \left( \sum_{y_1 \in Y} p(y_1, s_1|x_1) \hat{p}(z_1|y_1) \right) \left( \sum_{y_2 \in Y} p(y_2, s_2|x_2) \hat{p}(z_2|y_2) \right) \left( \sum_{y_3 \in Y} p(y_3, s_3|x_3) \hat{p}(z_3|y_3) \right)$$

$$= \sum_{y_1 \in Y} \sum_{y_2 \in Y} \sum_{y_3 \in Y} \left( p(y_1, s_1|x_1) \hat{p}(z_1|y_1) p(y_2, s_2|x_2) \hat{p}(z_2|y_2) p(y_3, s_3|x_3) \hat{p}(z_3|y_3) \right) .$$

D. The Stochastically Degraded FSBC: An Example

We now give an example of an actual scenario represented by Definition 8. Consider a scenario in which a base station transmits to two mobile units, located approximately on the same line-of-sight from
the base station (BS), as indicated by the dashed line in Figure 2. Let the BS transmit a BPSK signal and let the received signals be subject to additive Gaussian noise. Let decoding the messages at the receivers take place after hard-decision decoding of the channel symbols with a threshold at zero. 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–Rx₁ have a crossover probability \( \epsilon_1(A) = 0.1 \) and the channel BS–Rx₂ 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

\[
\epsilon_{12}(A) = \frac{\epsilon_2 - \epsilon_1}{1 - 2\epsilon_1} = 0.0625.
\]

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 \( \epsilon_{12}(B) = 0.0625 \). Hence, the degrading channel is the same for both states, irrespective of the state sequence. This satisfies condition (9).

Note that in this example the state sequence represents the traffic pattern, thus the state transitions reflect the average length of a vehicle and the average time 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. More generally, we can define a set of states for this scenario, e.g. \( S = \{1, 2, ..., K\} \), with \( Y = Z = \{0, 1\} \) and

\[
p(z, s_i|y_i, s_{i-1}) = p(s_i|s_{i-1})p(z_i|y_i, s_i)
\]

\[
p(z|y, s = k) = \begin{cases} 
\epsilon_{12}(k), & z = 1 - y \\
1 - \epsilon_{12}(k), & z = y 
\end{cases}
\]

\( \epsilon_{12}(k) \in (0, 0.5), k = 1, 2, ..., K \). This results in a collection of physically degraded BSBCs that can give more flexibility in modeling the scenario of Figure 2, as the degrading channel may now depend on the state. However, for this reason, this model does not satisfy our definition of stochastic degradedness in Definition 8.

\(^1\)The scenario parameters assumed in this example are: Two-ray propagation model, Rx decoding scheme is maximum-likelihood, 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–Rx₁ distance = 7.2 Km and BS–Rx₂ distance = 8 Km. We also assume a passing car increases the path attenuation by 3 dB.
E. The Discrete Broadcast Channel with Finite Memory

Definition 1 is the most general definition for a discrete broadcast channel with memory varying over a finite state space. It should be noted, however, that channel memory does not arise only from correlated changes in the propagation medium, such as correlated fading. Channel memory is also introduced by components in the transmitter and the receivers that have memory. In the transmitter these are the transmit filters (channel mask, IF and RF filters). In the receivers, memory is introduced by the loops, such as AGC and timing, and in particular the equalizer, all of which operate on the channel outputs [29]. Therefore, the channel has memory even when the medium is fixed (e.g. as in microwave communication in the LMDS frequency range [30]). The situation in which memory is only due to components in the transmit-receive chain can be modeled by the BC with finite memory. The point-to-point finite-memory channel was first considered in [4]. We now define the discrete, finite-memory broadcast channel:

Definition 9. The discrete broadcast channel with finite memory of length $K$ is defined by the triplet 
$$\{X^{K+1} \times Y^K \times Z^K, p(y_{K+1}, z_{K+1} | x^{K+1}, y^K, z^K), Y \times Z\}.$$ 
For a sequence of $n$ transmissions we have
$$p(y^n, z^n | x_0^n, x_{-K+1}^0, y_{-K+1}^0, z_{-K+1}^0) = \prod_{i=1}^n p(y_i, z_i | x_{i-K+1}^n, y_{i-K+1}^{i-1}, z_{i-K+1}^{i-1})$$
$$= \prod_{i=1}^n p(y_i, z_i | x_i^{i-K}, y_{i-K}^{i-1}, z_{i-K}^{i-1}),$$
where (a) follows from (2). Note that $(x^{0}_{-K+1}, y^{0}_{-K+1}, z^{0}_{-K+1})$ constitutes the initial state, but this state is generally not known at the transmitter and receivers at the beginning of the transmission, since this requires
the receivers to send their outputs to each other and to the transmitter. Defining \( S = Y^K \times Z^K \times X^K \) and \( s_{i-1} = (y_{i-K}, z_{i-K}, x_{i-K}) \), we see that this channel is a special case of the FSBC. Therefore, the results for this scenario can be obtained from the FSBC results. Since the channel memory is finite, this channel is indecomposable [4].

### III. Capacity Theorems for the Degraded FSBC

In this section we present the main capacity results for both the physically and stochastically degraded FSBC:

#### A. The Physically Degraded FSBC

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

\[
\| \times_{i=1}^n \mathcal{U}_i \| \leq \min \{ \| \mathcal{X} \|, \| \mathcal{Y} \|, \| \mathcal{Z} \| \}^n. \tag{12}
\]

Define the region \( \mathcal{R}_n \) as

\[
\mathcal{R}_n = \bigcup_{q_n \in \mathcal{Q}_n} \left\{ (R_0, R_1, R_2) : R_0 \geq 0, R_1 \geq 0, R_2 \geq 0, \right. \\
R_1 \leq \min_{s_0 \in \mathcal{S}_n} \frac{1}{n} I(X^n; Y^n|U^n, s_0^{n}), \\
R_0 + R_2 \leq \min_{s_0 \in \mathcal{S}_n} \frac{1}{n} I(U^n; Z^n|s_0^{n}) \bigg\}.
\]

The main result is stated in the following theorem:

**Theorem 1.** For the physically degraded FSBC \( \{ \mathcal{X} \times \mathcal{S}, p(y, z, s|x, s'), \mathcal{Y} \times \mathcal{Z} \times \mathcal{S} \} \) defined in Definition 7, the capacity region of the rate triplets \( (R_0, R_1, R_2) \) is given by

\[
C_{pd} = \lim_{n \to \infty} \mathcal{R}_n,
\]

and the limit exists\(^2\).

**Proof Outline:** The converse proof is based on manipulating Fano’s inequality. The details are given in Appendix B. The detailed achievability proof is deferred to Appendix C; we here provide a basic sketch of this proof to highlight its key elements:

\(^2\)See [23, Chapter 2] for the definition of the limit of a sequence of sets.
1) Fix \( n \) and a joint probability distribution \( p(u^n, x^n) \). Generate a superposition codebook with \( u^n \) as the cloud centers, indexed by the messages in \( M_2 \), and \( x^n \) as the cloud elements indexed by the message pairs in \( M_1 \times M_2 \).

2) Using maximum-likelihood decoding at Rx_2 according to

\[
\arg \max_{\hat{m}_2} \frac{1}{|S|} \sum_{s_0 \in S} p(z^n | u^n(\hat{m}_2), s_0)
\]

we show that a positive error exponent for decoding \( M_2 \) can be obtained as long as \( R_2 \leq \min_{s_0 \in S} \frac{1}{n} I(U^n; Z^n | s_0) - \frac{\log_2 |S|}{n} \triangleq R_{2,n}(p) \). 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 also for decoding \( M_2 \) at Rx_1 we can achieve a positive error exponent.

3) Using maximum-likelihood decoding at Rx_1 according to

\[
\arg \max_{\hat{m}_1} \frac{1}{|S|} \sum_{s_0 \in S} p(y^n | x^n(\hat{m}_1, m_2), s_0)
\]

we show that, given that \( M_2 \) was correctly decoded at Rx_1, a positive error exponent for decoding \( M_1 \) at Rx_1 can be achieved as long as \( R_1 \leq \min_{s_0 \in S} \frac{1}{n} I(X^n; Y^n | U^n, s_0') - \frac{\log_2 |S|}{n} \triangleq R_{1,n}(p) \).

4) Next, for a fixed \( n \) and some integer \( b \), we let \( n_0 = nb \). As in [15], we construct a distribution for \( (U^{n_0}, X^{n_0}) \) by taking the product of the basic distribution for a block of \( n \) symbols \( b \) times:

\[
p(u^{n_0}, x^{n_0}) = \prod_{b'=1}^{b} p(u^{b'n}_{(b'-1)n+1}^{b'n}x^{b'n}_{(b'-1)n+1}).
\]

For blocklength \( n_0 \) we generate a superposition codebook according to \( p(u^{n_0}, x^{n_0}) \). Now, we show that taking \( b \) large enough results in an 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 \( F_n(\lambda) \):

\[
C^n(\lambda) = \max_{p(u^n, x^n)} \left\{ \min_{s_0 \in S} \frac{1}{n} I(U^n; Z^n | s_0) + \lambda \min_{s_0' \in S} \frac{1}{n} I(X^n; Y^n | U^n, s_0') \right\}
\]

(14)

\[
F_n(\lambda) = C^n(\lambda) - (1 + \lambda) \frac{\log_2 |S|}{n}.
\]

(15)

The maximization in (14) is carried out subject to the cardinality bound on \( \times_{i=1}^n U_i \) given in (12). We show that \( F_n(\lambda) \) is sup-additive, hence

\[
C^\infty(\lambda) \triangleq \lim_{n \to \infty} C^n(\lambda) = \lim_{n \to \infty} F_n(\lambda) = \sup_n F_n(\lambda).
\]

(16)

Therefore, the boundary of the achievable region can be written as

\[
R_2(R_1) = \inf_{0 \leq \lambda \leq 1} (C^\infty(\lambda) - \lambda R_1).
\]

(17)
Comment 1. Note that we use $\times_{i=1}^{n} U_i$ and not $U^n$ 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 $U$, but the product $U_1 \times U_2 \times \ldots \times U_n$.

B. The Stochastically Degraded FSBC

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 [21, Chapter 14.6]) then the capacity region of the stochastically degraded FSBC is the same as the corresponding physically degraded FSBC:

Corollary 1. For the stochastically degraded FSBC of Definition 8, the capacity region is given by Theorem 1 where $p(y, z, s|x, s')$ is replaced by $p(y, s|x, s')\tilde{p}(z|y)$, such that Equation (10) is satisfied.

C. The Indecomposable FSBC

When the FSBC is indecomposable, the effect of the initial state becomes negligible as $n$ increases. Therefore, the maximum of each of the mutual information expressions over all $s_0 \in S$ asymptotically equals the minimum over all $s_0$. Hence, for all the initial states, the limits of the mutual information expressions as $n$ approaches infinity are the same. Before stating the result, let us define

$$\hat{R}_n = \sup_{q_n \in \mathcal{Q}_n} \left\{ (R_0, R_1, R_2) : R_0 \geq 0, R_1 \geq 0, R_2 \geq 0, \right. \left. R_1 \leq \frac{1}{n} I(X^n; Y^n|U^n)_{q_n}, \right. \left. R_0 + R_2 \leq \frac{1}{n} I(U^n; Z^n)_{q_n} \right\}. $$

The rate regions for this scenario are now characterized in the following theorem:

Theorem 2.

1) For the indecomposable physically degraded FSBC the capacity region is given by

$$\mathcal{C}_{pd}^{indecomp.} = \lim_{n \to \infty} \hat{R}_n, \quad (18)$$

and the limits exist.

2) For the indecomposable stochastically degraded FSBC, the capacity region is obtained from Equation (18), where $p(y, z, s|x, s')$ is replaced by $p(y, s|x, s')\tilde{p}(z|y)$, such that Equation (10) is satisfied.

Proof: See Appendix E.
D. Discussion

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

1) From the derivation of the capacity region for the physically degraded FSBC it follows that the capacity region \( C_{pd} \) is given as the limit of the intersection over \( S \) of regions defined for finite \( n \). This can be seen by letting

\[
\mathcal{R}_n(s_0) = \alpha \bigcup_{q_n \in Q_n} \left\{ (R_0, R_1, R_2) : R_0 \geq 0, R_1 \geq 0, R_2 \geq 0, \right. \\
R_1 \leq \frac{1}{n} I(X^n; Y^n | U^n, s_0)_{q_n}, \\
R_0 + R_2 \leq \frac{1}{n} I(U^n; Z^n | s_0)_{q_n} \big\}
\]

Then,

\[
\mathcal{R}_n = \bigcap_{s_0 \in S} \mathcal{R}_n(s_0)
\]  \hspace{1cm} (20)

and \( C_{pd} = \lim_{n \to \infty} \bigcap_{s_0 \in S} \mathcal{R}_n(s_0) \). It is natural to ask if the capacity region can also be written as the intersection over \( S \) of the limit regions for all \( s_0 \in S \). If this holds then the capacity region for the physically degraded FSBC with initial state unknown to the receivers and transmitter can be interpreted as the intersection of the capacity regions when the initial state is known at the receivers but not at the transmitter. Note that the fact that \( \lim_{n \to \infty} \bigcap_{s_0 \in S} \mathcal{R}_n(s_0) \) exists does not imply that for each \( s_0 \in S \), \( \lim_{n \to \infty} \mathcal{R}_n(s_0) \) exists. However, if all the limits \( \{ \lim_{n \to \infty} \mathcal{R}_n(s_0) \}_{s_0 \in S} \) exist, we have the following proposition:

**Proposition 1.** If all the limits \( \{ \lim_{n \to \infty} \mathcal{R}_n(s_0) \}_{s_0 \in S} \) exist, the capacity region \( C_{pd} \) defined in (13) can be written as

\[
C_{pd} = \lim_{n \to \infty} \bigcap_{s_0 \in S} \mathcal{R}_n(s_0) = \bigcap_{s_0 \in S} \lim_{n \to \infty} \mathcal{R}_n(s_0). \]  \hspace{1cm} (21)

**Proof:** Assume first

\[
(R_0, R_1, R_2) \triangleq R \in \lim_{n \to \infty} \bigcap_{s_0 \in S} \mathcal{R}_n(s_0).
\]

This implies that for all \( \epsilon > 0 \) there exists \( N(\epsilon) \) such that for all \( n > N(\epsilon) \), and for all \( s_0 \in S \), \( R - \epsilon \in \mathcal{R}_n(s_0) \). Therefore, for every \( s_0 \), \( R - \epsilon \in \lim_{n \to \infty} \mathcal{R}_n(s_0) \) and as \( S \) is finite we have \( \lim_{n \to \infty} \bigcap_{s_0 \in S} \mathcal{R}_n(s_0) \subseteq \bigcap_{s_0 \in S} \lim_{n \to \infty} \mathcal{R}_n(s_0) \). Alternatively, \( R \in \bigcap_{s_0 \in S} \lim_{n \to \infty} \mathcal{R}_n(s_0) \) implies that for every \( \epsilon > 0 \) and for every \( s_0 \in S \) there exists \( N(\epsilon, s_0) \) such that for all \( n > N(\epsilon, s_0) \) we have that \( R - \epsilon \in \mathcal{R}_n(s_0) \). Since \( S \) is finite, then taking \( N(\epsilon) = \max_{s_0 \in S} N(\epsilon, s_0) \), we obtain that
for all $n > N(\epsilon)$, $\mathbf{R} - \epsilon \in \bigcap_{s_0 \in S} \mathcal{R}_n(s_0)$ hence $\lim_{n \to \infty} \bigcap_{s_0 \in S} \mathcal{R}_n(s_0) \supseteq \bigcap_{s_0 \in S} \lim_{n \to \infty} \mathcal{R}_n(s_0)$, implying these sets are equal.

Thus, the capacity region can be viewed as the intersection of all the capacity regions obtained when the initial state is known at the receivers (but not at the transmitter). This is illustrated in Figure 3.

![Achievable Region Diagram](image)

Fig. 3. Achievable region obtained as an intersection of achievable regions with initial state known only at the receivers.

2) Since the limit defining the region $\mathcal{C}_{pd}$ exists, then as $n$ increases, the achievable region increases as well. Although this point seems trivial (and in the discrete, memoryless case it always holds), it is important to note that when the channel has memory this property is not guaranteed when the limits cannot be shown to exist\(^3\). Thus, for the FSBC, larger blocklengths yield a larger achievable rate region. Since a larger $n$ corresponds to a higher complexity code, in practical designs the rate increase associated with increasing $n$ can be weighed against the resulting complexity cost.

3) 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.

4) As noted in the discussion following Definition 5, for indecomposable channels both definitions of the achievable rate lead to the same capacity region. This happens because when the channel is indecomposable, the effect of the initial channel state becomes negligible over time.

\(^3\)Consider for example a non-stationary channel with noise that oscillates with time.
5) For the discrete finite-memory BC each receiver can send its last $K$ channel outputs to the other receiver and to the transmitter, so that at the next message interval the initial state will be known to all. However, as this channel is indecomposable it is not necessary, in the limit of large blocklength, to design the encoder and decoders based on knowledge of the initial state to achieve the maximum performance.

6) 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. Moreover, letting $n = m_1 + m_2$, from Equation (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 $p(x^{m_1}|u^n) \neq p(x^{m_1}|u^{m_1})$ we have 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 the FSBC and the point-to-point and MAC channels. Consider, for example, the expression

$$\max_{p(u^n, x^n)} \left\{ \max_{s_0 \in S} \frac{1}{n} I(U^n; Z^n|s_0) + \lambda \max_{s'_0 \in S} \frac{1}{n} I(X^n; Y^n|U^n, s'_0) \right\},$$

which serves as an upper bound on the achievable region. While in the point-to-point channel the corresponding upper bound converges for all channels (see [5, Theorem 4.6.1]), for the FSBC (22) can be shown to converge only for 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 in contrast also to the DMBC.

IV. CONCLUSIONS

We have defined the degraded finite-state broadcast channel as well as its capacity region for both indecomposable and non-indecomposable channels. Specifically, we first defined the notion of degradedness for the broadcast channel with memory. Then, we considered two possible definitions for error probability based on either worst-case (relative to initial state) or averaged (relative to the initial state distribution) error. Average error probability is best suited to indecomposable channels since the effect of the initial channel state becomes negligible over time, whereas the worst case error probability is more appropriate for non-indecomposable broadcast channels. When the channel is indecomposable, in fact,
both definitions lead to the same achievable regions. 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.

We derived the capacity region of the general degraded FSBC and of the indecomposable degraded
FSBC 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, instead of generating a codeword by independently selecting sym-
bols. A key element is verifying that in the limit as the blocklength \( n \) increases to infinity, the achievable
region converges. Convergence is crucial 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 the point-to-point FSC feedback helps [9], then also for the degraded FSBC feedback
can increase the capacity region.

APPENDIX A

AUXILIARY LEMMA

**Lemma A.1** ([5, Lemma 1 in Appendix 4A]). Let \( X, Y, Z, S \) be a joint ensemble. If \( S \) has a finite
cardinality then

\[
|I(X; Y | Z, S) - I(X; Y | Z)| \leq \log_2 ||S||.
\]

This lemma implies that the RVs \( A^n, B^n, C^n, S \) satisfy

\[
I(A^n; B^n | C^n) \geq I(A^n; B^n | C^n, S) - \log_2 ||S|| \tag{A.1}
\]

\[
I(A^n; B^n | C^n, S) \geq I(A^n; B^n | C^n) - \log_2 ||S||. \tag{A.2}
\]

APPENDIX B

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 \( R \) into private and common rates, as in [21,
Theorem 14.6.4]. The converse is stated in the following lemma:
Lemma B.1. If for some $\lambda > 0$,

$$R_2 + \lambda R_1 > C^\infty(\lambda) + \epsilon,$$  

(B.1)

where $C^\infty(\lambda)$ is defined in (16), then there exist initial states $\tilde{s}_0, \tilde{s}_0' \in \mathcal{S}$ for which

$$P^{(n)}_{e_2}(\tilde{s}_0)R_2 + \lambda P^{(n)}_{e_1}(\tilde{s}_0')R_1 > \epsilon - \frac{1}{n}(1 + \lambda)(1 + \log_2 ||\mathcal{S}||).$$  

(B.2)

The implication of (B.2), as explained in [16], is that for large values of $n$, $\max_{s_0 \in \mathcal{S}} P^{(n)}_{e}(s_0)$ cannot be made arbitrarily small, outside the region whose boundary is given in (17).

Proof: Recall that $P^{(n)}_{e_2}(s_0)$ and $P^{(n)}_{e_1}(s_0)$ 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, Equation 4.6.16]) we have that for any given initial state $s_0$

$$H(M_2|Z^n, s_0) \leq P^{(n)}_{e_2}(s_0)nR_2 + 1$$

(B.3a)

$$H(M_1|Y^n, s_0') \leq P^{(n)}_{e_1}(s_0')nR_1 + 1.$$  

(B.3b)

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

$$\min_{s_0 \in \mathcal{S}} I(M_2; Z^n|s_0) = \min_{s_0 \in \mathcal{S}} \{H(M_2|s_0) - H(M_2|Z^n, s_0)\}$$

$$= nR_2 - \max_{s_0 \in \mathcal{S}} H(M_2|Z^n, s_0)$$

(B.4)

$$\min_{s_0' \in \mathcal{S}} I(M_1; Y^n|M_2, s_0') = \min_{s_0' \in \mathcal{S}} \{H(M_1|M_2, s_0') - H(M_1|Y^n, M_2, s_0')\}$$

$$= nR_1 - \max_{s_0' \in \mathcal{S}} H(M_1|Y^n, M_2, s_0')$$

$$\geq nR_1 - \max_{s_0 \in \mathcal{S}} H(M_1|Y^n, s_0').$$  

(B.5)

Next, we show that

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

(B.6)

First note that

$$I(M_2; Z^n|s_0) = H(Z^n|s_0) - H(Z^n|M_2, s_0)$$

$$= H(Z^n|s_0) - H(Z^n|U^n, s_0)$$

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

April 30, 2008

DRAFT
where $U_i = M_2$, $i = 1, 2, \ldots, n$. We also have that
\[
I(M_1; Y^n | M_2, s'_0) = H(Y^n | M_2, s'_0) - H(Y^n | M_1, M_2, s'_0)
\]
\[
\leq H(Y^n | M_2, s'_0) - H(Y^n | X^n, M_1, M_2, s'_0)
\]
\[
= H(Y^n | U^n, s'_0) - H(Y^n | X^n, U^n, s'_0)
\]
\[
= I(X^n; Y^n | U^n, s'_0),
\]
where our definition of $U^n$ satisfies the Markov relationship $U^n | s'_0 \leftrightarrow X^n | s'_0 \leftrightarrow Y^n | s'_0$. Combining both derivations we have that for our choice of $U^n$:
\[
\min_{s_0 \in S} I(M_2; Z^n | s_0) + \lambda \min_{s_0' \in S} I(M_1; Y^n | M_2, s'_0)
\]
\[
\leq \min_{s_0 \in S} I(U^n; Z^n | s_0) + \lambda \min_{s_0' \in S} I(X^n; Y^n | U^n, s'_0)
\]
\[
\quad \quad \quad \leq nC^\infty(\lambda), \quad \text{(a)}
\]
where (a) is because $C^\infty(\lambda)$ is obtained by maximizing over all joint distributions $p(u^n, x^n)$ subject to the cardinality constraint (12) (we show in Appendix C-H that this is enough to achieve the maximum), hence (B.6) is verified. Therefore,
\[
\min_{s_0 \in S} I(M_2; Z^n | s_0) + \lambda \min_{s_0' \in S} I(M_1; Y^n | M_2, s'_0) - (1 + \lambda) \log_2 ||S|| \leq n \left( C^\infty(\lambda) - (1 + \lambda) \frac{\log_2 ||S||}{n} \right)
\]
\[
\leq nC^\infty(\lambda), \quad \text{(B.7)}
\]
where (B.7) is shown in Appendix D.

Plugging (B.4) and (B.5) into (B.7) yields
\[
nR_2 - H(M_2 | Z^n, s_{0,n}) + \lambda(nR_1 - H(M_1 | Y^n, s'_{0,n})) - (1 + \lambda) \log_2 ||S|| \leq nC^\infty(\lambda)
\]
\[
\Rightarrow H(M_2 | Z^n, s_{0,n}) + \lambda H(M_1 | Y^n, s'_{0,n}) + (1 + \lambda) \log_2 ||S|| \geq n \left( R_2 + \lambda R_1 - C^\infty(\lambda) \right)
\]
\[
\Rightarrow \mathcal{H}(M_2 | Z^n, s_{0,n}) + \lambda \mathcal{H}(M_1 | Y^n, s'_{0,n}) > n \left( \epsilon - (1 + \lambda) \frac{\log_2 ||S||}{n} \right),
\]
where the last line follows from (B.1). Combined with Fano’s inequalities (B.3), we have
\[
\mathcal{P}_{e_2}^{(n)}(s_{0,n}) nR_2 + \lambda \left( \mathcal{P}_{e_1}^{(n)}(s'_{0,n}) nR_1 + 1 \right) > n\epsilon - (1 + \lambda) \log_2 ||S||
\]
which means that for large values of $n$, at least one of the states $s_{0,n}$, $s'_{0,n}$ results in a probability of error (at the respective receiver) that is bounded away from zero, thus completing the proof.

April 30, 2008
APPENDIX C

ACHIEVABILITY OF THE RATES OF THEOREM 1

We prove that all the rates of Theorem 1 are achievable for the physically degraded FSBC. The derivation requires that condition (7) holds, which was shown in Section II-C to be true when the channel is physically degraded. In the derivation we shall consider only the two private messages case since the common message can be incorporated by splitting the rate to Rx2 into private and common rates, as in [21, Theorem 14.6.4]. Recall that the transmitter and receivers do not know the channel states.

A. Proof Outline

The main steps of the proof are as follows:

1) For a given $n$, generate a superposition codebook using $p(u^n, x^n)$, where $u^n$ are the cloud centers indexed by the messages in $M_2$ and $x^n$ are the cloud elements indexed by the message pairs in $M_1 \times M_2$.

2) Rx$_2$, the weak receiver, uses maximum-likelihood (ML) decoding to decode the message $M_2$ from its channel output $Z^n$. This gives rise to a rate bound on $R_2$:

$$R_2 \leq \min_{s_0 \in \mathcal{S}} \frac{1}{n} I(U^n; Z^n | s_0) - \frac{\log_2 |\mathcal{S}|}{n} \triangleq R_{2,n}(p). \quad (C.1)$$

3) Rx$_1$, the strong receiver, uses ML decoding to decode the message $M_2$ from its channel output $Y^n$. This gives rise to a second rate constraint on $R_2$. We show that due to degradedness, this bound is looser than (C.1). Next, Rx$_1$ proceeds to decode $M_1$ using $Y^n$ and the decoded $M_2$. This results in a rate constraint on $R_1$:

$$R_1 \leq \min_{s_0 \in \mathcal{S}} \frac{1}{n} I(X^n, Y^n; U^n | s_0) - \frac{\log_2 |\mathcal{S}|}{n} \triangleq R_{1,n}(p). \quad (C.2)$$

4) Using the product extension of the basic distribution $p(u^n, x^n)$ to blocklength $b = Kn$, $K \in \mathbb{N}$, we show that the rate pair $(R_{1,n}(p), R_{2,n}(p))$ defined in (C.1) and (C.2), is achievable.

5) We next show that the achievable region is sup-additive, implying that taking $n \to \infty$ results in the largest achievable region.

6) We derive a bound on the cardinality of the auxiliary RV $U^n$, thus asserting that the optimization problem for finding the maximum rate pairs has a finite-dimension domain for any given $n$. 

April 30, 2008 DRAFT
Before giving the details of the proof we write the average probability of error when the initial channel state is \( s_0 \) as:

\[
\begin{align*}
P_e^{(n)}(s_0) & \leq P_e^{(n)}(M_1|s_0) + P_e^{(n)}(M_2|s_0) \\
& \leq P_{e_1}^{(n)}(M_1, M_2|s_0) + P_{e_2}^{(n)}(M_2|s_0) \\
& = P_{e_{12}}^{(n)}(M_2|s_0) + P_{e_{11}}^{(n)}(M_1|M_2, s_0) + P_{e_2}^{(n)}(M_2|s_0) \triangleq \bar{P}_e^{(n)}(s_0),
\end{align*}
\]

where \( P_{e_i}^{(n)}(M_i|s_0), i = 1, 2 \) is the average probability of error for decoding \( M_i \) at \( \text{Rx}_i \) when the initial state is \( s_0 \), and \( P_{e_{11}}^{(n)}(M_1|M_2, s_0) \triangleq \text{Pr}( \text{decoding error for } M_1|M_2 \text{ decoded correctly at } \text{Rx}_1, s_0) \). In the following sections we provide bounds on the probabilities in (C.3).

B. Code Construction and Encoding

- Fix a blocklength \( n \) and a joint distribution \( p(u^n, x^n) \). For each message \( m_2 \in M_2 \), independently generate a codeword according to \( \text{Pr}(u(m_2)) = p(u^n) \). Next, for each message \( m_2 \in M_2 \) generate a codebook with \( 2^{nR_2} \) codewords \( x(m_1, m_2) \), \( m_1 \in M_1 \) where each codeword is generated according to the probability \( \text{Pr}(x(m_1, m_2)) = p(x^n|u^n(m_2)) \), and is selected independently of the codewords generated for the other \( m_1 \) messages.

- For transmitting the message pair \( (m_1, m_2) \) the transmitter outputs \( x(m_1, m_2) \).

C. Decoding at the Weak Receiver \( \text{Rx}_2 \)

Since the messages \( m_2 \) are selected uniformly, the decoder that minimizes the probability of error is the maximum likelihood decoder [25]. A natural choice of a decoding rule for the BC with memory is to extend the decoding rule used in the DMBC to the scenario with memory. As in [5, Section 5.9], the decoder overcomes its ignorance of the initial state by averaging over all initial channel states \( s_0 \in S \). The derivation follows the essential steps of [5, Section 5.9].

First, define

\[
\begin{align*}
\tilde{p}(z^n|u^n) & \triangleq \sum_{s_0 \in S} \frac{1}{||S||} p(z^n|u^n, s_0) \\
& = \sum_{s_0 \in S} \frac{1}{||S||} \sum_{y^n,x^n,s^n} p(z^n, y^n, x^n, s^n|u^n, s_0) \\
& = \sum_{s_0 \in S} \frac{1}{||S||} \sum_{y^n,x^n,s^n} p(x^n|u^n)p(z^n, y^n, s^n|x^n, s_0).
\end{align*}
\]
Let the decoding rule be the ML decoder according to $\hat{p}(z^n|u^n)$. Then, for a received sequence $z$,

$$\text{if } \forall m_2' \in M_2, m_2' \neq \hat{m}_2, \quad \hat{p}(z|u(\hat{m}_2)) \geq \hat{p}(z|u(m_2')) \quad \Rightarrow g_z(z) = \hat{m}_2, $$

with ties broken arbitrarily. The probability of error when message $m_2$ is transmitted using $u(m_2)$ and the initial state is $s_0$ is given by

$$P_e^{(n)}(m_2|u(m_2), s_0) = \sum_{z: g_z(z) \neq m_2} \sum_{Y^n, X^n, S^n} p(z, y, x, s|u(m_2), s_0),$$

see (C.4). Note that this probability, for a given $u(m_2)$ and a given $z$, is the average over all possible codewords $x$, state sequences $s$ and received sequences at the strong receiver $Rx_1, y$. Proceeding as in [5], we consider the error event

$$E_{m_2'} \triangleq \{ m_2' \text{ is decoded when } m_2 \text{ is transmitted using codeword } u(m_2), \text{ and } z \text{ is received at } Rx_2 \}. $$

The probability of $E_{m_2'}$ averaged over all possible selections of $u(m_2')$ is\footnote{As $z$ is given, the initial state does not affect $Pr(E_{m_2'})$.}

$$\Pr(E_{m_2'}) = \Pr\left(m_2' \text{ decoded } | m_2, u(m_2), z\right) = \sum_{u(m_2'): \hat{p}(z|u(m_2')) \geq \hat{p}(z|u(m_2))} p(u(m_2')) \leq \sum_{u(m_2') \in X^n U_i} p(u(m_2')) \frac{\hat{p}(z|u(m_2'))^s}{\hat{p}(z|u(m_2))}^s,$$

(C.5)

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

Now, the probability of error when $m_2$ is transmitted using $u(m_2)$ and $z$ is received is

$$\Pr\left(\text{error } | m_2, u(m_2), z\right) = \Pr\left( \bigcup_{m_2' \neq m_2} E_{m_2'} \right)$$

\[
\leq \sum_{m_2' \neq m_2} \Pr\left(E_{m_2'}\right) \leq \sum_{m_2' \neq m_2} \sum_{u \in X^n U_i} p(u) \frac{\hat{p}(z|u)^s}{\hat{p}(z|u(m_2))}^s = \left(\left|\left| M_2\right|\right| - 1\right) \sum_{u \in X^n U_i} p(u) \frac{\left(\sum_{s_0 \in S, |S|} p(z|u, s_0)\right)^s}{\left(\sum_{s_0 \in S, |S|} p(z|u(m_2), s_0)\right)^s} \right)^\rho, \quad (C.6)
\]
where in (a) \( 0 < \rho \leq 1 \) and (b) follows from the bound (C.5) by noting that \( u(m'_2) \) functions as a dummy variable of summation.

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

\[
P_{e2}^{(n)}(m_2|s_0) = \sum_{u(m_2) \in \mathcal{X}_i, i = 1} p(u(m_2)) \sum_{Z^n} p(z|u(m_2), s_0) \Pr(\text{error} | m_2, u(m_2), z)
\]

\[
\leq ||S|| \sum_{Z^n} \sum_{u(m_2) \in \mathcal{X}_i, i = 1} p(u(m_2)) \left( \sum_{s_0 \in S} \frac{1}{||S||} p(z|u(m_2), s_0) \right) \Pr(\text{error} | m_2, u(m_2), z),
\]

as summing over all initial states \( s_0 \) results in an upper bound. Using the bound (C.6) in this expression, we can upper bound the average probability of error for the message \( m_2 \) for any \( s_0 \) by

\[
P_{e2}^{(n)}(m_2|s_0) \leq ||S|| \sum_{Z^n} \sum_{u(m_2) \in \mathcal{X}_i, i = 1} p(u(m_2)) \bar{p}(z|u(m_2)) \times
\]

\[
\left[ (||M_2|| - 1) \sum_{u \in \mathcal{X}_i, i = 1} p(u) \frac{\bar{p}(z|u)}{\bar{p}(z|u(m_2))} \right]^\rho
\]

\[
\leq ||S|| (||M_2|| - 1)^\rho \sum_{Z^n} \sum_{x_i \in \mathcal{X}_i, i = 1} p(u) \left( \sum_{s_0 \in S} \frac{1}{||S||} p(z|u, s_0) \right)^\frac{1}{1+\rho}
\]

\[
\leq ||S|| (||M_2|| - 1)^\rho ||S||^\rho \sum_{Z^n} \sum_{x_i \in \mathcal{X}_i, i = 1} p(u) p(z|u, s_0)^\frac{1}{1+\rho},
\]

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

Next, define

\[
E_{n,2}(\rho, p(u, x), s_0) = -\frac{1}{n} \log_2 \sum_{Z^n} \left[ \sum_{x_i \in \mathcal{X}_i, i = 1} p(u) p(z|u, s_0) \right]^{1+\rho}
\]

\[
F_{n,2}(\rho, p(u, x)) \triangleq \min_{s_0 \in S} E_{n,2}(\rho, p(u, x), s_0) - \rho \frac{\log_2 ||S||}{n},
\]

\( 0 \leq \rho \leq 1 \). Then, from (C.8), the probability of error averaged over all codebooks generated according to the distribution \( p(u^n, x^n) \) is given by

\[
P_{e2}^{(n)}(m_2|s_0) \leq ||S|| 2^{-n(F_{n,2}(\rho, p(u, x)) - \rho R_2)}.
\]

\(^5\) We use the following lemma from [5, Section 5.6]: Let \( p(A_1), p(A_2), \ldots, p(A_M) \) be the probabilities of a set of events and \( p \left( \bigcup_{m=1}^M A_m \right) \) be the probability of their union. For any \( \rho \), \( 0 < \rho \leq 1 \), \( p \left( \bigcup_{m=1}^M A_m \right) \leq \left[ \sum_{m=1}^M p(A_m) \right]^\rho \).
The next step is to show that for a fixed initial state $s_0$, $\exists \rho, 0 < \rho \leq 1$ for which $E_{n,2}(\rho, p(u, x), s_0) - \rho \frac{\log_2 ||S||}{n} - \rho R_2$ is positive as long as

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

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

$$\frac{\partial}{\partial \rho} \left( E_{n,2}(\rho, p(u, x), s_0) - \rho \frac{\log_2 ||S||}{n} - \rho R_2 \right) = \frac{\partial}{\partial \rho} E_{n,2}(\rho, p(u, x), s_0) - \frac{\log_2 ||S||}{n} - R_2.$$

Noting that $p(u) = p(u|s_0)$ since $u$ is selected independently of the initial state, we arrive at

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

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

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

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

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

We give the details of this argument when considering the rate bound on $R_1$ in Appendix C-E, since that case is more general.

Thus we obtain that for a given $n$, if

$$R_2 \leq \min_{s_0 \in S} \frac{1}{n} I(U^n; Z^n|s_0) - \frac{\log_2 ||S||}{n} \triangleq R_{2,n}(p), \quad (C.10)$$

there exists $\rho^* > 0$ for which $E_{n,2}(\rho^*, p(u, x)) - \rho^* R_2$ is positive, hence $\tilde{F}_{e2}^{(n)}(m_2|s_0)$ in (C.9) has a positive error exponent. As a last comment, note that the expression in (C.10) can be negative. However, in Appendix D 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 non-zero, then there exists 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.
D. Decoding the Message $m_2$ at the Strong Receiver $R_{x_1}$

Define
\[
\hat{p}(y^n|u^n) = \sum_{s_0 \in \mathcal{S}} \frac{1}{||\mathcal{S}||} p(y^n|u^n, s_0)
\]
\[
= \sum_{s_0 \in \mathcal{S}} \frac{1}{||\mathcal{S}||} \sum_{z^n, x^n, s^n} p(z^n, y^n, x^n, s^n|u^n, s_0)
\]
\[
= \sum_{s_0 \in \mathcal{S}} \frac{1}{||\mathcal{S}||} \sum_{z^n, x^n, s^n} p(x^n|u^n)p(z^n, y^n, s^n|x^n, s_0).
\]

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

if $\forall m_2 \in \mathcal{M}_2, m_2 \neq \hat{m}_2$, $\hat{p}(y|u(\hat{m}_2)) \geq \hat{p}(y|u(m_2')) \Rightarrow$ $R_{x_1}$ decides on $\hat{m}_2$,

with ties broken arbitrarily. The average probability of error when $m_2$ is transmitted using $u(m_2)$ and $s_0$ is the initial state is given by
\[
P_{e_{12}}^{(n)}(m_2|u(m_2), s_0) = \sum_{y: \exists m_2 \neq m_2', \hat{p}(y|u(m_2')) \geq \hat{p}(y|u(m_2))} \sum_{z^n, x^n, s^n} p(z, y, x, s|u(m_2), s_0).
\]

Applying the same steps used in the derivation of the probability of error at $R_{x_2}$ to this case, we conclude that the error exponent for decoding the message $m_2$ at $R_{x_1}$ can be made positive for every $s_0 \in \mathcal{S}$ as long as
\[
R_2 \leq \min_{s_0 \in \mathcal{S}} \frac{1}{n} I(U^n; Y^n|s_0) - \frac{\log_2 ||\mathcal{S}||}{n}. \tag{C.11}
\]

However, the physically degraded FSBC satisfies (7), thus
\[
p(y^n, z^n|u^n, s_0) = \sum_{x^n} p(y^n, z^n, x^n|u^n, s_0)
\]
\[
= \sum_{x^n} p(y^n, x^n|u^n, s_0)p(z^n|y^n, x^n, u^n, s_0)
\]
\[
= \sum_{x^n} p(y^n, x^n|u^n, s_0)p(z^n|y^n, s_0)
\]
\[
= p(y^n|u^n, s_0)p(z^n|y^n, s_0), \quad \forall s_0 \in \mathcal{S},
\]

and by the data processing inequality we obtain $I(U^n; Y^n|s_0') \geq I(U^n; Z^n|s_0')$, where $s_0'$ minimizes (C.11). Thus,
\[
\min_{s_0 \in \mathcal{S}} \frac{1}{n} I(U^n; Z^n|s_0) \leq \frac{1}{n} I(U^n; Z^n|s_0') \leq \frac{1}{n} I(U^n; Y^n|s_0) = \min_{s_0 \in \mathcal{S}} \frac{1}{n} I(U^n; Y^n|s_0).
\]

It follows that, if for decoding $m_2$ at $R_{x_1}$ we can find a $\rho^* > 0$ that results in a positive error exponent whenever $R_2 < R_{2,n}(p)$, then it is possible to find a (possibly different) $\rho^* > 0$ that results in a positive error exponent for decoding $m_2$ at $R_{x_1}$.
E. Decoding the Message $m_1$ at the Strong Receiver $Rx_1$

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

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

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

$$\text{if } \forall m'_1 \in M_1, m'_1 \neq \hat{m}_1 \quad \tilde{p}(y|x(\hat{m}_1, \hat{m}_2)) \geq \tilde{p}(y|x(m'_1, \hat{m}_2)) \quad \Rightarrow g_y(y) = \hat{m}_1,$$

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

$$P_{e_{11}}^{(n)}(m_1|\hat{m}_2, s_0) = \sum_{y: \exists m'_1 \neq m_1, \tilde{p}(y|x(m_1, \hat{m}_2)) \leq \tilde{p}(y|x(m'_1, \hat{m}_2))} \sum_{z^n,s^n,s_0 \in S} p(z,y,s|x(m_1, \hat{m}_2), s_0).$$

From now on we assume $\hat{m}_2 = m_2$. The bound on the probability of error averaged over all selections of the codebook $\{x(m_1,m_2)\}_{m_1 \in M_1}$ for a fixed $u(m_2)$ is given by (cf. (C.7))

$$\bar{P}_{e_{11}}^{(n)}(m_1|u(m_2), s_0) \leq ||S| (||M_1|| - 1)^{\theta} \sum_{y^n} \left[ \sum_{x^n} p(x|u(m_2)) \left( \sum_{s_0 \in S} \frac{1}{||S||} p(y|x,s_0) \right)^{\frac{1}{1+\rho}} \right]^{1+\rho}.$$

The bound averaged over all possible selections of $u(m_2)$ is given by

$$P_{e_{11}}^{(n)}(m_1|m_2, s_0) = \sum_{x^n \in U_2} P_{e_{11}}^{(n)}(m_1|u(m_2), s_0).$$

Bounding the expression using similar steps to those leading to (C.8) we obtain

$$P_{e_{11}}^{(n)}(m_1|m_2, s_0) \leq ||S| (||M_1|| - 1)^{\theta} \sum_{x^n \in U_2} p(u(m_2)) \sum_{y^n} \left[ \sum_{x^n} p(x|u(m_2)) \left( \sum_{s_0 \in S} \frac{1}{||S||} p(y|x,s_0) \right)^{\frac{1}{1+\rho}} \right]^{1+\rho} \leq ||S| (||M_1|| - 1)^{\theta} \sum_{s_0 \in S} \frac{1}{||S||} \sum_{x^n \in U_2} p(u(m_2)) \sum_{y^n} \left[ \sum_{x^n} p(x|u(m_2)) p(y|x,s_0) \right]^{\frac{1}{1+\rho}} \leq ||S|| \left( -\rho \log_{2} \left( \frac{||S||}{n} \right) + \min_{s_0 \in S} - \frac{1}{n} \log_2 \sum_{x^n \in U_2} p(u) \sum_{y^n} \left[ \sum_{x^n} p(x|u) p(y|x,s_0) \right]^{\frac{1}{1+\rho}} \right)^{1+\rho} \leq ||S||^{2} \left( -\rho R_1 - \rho \log_2 \left( \frac{||S||}{n} \right) + \min_{s_0 \in S} - \frac{1}{n} \log_2 \sum_{x^n \in U_2} p(u) \sum_{y^n} \left[ \sum_{x^n} p(x|u) p(y|x,s_0) \right]^{\frac{1}{1+\rho}} \right)^{1+\rho} \right). \quad (C.12)
As in the analysis for $R_{x_2}$, for every initial state $s_0$ we find the maximum $R_1$ that allows a positive error exponent. Then, minimizing $R_1$ over all initial states guarantees a positive error exponent for all $s_0 \in S$. To see this define

$$E_{n,11}(\rho, p(u, x), s_0) = \frac{1}{n} \log_2 \sum_{x_i^n = t} p(u) \sum_{y^n} \left[ \sum_{x^n} p(x|u)p(y|x, s_0) \frac{1}{1+\rho} \right]^{1+\rho}$$

(C.13)

$$F_{n,11}(\rho, p(u, x)) = \min_{s_0 \in S} E_{n,11}(\rho, p(u, x), s_0) - \rho \frac{\log_2 ||S||}{n},$$

(C.14)

and rewrite (C.12) as

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

(C.15)

To find a $\rho > 0$ that results in a positive value of $F_{n,11}(\rho, p(u, x)) - \rho R_1$ we first equate the first derivative of $E_{n,11}(\rho, p(u, x), s_0)$ w.r.t. $\rho$ to zero. Differentiating $E_{n,11}(\rho, p(u, x), s_0)$ w.r.t. $\rho$ results in

$$-\frac{1}{n} \frac{\partial}{\partial \rho} \log_2 \sum_{y^n} \sum_{x_i^n = t} p(u) \left[ \sum_{x^n} p(x|u)p(y|x, s_0) \frac{1}{1+\rho} \right]^{1+\rho} = -\frac{1}{n} \sum_{y^n} \sum_{x_i^n = t} p(u) \left[ \sum_{x^n} p(x|u)p(y|x, s_0) \frac{1}{1+\rho} \right]^{1+\rho} \ln 2 \times$$

$$\left[ \sum_{y^n} \sum_{x_i^n = t} p(u) \left\{ \sum_{x^n} p(x|u)p(y|x, s_0) \frac{1}{1+\rho} \right\}^{1+\rho} \times$$

$$\left( \ln \sum_{x^n} p(x|u)p(y|x, s_0) \frac{1}{1+\rho} - \frac{1}{1+\rho} \sum_{x^n} p(x|u)p(y|x, s_0) \ln(p(y|x, s_0)) \right) \right] \right].$$

Setting $\rho = 0$ we obtain

$$-\frac{1}{n} \frac{\partial}{\partial \rho} \log_2 \sum_{y^n} \sum_{x_i^n = t} p(u) \left[ \sum_{x^n} p(x|u)p(y|x, s_0) \frac{1}{1+\rho} \right]^{1+\rho} \bigg|_{\rho=0}$$

$$= -\frac{1}{n} \sum_{y^n} \sum_{x_i^n = t} p(u) \left[ \sum_{x^n} p(x|u)p(y|x, s_0) \right] \ln 2 \times$$

$$\left[ \sum_{y^n} \sum_{x_i^n = t} p(u) \left\{ \sum_{x^n} p(x|u)p(y|x, s_0) \right\} \times$$

$$\left( \ln \sum_{x^n} p(x|u)p(y|x, s_0) - \frac{1}{\sum_{x^n} p(x|u)p(y|x, s_0)} \ln(p(y|x, s_0)) \right) \right] \right]$$

$$= \frac{1}{n} I(X^n; Y^n|U^n, s_0).$$

The next step is to show that $R_1 \leq \frac{1}{n} I(X^n; Y^n|U^n, s_0) - \frac{\log_2 ||S||}{n}$ guarantees the existence of a positive error exponent when the initial state is $s_0$. Basically, one can repeat the argument in Appendix C-C.
However, since the expression (C.13) contains an auxiliary RV with an extra summation, we provide all the details of the proof. The conclusion is stated in the following lemma:

**Lemma C.1.** For any superposition codebook satisfying the Markov relationship $U^n|s_0 \leftrightarrow X^n|s_0 \leftrightarrow Y^n|s_0$, as long as

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

(C.16)

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

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

This implies that for $R_1 \leq \min_{s_0 \in S} \frac{1}{n} I(X^n; Y^n|U^n, s_0) - \frac{\log_2 ||S||}{n}$ we can find a $\rho^* > 0$ such that

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

i.e. the error exponent in (C.15) is positive.

**Proof:** We first highlight the main elements of the proof. This will be followed by the detailed proof.

1) **Proof Outline:**

- Note that $\left[ E_{n,11}(\rho, p(u, x), s_0) - \rho \frac{\log_2 ||S||}{n} - \rho R_1 \right]_{\rho=0} = 0$.
- We show that $E_{n,11}(\rho, p(u, x), s_0) - \rho \frac{\log_2 ||S||}{n} - \rho R_1$ is concave in $\rho$. Thus if it decreases with $\rho$ at $\rho = 0$, it will keep decreasing and the error exponent will be negative (i.e. the bound in (C.15) is greater than 1). Therefore, for the bound to be useful, this expression must increase with $\rho$ at $\rho = 0$.
- The derivative at zero is $\left. \frac{\partial E_{n,11}(\rho, p(u, x), s_0)}{\partial \rho} - \frac{\log_2 ||S||}{n} - R_1 \right|_{\rho=0}$. Making it positive gives an upper bound on $R_1$.
- Furthermore, $\frac{\partial E_{n,11}(\rho, p(u, x), s_0)}{\partial \rho}$ is analytic in $\rho$ at $\rho \geq 0$. Therefore, $E_{n,11}(\rho, p(u, x), s_0) - \rho \frac{\log_2 ||S||}{n} - \rho R_1$ is continuous.
- Positive derivative at zero, together with the continuity of the expression and its first derivative, imply that there is a region of positive $\rho$ for which the error exponent is positive. This allows us to pick $\rho > 0$ and get a positive error exponent.

2) **Proof Details:** The stationary point is obtained when the first derivative is zero. Therefore,

$$R_1 = \left. \frac{\partial E_{n,11}(\rho, p(u, x), s_0)}{\partial \rho} - \frac{\log_2 ||S||}{n} \right|_{\rho=0},$$

is a stationary point. The first step is to show that $E_{n,11}(\rho, p(u, x), s_0) - \rho R_1 - \rho \frac{\log_2 ||S||}{n}$ is a concave function of $\rho$. Since $\rho R_1 + \rho \frac{\log_2 ||S||}{n}$ is linear in $\rho$, it is also convex, and it is left to show that $E_{n,11}(\rho, p(u, x), s_0)$ is concave in $\rho$.
Concavity implies that for \( \rho_3 = \theta \rho_1 + (1 - \theta) \rho_2, \rho_1, \rho_2 > 0, 0 < \theta < 1, \)

\[ E_{n,11}(\rho_3, p(u, x), s_0) \geq \theta E_{n,11}(\rho_1, p(u, x), s_0) + (1 - \theta)E_{n,11}(\rho_2, p(u, x), s_0). \]

We start with [5, Equation (5B.6)] in the proof of the lemma in [5, Appendix 5B] \(^6\). This equation states that for a fixed \( s_0, u \) and \( y, \) a distribution \( p(x|u) \) on \( \mathcal{X}^n, \) a set of positive numbers \( p(y|x, s_0), \) a pair of positive numbers \( s, r, \) a number \( \theta, 0 < \theta < 1, \) and \( t = \theta s + (1 - \theta) r, \) that

\[
\left[ \sum_{x \in \mathcal{X}^n} p(x|u)p(y|x, s_0)^{t} \right]^{s} \leq \left[ \sum_{x \in \mathcal{X}^n} p(x|u)p(y|x, s_0)^{\frac{t}{1+\rho_3}} \right]^{s} \left[ \sum_{x \in \mathcal{X}^n} p(x|u)p(y|x, s_0)^{\frac{t}{1+\rho_1}} \right]^{(1+\rho_1)\theta} \times \\
\left[ \sum_{x \in \mathcal{X}^n} p(x|u)p(y|x, s_0)^{\frac{t}{1+\rho_2}} \right]^{(1+\rho_2)(1-\theta)}.
\]  

(C.17)

Using (C.17) with \( t = 1 + \rho_3, s = 1 + \rho_1 \) and \( r = 1 + \rho_2 \) we have

\[
\left[ \sum_{x \in \mathcal{X}^n} p(x|u)p(y|x, s_0)^{1+\rho_3} \right]^{(1+\rho_3)\theta} \leq \left[ \sum_{x \in \mathcal{X}^n} p(x|u)p(y|x, s_0)^{1+\rho_1} \right]^{(1+\rho_1)\theta} \times \\
\left[ \sum_{x \in \mathcal{X}^n} p(x|u)p(y|x, s_0)^{1+\rho_2} \right]^{(1+\rho_2)(1-\theta)}.
\]

Multiplying both sides by \( p(u) \) and summing over all \( (u, y) \) we get

\[
\sum_{u \in \mathcal{X}^n} \sum_{y \in \mathcal{Y}^n} p(u) \left[ \sum_{x \in \mathcal{X}^n} p(x|u)p(y|x, s_0)^{\frac{1}{1+\rho_3}} \right]^{(1+\rho_3)\theta} \leq \sum_{u \in \mathcal{X}^n} \sum_{y \in \mathcal{Y}^n} p(u) \left[ \sum_{x \in \mathcal{X}^n} p(x|u)p(y|x, s_0)^{\frac{1}{1+\rho_1}} \right]^{(1+\rho_1)\theta} \times \\
\left[ \sum_{x \in \mathcal{X}^n} p(x|u)p(y|x, s_0)^{\frac{1}{1+\rho_2}} \right]^{(1+\rho_2)(1-\theta)}
\]

\[
= \sum_{u \in \mathcal{X}^n} \sum_{y \in \mathcal{Y}^n} \left[ p(u)^{\frac{1}{1+\rho_1}} \sum_{x \in \mathcal{X}^n} p(x|u)p(y|x, s_0)^{\frac{1}{1+\rho_1}} \right]^{(1+\rho_1)\theta} \times \\
\left[ p(u)^{\frac{1}{1+\rho_2}} \sum_{x \in \mathcal{X}^n} p(x|u)p(y|x, s_0)^{\frac{1}{1+\rho_2}} \right]^{(1+\rho_2)(1-\theta)}
\]

where we expressed \( p(u) = p(u)^\theta p(u)^{1-\theta}. \) We now apply Holder’s inequality (see [24, Equation 2.8.3]):

\[
\sum_j a_j b_j \leq \left[ \sum_j a_j^\frac{1}{\theta} \right]^\theta \left[ \sum_j b_j^{\frac{1}{1-\theta}} \right]^{1-\theta}
\]

\(^6\)The equation [5, Equation (5B.6)] follows from Holder’s inequality, see [24, Section 2.8, Theorem 12], [5, Problem 4.15c].
for non-negative $a_j$’s and $b_j$’s and a finite number of values $j, \theta \in (0, 1)$. This results in:

$$\sum_{u \in X^n} \sum_{y \in Y^n} \left[ p(u)^{1+\rho_1} \sum_{x \in X^n} p(x|u)p(y|x, s_0)^{1+\rho_1} \right]^{(1+\rho_1)\theta} \left[ p(u)^{1+\rho_2} \sum_{x \in X^n} p(x|u)p(y|x, s_0)^{1+\rho_2} \right]^{(1+\rho_2)(1-\theta)}$$

$$\leq \left[ \sum_{u \in X^n} \sum_{y \in Y^n} \left[ p(u)^{1+\rho_1} \sum_{x \in X^n} p(x|u)p(y|x, s_0)^{1+\rho_1} \right]^{(1+\rho_1)} \right]^\theta \times \left[ \sum_{u \in X^n} \sum_{y \in Y^n} \left[ p(u)^{1+\rho_2} \sum_{x \in X^n} p(x|u)p(y|x, s_0)^{1+\rho_2} \right]^{(1+\rho_2)} \right]^{(1-\theta)}$$

$$= \left[ \sum_{u \in X^n} \sum_{y \in Y^n} p(u) \left[ \sum_{x \in X^n} p(x|u)p(y|x, s_0)^{1+\rho_1} \right] \right]^{(1+\rho_1)} \times \left[ \sum_{u \in X^n} \sum_{y \in Y^n} p(u) \left[ \sum_{x \in X^n} p(x|u)p(y|x, s_0)^{1+\rho_2} \right] \right]^{(1+\rho_2)}$$

where we treated the double summation as a single summation over a larger alphabet. Taking $\log_2$ of both sides we obtain

$$-n E_{n,11}(\rho_3, p(u, x), s_0) \leq -n E_{n,11}(\rho_1, p(u, x), s_0)\theta - n E_{n,11}(\rho_2, p(u, x), s_0)(1-\theta)$$

which is the desired result.

Finally, we note that $E_{n,11}(0, p(u, x), s_0) = 0$, and if $I(X^n; Y^n|U^n, s_0) > 0$, then $E_{n,11}(\rho, p(u, x), s_0) > 0$ for $\rho > 0$. To see the latter, we recall the first statement in the proof of the lemma in [5, Appendix 5B]. This statement asserts that

$$\left[ \sum_{x \in X^n} p(x|u)p(y|x, s_0)^{1+\rho} \right]^{1+\rho}$$

is non-increasing in $\rho$. Since $\frac{1}{n} I(X^n; Y^n|U^n, s_0) > 0$ we have that for at least one $u$ s.t. $p(u|s_0) = p(u) > 0$ then $I(X^n; Y^n|u, s_0) > 0$. Hence, from the properties of mutual information it follows that $Y|u, s_0$ is not independent of $X|u, s_0$, i.e. $p(y|x, u, s_0) = p(y|x, s_0)$ is not constant for all $x$ with $p(x|u, s_0) = p(x|u) > 0$. Thus from the lemma, $\left[ \sum_{x \in X^n} p(x|u)p(y|x, s_0)^{1+\rho} \right]^{1+\rho}$ is strictly decreasing with $\rho$, and

---

7 Lemma in [5, Appendix 5B]: Let $Q = [Q(0), Q(1), \ldots, Q(K-1)]$ be a probability vector and let $a_0, a_1, \ldots, a_{K-1}$ be a set of non-negative numbers. Then the function

$$f(s) = \ln \left[ \sum_{k=0}^{K-1} Q(k)a_k^s \right]$$

is non-increasing and convex with $s$, for $s > 0$. Moreover, $f(s)$ is strictly decreasing unless all $a_k$ for which $Q(k) > 0$ are equal. The convexity is strict unless all the non-zero $a_k$ for which $Q(k) > 0$ are equal.
hence after adding the average over \( u \) we have that 
\[
\sum_{y^n} \sum_{x^n \in \mathcal{U}_n} p(u) \left( \sum_{x \in \mathcal{X}^n} p(x|u)p(y|x, s_0) \right)^{1+\rho} 
\]
is strictly decreasing with \( \rho \) (since the other values in the averaging are either strictly decreasing or non-increasing). Therefore \( E_{n,11}(\rho, p(u, x), s_0) \) is strictly increasing with \( \rho \). This implies that 
\[
\frac{\partial E_{n,11}(\rho, p(u, x), s_0)}{\partial \rho} > 0, 
\]
and hence, \( E_{n,11}(\rho, p(u, x), s_0) > 0 \) for \( \rho > 0 \).

Combined with the concavity of \( E_{n,11}(\rho, p(u, x), s_0) \) this implies that 
\[
\frac{\partial E_{n,11}(\rho, p(u, x), s_0)}{\partial \rho} 
\]
is decreasing with \( \rho \), thus 
\[
\left( \frac{\partial E_{n,11}(\rho, p(u, x), s_0)}{\partial \rho} - \frac{\log_2 ||S||}{n} \right) \mid_{\rho = 0} > 0 
\]
is maximized at \( \rho = 0 \) for which we obtain the value 
\[
\frac{1}{n} I(X^n; Y^n|U^n, s_0) - \frac{\log_2 ||S||}{n} 
\]
In conclusion, making 
\[
\left( \frac{\partial E_{n,11}(\rho, p(u, x), s_0)}{\partial \rho} - R_1 - \frac{\log_2 ||S||}{n} \right) \mid_{\rho = 0} > 0 
\]
is possible when 
\[
R_1 < \left( \frac{\partial E_{n,11}(\rho, p(u, x), s_0)}{\partial \rho} \right) \mid_{\rho = 0} - \frac{\log_2 ||S||}{n} 
\]
By continuity we conclude that for this range of \( R_1 \) there exists a \( \rho > 0 \) such that \( E_{n,11}(\rho, p(u, x), s_0) - R_1 \left( R_1 + \frac{\log_2 ||S||}{n} \right) > 0 \), and furthermore, by convexity this is the largest possible range for \( R_1 \).

Since \( R_1 \leq \frac{1}{n} I(X^n; Y^n|U^n, s_0) - \frac{\log_2 ||S||}{n} \) allows a positive error exponent for initial state \( s_0 \), thus letting 
\[
R_1 \leq \min_{s_0' \in S} \frac{1}{n} I(X^n; Y^n|U^n, s_0') - \frac{\log_2 ||S||}{n} \triangleq R_{1,n}(p) 
\]
guarantees a positive error exponent for every initial state, and therefore also under the minimum in the expression of the error exponent (C.14). Again, we deal with negative \( R_{1,n}(p) \) by taking \( n \) large enough.

In combination with Appendix C-D we conclude that if also \( R_2 \leq R_{2,n}(p) \) then it is possible to achieve positive error exponents for both \( \tilde{P}_{e_{11}}(m_1|m_2) \) and \( \tilde{P}_{e_{12}}(m_2|s_0) \), simultaneously.

**E. Proving that the Rate Pair \( (R_{1,n}(p), R_{2,n}(p)) \) is Achievable**

Combining the results of Appendix C-C – C-E we have that for the physically degraded FSBC, any rate pair \( (R_1, R_2) \) that belongs to the convex hull of the region of all positive rate pairs \( (R_1, R_2) \) satisfying 
\[
R_1 \leq \min_{s_0' \in S} \frac{1}{n} I(X^n; Y^n|U^n, s_0') - \frac{\log_2 ||S||}{n} 
\]
\[
R_2 \leq \min_{s_0 \in S} \frac{1}{n} I(U^n; Z^n|s_0) - \frac{\log_2 ||S||}{n}, 
\]
for some joint distribution \( p(u^n, x^n) \), results in positive error exponents.

We now show that the rates of Equations (C.18) – (C.19) are achievable. It is enough to show that given a maximum average probability of error \( \epsilon > 0 \), then for the positive rate-pair 
\[
(R_{1,n}(p), R_{2,n}(p)) = \left( \min_{s_0' \in S} \frac{1}{n} I(X^n; Y^n|U^n, s_0') - \frac{\log_2 ||S||}{n} , \min_{s_0 \in S} \frac{1}{n} I(U^n; Z^n|s_0) - \frac{\log_2 ||S||}{n} \right) 
\]
we can find a blocklength \( b_0 \) such that a code \( (R_{1,n}(p), R_{2,n}(p), b) \) with \( \tilde{P}_e(b, s_0) \leq \epsilon \) can be constructed for all \( b > b_0 \) and \( s_0 \in S \).

April 30, 2008 DRAFT
We recall here that we actually have the same bound for all initial states \( s_0 \). Therefore the probability of error bound 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 be

\[
p(u^b, x^b) = \prod_{k=1}^{K} q_n(u_{(k-1)n+1}^k, x_{(k-1)n+1}^k) \equiv \prod_{k=1}^{K} q_n(u_k, x_k).
\]  

(C.20)

We now prove the following lemma:

**Lemma C.2.** For \( F_{n,11}(\rho, p(u^b, x^b)) \) defined in (C.14) under the distribution law (C.20) it holds that

\[
F_{b,11}\left( \rho, \prod_{k=1}^{K} q_n(u_k, x_k) \right) = F_{b,11}(\rho, p(u^b, x^b)) \geq F_{n,11}(\rho, q_n(u^n, x^n)).
\]  

(C.21)

**Proof:** The proof uses the same essential steps as in Gallager’s derivation [5, Section 5.9] and [15, Lemma 19]. However we have to make sure that also with the added complexity of the superposition codebook, the main steps still hold (this is not trivial, see Appendix E).

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 \( p_N(u^N, x^N) = p_M(u^M, x^M) p_L(u^N_{M+1}, x^N_{M+1}) \). We denote \( (u_1, x_1) \equiv (u^M, x^M) \) and \( (u_2, x_2) \equiv (u^N_{M+1}, x^N_{M+1}) \). Then,

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

Let \( s_{0,N} \) be the minimizing state. Raising both sides to the power of two we obtain from definition (C.13) that

\[
2^{-NF_{N,11}(\rho, p_N(u^N, x^N))} = ||S||^\rho \sum_{X_{1:n}, u \in U} p(u) \sum_{Y^N} \left[ \sum_{X^N} p(x|u) p(y|x, s_{0,N}) \right]^{\frac{1}{1+\rho}}.
\]  

(C.22)
Now write

\[
\sum_{\mathbf{u} \in \mathcal{U}} p(\mathbf{u}) \sum_{\mathcal{Y}^N} \left[ \sum_{\mathcal{X}^N} p(\mathbf{x} | \mathbf{u}) p(\mathbf{y} | \mathbf{x}, s_{0,N}) \right]^\frac{1}{1+p}
\]

\[
\quad \equiv \sum_{\mathbf{u}_1 \in \mathcal{U}_1} \sum_{\mathbf{u}_2 \in \mathcal{U}_2} p_M(\mathbf{u}_1)p_L(\mathbf{u}_2) \sum_{\mathcal{Y}_1^{M-1}} \sum_{\mathcal{Y}_2^L} \left[ \sum_{\mathcal{X}_1^{M-1}} p(\mathbf{x}_1, \mathbf{x}_2 | \mathbf{u}_1, \mathbf{u}_2) p(\mathbf{y}_1, \mathbf{y}_2 | \mathbf{x}_1, \mathbf{x}_2, s_{0,N}) \right]^\frac{1}{1+p}
\]

\[
\equiv \sum_{\mathbf{u}_1 \in \mathcal{U}_1} \sum_{\mathbf{u}_2 \in \mathcal{U}_2} p_M(\mathbf{u}_1)p_L(\mathbf{u}_2) \sum_{\mathcal{Y}_1^{M-1}} \sum_{\mathcal{Y}_2^L} \left[ \sum_{\mathcal{X}_1^{M-1}} \sum_{\mathcal{X}_2^L} p_M(\mathbf{x}_1 | \mathbf{u}_1)p_L(\mathbf{x}_2 | \mathbf{u}_2) p(\mathbf{y}_1, \mathbf{y}_2 | \mathbf{x}_1, \mathbf{x}_2, s_{0,N}) \right]^\frac{1}{1+p}
\]

\[
= \sum_{\mathbf{u}_1 \in \mathcal{U}_1} \sum_{\mathbf{u}_2 \in \mathcal{U}_2} p_M(\mathbf{u}_1)p_L(\mathbf{u}_2) \sum_{\mathcal{Y}_1^{M-1}} \sum_{\mathcal{Y}_2^L} \left[ \sum_{\mathcal{X}_1^{M-1}} \sum_{\mathcal{X}_2^L} p_M(\mathbf{x}_1 | \mathbf{u}_1)p_L(\mathbf{x}_2 | \mathbf{u}_2) \left[ \sum_{s_{0,N} \in \mathcal{S}} p(\mathbf{y}_1, \mathbf{y}_2, s_{0,N} | \mathbf{x}_1, \mathbf{x}_2, s_{0,N}) \right] \right]^\frac{1}{1+p}
\]

\[
\equiv \sum_{\mathbf{u}_1 \in \mathcal{U}_1} \sum_{\mathbf{u}_2 \in \mathcal{U}_2} p_M(\mathbf{u}_1)p_L(\mathbf{u}_2) \sum_{\mathcal{Y}_1^{M-1}} \sum_{\mathcal{Y}_2^L} \left[ \sum_{\mathcal{X}_1^{M-1}} \sum_{\mathcal{X}_2^L} \sum_{s_{0,N} \in \mathcal{S}} p_M(\mathbf{x}_1 | \mathbf{u}_1)p_L(\mathbf{x}_2 | \mathbf{u}_2) p(\mathbf{y}_1, s_{0,N} | \mathbf{x}_1) p(\mathbf{y}_2 | s_{0,N} | \mathbf{x}_2) \right]^\frac{1}{1+p}
\]

\[
\leq \left| \mathcal{S} \right|^{1+p} \sum_{\mathbf{u}_1 \in \mathcal{U}_1} \sum_{\mathbf{u}_2 \in \mathcal{U}_2} p_M(\mathbf{u}_1)p_L(\mathbf{u}_2) \sum_{\mathcal{Y}_1^{M-1}} \sum_{\mathcal{Y}_2^L} \left[ \sum_{\mathcal{X}_1^{M-1}} \sum_{\mathcal{X}_2^L} \sum_{s_{0,N} \in \mathcal{S}} p_M(\mathbf{x}_1 | \mathbf{u}_1)p_L(\mathbf{x}_2 | \mathbf{u}_2) p(\mathbf{y}_1, s_{0,N} | \mathbf{x}_1) p(\mathbf{y}_2 | s_{0,N} | \mathbf{x}_2) \right]^\frac{1}{1+p}
\]
\[
\begin{align*}
&(e) \quad ||S||^p \sum_{s_M \in S} \sum_{x_M} \sum_{y_L} p_M(u_1)p_L(u_2) \sum_{y_M} \sum_{x_L} \left( \sum_{s_M \in S} \sum_{x_M} p_M(x_1|u_1)p(y_1, s_M|x_1, s_{0,N})^{\frac{1}{1+p}}p(y_2|s_M, x_2)^{\frac{1}{1+p}} \right)^{1+p} \\
&= ||S||^p \sum_{s_M \in S} \sum_{x_M} \sum_{y_L} p_M(u_1) \left( \sum_{x_M} \sum_{y_M} p_M(x_1|u_1)p(y_1, s_M|x_1, s_{0,N})^{\frac{1}{1+p}} \right)^{1+p} \\
&\quad \times \sum_{x_M} \sum_{y_M} p_L(u_2) \left( \sum_{y_L} \sum_{x_L} p_L(x_2|u_2)p(y_2|s_M, x_2)^{\frac{1}{1+p}} \right)^{1+p} \\
&= \sum_{s_M \in S} \sum_{x_M} p_M(u_1) \left( \sum_{x_M} \sum_{y_M} p_M(x_1|u_1)p(y_1, s_M|x_1, s_{0,N})^{\frac{1}{1+p}} \right)^{1+p} 2^{-L(E_{L,11}(\rho, p_L(u_2, x_2), s_M) - \frac{\log_2 ||S||}{L})} \\
&\leq \sum_{s_M \in S} \sum_{x_M} p_M(u_1) \left( \sum_{x_M} \sum_{y_M} p_M(x_1|u_1)p(y_1, s_M|x_1, s_{0,N})^{\frac{1}{1+p}} \right)^{1+p} 2^{-LF_{L,11}(\rho, p_L(u_2, x_2))} \\
&\leq \sum_{x_M} \sum_{y_M} \left( \sum_{x_M} \sum_{y_M} p_M(x_1|u_1)p(y_1, s_M|x_1, s_{0,N})^{\frac{1}{1+p}} \right)^{1+p} 2^{-LF_{L,11}(\rho, p_L(u_2, x_2))} \\
&= \sum_{x_M} \sum_{y_M} \left( \sum_{x_M} \sum_{y_M} p_M(x_1|u_1)p(y_1, s_M|x_1, s_{0,N})^{\frac{1}{1+p}} \right)^{1+p} 2^{-LF_{L,11}(\rho, p_L(u_2, x_2))},
\end{align*}
\]

where (a) and (b) are because

\[
p_N(x_1, x_2, u_1, u_2) = p_M(u_1, x_1)p_L(u_2, x_2)
\]

\[
\Rightarrow p_N(u_1, u_2) = p_M(u_1)p_L(u_2)
\]

\[
\Rightarrow p_N(x_1, x_2|u_1, u_2) = p_M(x_1|u_1)p_L(x_2|u_2),
\]

(c) is because \(x_1\) and \(x_2\) are independent and therefore \(X_2 \perp (X_1, Y_1, S_M, S_0)\) when \(Y_2, Z_2\) and \(S_2\) are not given, and also because \(Y_2\) is independent of \(S_0, X_1\) and \(Y_1\) when \(S_M\) is given. In (d) we used \((\sum_i a_i)^r \leq \sum_i a_i^r, a_i \geq 0, 0 < r \leq 1\) ([24, Section 2.10, Theorem 19]), and in (e) we used \((\sum_i P_i a_i)^r \leq \sum_i P_i a_i^r, a_i \geq 0, r \geq 1\) and \(\{P_i\}\) is a p.m.f. ([24, Section 2.9, Theorem 16]). For (f) we used the fact that \(F_{L,11}(\rho, p_L(u_2, x_2))\) is evaluated with the initial state \(s_0 \in S\) that minimizes \(E_{L,11}(\rho, p(u, x), s_0)\) and (g) follows from Minkowski’s inequality: \(\sum_k \left( \sum_j Q_j a_{jk} \right)^r \leq \left( \sum_j Q_j \left( \sum_k a_{jk} \right)^\frac{r}{2} \right)^r, r \geq 1, a_{jk} \geq 0,\)
\{Q_j\} is a p.m.f. ([24, Section 2.11, Theorem 24]). Plugging this back into (C.22) yields
\[
2^{-NF_{N,11}(\rho, p_N(u^N, x^N))} \leq ||S||^p \sum_{x_{i=1}^M} p_M(u_1) \sum_{\chi_{i=1}^M} \left( \sum_{x_{i=1}^M} p_M(x_1|u_1)p(y_1|x_1, s_{0N}) \right)^{\frac{1}{\alpha + \rho}} 2^{-L_{F,11}(\rho, p_L(u_2, x_2))}
\]
\[
= 2^{-M(E_{M,11}(\rho, p_M(u_1, x_1), y_{0N}) - \rho \log_2 ||S||)} 2^{-L_{F,11}(\rho, p_L(u_2, x_2))}
\]
\[
\leq 2^{-M(F_{M,11}(\rho, p_M(u_1, x_1))} 2^{-L_{F,11}(\rho, p_L(u_2, x_2))}
\]
\[\Rightarrow NF_{N,11}(\rho, p_N(u^N, x^N)) \geq MF_{M,11}(\rho, p_M(u_1, x_1)) + LF_{L,11}(\rho, p_L(u^L, x^L))\]

Therefore
\[
bF_{b,11} \left( \rho, \prod_{k=1}^{K} q_n(u_k, x_k) \right) \geq nF_{n,11}(\rho, q_n(u, x)) + (K - 1)nF_{(K-1)n,11}(\rho, \prod_{k=1}^{K-1} q_n(u_k, x_k))
\]
\[
\geq 2nF_{n,11}(\rho, q_n(u, x)) + (K - 2)nF_{(K-2)n,11}(\rho, \prod_{k=1}^{K-2} q_n(u_k, x_k))
\]
\[\geq KF_{n,11}(\rho, q_n(u, x)),\]
where (a) follows by continuing the process in the first 2 steps.

Using Lemma C.2 in (C.15) we can write the bound on \(\tilde{P}_{e_{11}}(m_1|m_2, s_0)\) as
\[
\tilde{P}_{e_{11}}(m_1|m_2, s_0) \leq ||S||^{2-Kn(F_{K,n,11}(\rho, \Pi_{n=1}^{K} q_n(u, x_n)) - \rho R_1(q_n))}
\]
\[
\leq ||S||^{2-Kn(F_{n,11}(\rho, q_n(u, x)) - \rho R_1(q_n))}
\]

From Lemma C.1 we conclude that there exists a positive \(\rho_{11}\) such that \(F_{n,11}(\rho_{11}, q_n(u, x)) - \rho_{11} R_1(q_n) = \epsilon_{11}\) for some \(\epsilon_{11} > 0\). Hence, the probability \(\tilde{P}_{e_{11}}(m_1|m_2, s_0)\) can be made arbitrarily small by taking \(K\) large enough. The same considerations can be repeated for \(\tilde{P}_{e_{12}}(m_2|s_0)\) and \(\tilde{P}_{e_2}(m_2|s_0)\). Thus, taking \(K\) large enough we have that \(\tilde{P}_{e_{11}}(m_1|m_2, s_0) + \tilde{P}_{e_{12}}(m_2|s_0) + \tilde{P}_{e_2}(m_2|s_0)\) 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 \(\tilde{P}_{e_{11}}(m_1|m_2, s_0) \leq \epsilon, \forall s_0 \in S\) allows us to generate codes for any blocklength \(b > Kn\) with \(\tilde{P}_{e_{11}}(s_0) \leq \epsilon, \forall s_0 \in S\). 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 \(\max\{R_{1,n}(q_n), R_{2,n}(q_n)\} < \frac{\delta}{2}\), and adding zeros when the blocklength is not an integer multiple of \(K\).

We omit the details of the derivation that shows
\[
F_{Kn,2}(\rho, p(u^{Kn}, x^{Kn})) \geq F_{n,2}(\rho, q_n(u, x)),\]
(C.23)
except the critical step of introducing $s_M$ into the channel that connects $U^n$ and $Z^n$, which is characterized by $p(z|u, s_0)$. This p.m.f. is used in the expressions for $F_{K_n,2}$ and $F_{n,2}$:

$$p(z|u, s_0) = p(z_1, z_2|u_1, u_2, s_0)$$

$$= \sum_{s_M} p(z_1, z_2, s_M|u_1, u_2, s_0)$$

$$= \sum_{X^n, Y^n, S^n} \sum_{s_M \in S} p(z_1, z_2, y_1, y_2, s_1, s_2, x_1, x_2|u_1, u_2, s_0)$$

$$= \sum_{X^n, Y^n, S^n} \sum_{s_M \in S} p(z_1, z_2, y_1, y_2, s_1, s_2|x_1, x_2, u_1, u_2, s_0)p(x_1, x_2|u_1, u_2, s_0)$$

$$= \sum_{X^n, Y^n, S^n} \sum_{s_M \in S} p(z_1, z_2, y_1, y_2, s_1, s_2|x_1, x_2, s_0)p_M(x_1|u_1)p_L(x_2|u_2)$$

$$= \sum_{X^n, Y^n, S^n} \sum_{s_M \in S} p(z_1, y_1, s_1|x_1, x_2, s_0)p(z_2, y_2, s_2|x_1, x_2, s_0)p_M(x_1|u_1)p_L(x_2|u_2)$$

$$= \sum_{X^n, Y^n, S^n} \sum_{s_M \in S} p(z_1, y_1, s_1|x_1, x_2, s_0)p(z_2, x_2|x_1, s_2)p_M(x_1|u_1)p_L(x_2|u_2, s_M)$$

$$= \sum_{X^n, Y^n, S^n} \sum_{s_M \in S} p(z_1, y_1, s_1|x_1, s_0)p(z_2, x_2|x_1, s_0)p_M(x_1|u_1)$$

$$= \sum_{X^n, Y^n, S^n} \sum_{s_M \in S} p(z_1, y_1, s_1|x_1, s_0)p(z_2|s_M, x_2)p_M(x_1|u_1)p_L(x_2|u_2, s_M)$$

$$= \sum_{s_M \in S} p(z_1, s_M|u_1, s_0)p(z_2|s_M, u_2). \quad (C.24)$$

The expansion (C.24) is used to show that (C.23) holds.

\[ G. \text{ The Boundary of the Achievable Region} \]

To study the asymptotic properties of the achievable region, we first characterize its boundary. Begin by writing the achievable region for a given $n$ as

$$R_n = \text{co} \bigcup_{q_n \in Q_n} R_n(q_n),$$

$$R_n(q_n) \triangleq \left\{ (R_1, R_2) : 0 \leq R_1 \leq R_{1,n}(q_n), 0 \leq R_2 \leq R_{2,n}(q_n) \right\},$$

$$R_{1,n}(q_n) \triangleq \min_{s_0 \in S} \frac{1}{n} I(X^n; Y^n| U^n, s'_0, q_n) - \frac{\log_2 ||S||}{n},$$

$$R_{2,n}(q_n) \triangleq \min_{s_0 \in S} \frac{1}{n} I(U^n; Z^n| s_0, q_n) - \frac{\log_2 ||S||}{n},$$

April 30, 2008
where $Q_n$ is defined in Section III. Now, fix $\lambda \geq 0$ 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, for a given $n$. Hence, for a fixed $\lambda \geq 0$, the line
\[
R_2^\lambda(R_1) = \max_{p(u^n, x^n)} \{ R_{2,n}(p) + \lambda R_{1,n}(p) \} - \lambda R_1
\]
upper bounds the region of positive error exponents, otherwise for the same $\lambda$ there exists $\tilde{p}(u^n, x^n)$ for which $R_{2,n}(\tilde{p}) + \lambda R_{1,n}(\tilde{p}) > \max_{p(u^n, x^n)} \{ R_{2,n}(p) + \lambda R_{1,n}(p) \}$, thus contradicting the maximization. Following [16], [26, Lemma 3] we can write the boundary of $\mathcal{R}_n$ as the least upper bound:
\[
\tilde{R}_2(R_1) = \inf_{\lambda \geq 0} \left\{ \max_{p(u^n, x^n)} \left\{ \min_{s_0 \in S} \frac{1}{n} I(U^n; Z^n | s_0) - \frac{\log_2 ||S||}{n} \right. \right. \\
+ \lambda \left( \min_{s_0 \in S} \frac{1}{n} I(X^n; Y^n | U^n, s_0') - \frac{\log_2 ||S||}{n} \right) \left. \right\} - \lambda R_1. \tag{C.25}
\]
This situation is illustrated in Figure 4.

![Figure 4](image)

Fig. 4. Lines bounding the achievable region for the FSBC, and the resulting achievable region.

Note that when $\lambda = 0$, $R_2^0(R_1) = \max_{p(u^n, x^n)} \min_{s_0 \in S} \frac{1}{n} I(U^n; Z^n | s_0) - \frac{\log_2 ||S||}{n}$ is the resulting upper bound, irrespective of $R_1$. Next, consider $\lambda = 1$ and denote with $(\tilde{p}, \tilde{s}_0, \tilde{s}_0')$ the triplet that achieves the
max-min solution, where \( \tilde{s}_0 \) minimizes \( I(U^n; Z^n|s_0) \) and \( \tilde{s}'_0 \) minimizes \( I(X^n; Y^n|U^n, s'_0) \). Thus, we have

\[
R_2^1(R_1) = \left\{ -\frac{1}{n} I(U^n; Z^n|\tilde{s}_0) - \frac{1}{n} I(X^n; Y^n|U^n, \tilde{s}'_0) - 2 \frac{\log_2 ||S||}{n} \right\} - R_1
\]

\[
\leq \left\{ -\frac{1}{n} I(U^n; Z^n|\tilde{s}_0) - \frac{1}{n} I(X^n; Y^n|U^n, \tilde{s}'_0) - 2 \frac{\log_2 ||S||}{n} \right\} - R_1
\]

\[
\leq \left\{ -\frac{1}{n} I(U^n; Y^n|\tilde{s}'_0) - \frac{1}{n} I(X^n; Y^n|U^n, \tilde{s}'_0) - 2 \frac{\log_2 ||S||}{n} \right\} - R_1
\]

\[
= \frac{1}{n} I(X^n; Y^n|\tilde{s}'_0) - 2 \frac{\log_2 ||S||}{n} - R_1,
\]

irrespective of \( p(u^n) \). Hence, the bound is achieved using an atomic \( p(u^n) \): if \((p', s'_0)\) achieve that max-min solution for \( \max_{p(x^n)} \min_{s_0 \in S} \frac{1}{n} I(X^n; Y^n|s_0) \) then

\[
-\frac{1}{n} I(X^n; Y^n|s'_0)p' \geq -\frac{1}{n} I(X^n; Y^n|s'_0)p
\]

\[
\geq -\frac{1}{n} I(U^n; Z^n|s'_0)p + \frac{1}{n} I(X^n; Y^n|U^n, s'_0)p
\]

\[
\geq -\frac{1}{n} I(X^n; Y^n|U^n, s'_0)p + \frac{1}{n} I(X^n; Y^n|U^n, s'_0)p
\]

\[
\geq -\frac{1}{n} I(U^n; Z^n|s_0)p + \frac{1}{n} I(X^n; Y^n|U^n, s'_0)p.
\]

(C.26)

Denote \( C_0(1) = \max_{p(x^n)} \min_{s_0 \in S} \frac{1}{n} I(X^n; Y^n|s_0) - \frac{\log_2 ||S||}{n} \). 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

\[
R_2^\lambda(R_1) = \left\{ \lambda \frac{1}{n} I(U^n; Z^n|\tilde{s}_0) + \frac{1}{n} I(X^n; Y^n|U^n, \tilde{s}'_0) - (1 + \lambda) \frac{\log_2 ||S||}{n} \right\}
\]

\[
\quad \quad \quad - (\lambda - 1) \frac{1}{n} I(U^n; Z^n|\tilde{s}_0) \right\} - \lambda R_1
\]

\[
\leq \left\{ -(\lambda - 1) \frac{1}{n} I(U^n; Z^n|\tilde{s}_0) + \frac{1}{n} I(X^n; Y^n|\tilde{s}'_0) - (1 + \lambda) \frac{\log_2 ||S||}{n} \right\}
\]

\[
\leq \lambda \frac{1}{n} I(X^n; Y^n|\tilde{s}'_0) - (1 + \lambda) \frac{\log_2 ||S||}{n} - \lambda R_1
\]

\[
\leq \lambda(C_0(1) - R_1) - \frac{\log_2 ||S||}{n},
\]

(C.27)

where the last inequality can be shown similarly to (C.26). Thus, for \( \lambda > 1 \), \( R_2^\lambda(R_1) = \lambda(C_0(1) - R_1) - \frac{\log_2 ||S||}{n} \). Clearly when \( \lambda > 1 \), \( R_2^\lambda(C_0(1)) = R_2^\lambda(C_0(1)) = \frac{\log_2 ||S||}{n} \). However, when \( R_1 \) decreases, then the lines for \( \lambda > 1 \) pass to the right of \( R_2^\lambda(R_1) \). \( R_1 < C_0(1) \), and thus they do not constitute tight outer bounds. Therefore, it is enough to consider only \( 0 \leq \lambda \leq 1 \), to get a complete characterization of the region.

April 30, 2008

DRAFT
Define next
\[ C^n(\lambda) \triangleq \max_{p(u^n, x^n)} \left\{ \min_{s_0 \in \mathcal{S}} \frac{1}{n} I(U^n; Z^n | s_0) + \lambda \min_{s'_0 \in \mathcal{S}} \frac{1}{n} I(X^n; Y^n | U^n, s'_0) \right\} . \] (C.28)

We have the following lemma:

**Lemma C.3.** \( C^n(\lambda) \) defined in (C.28) converges as \( n \to \infty \) to a finite limit given by
\[ C^\infty(\lambda) \triangleq \lim_{n \to \infty} C^n(\lambda) = \sup_n \left[ C^n(\lambda) - (1 + \lambda) \frac{\log_2 ||\mathcal{S}||}{n} \right] . \] (C.29)
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 D. 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 page 7]. The fact that \( C^\infty(\lambda) \) provides the largest achievable region follows from its sup-additivity, also shown in Appendix D.

Lemma C.3 implies that the boundary of the achievable region, \( R_2(R_1) \), is characterized by
\[ R_2(R_1) = \inf_{0 \leq \lambda \leq 1} \{ C^\infty(\lambda) - \lambda R_1 \} . \] (C.30)

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 \( n \geq n(\epsilon, \delta) \), an \( (R_1 - \delta, R_2(R_1) - \delta, n) \) code with an average probability of error that is smaller than \( \epsilon \) can be constructed.

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.

**H. Cardinality Bounds**

From the derivation in [16], it follows that when \( \lambda \) and \( s_0 \) are fixed, then maximizing
\[ \frac{1}{n} I(U^n; Z^n | s_0) - \frac{\log_2 ||\mathcal{S}||}{n} + \lambda \left( \frac{1}{n} I(X^n; Y^n | U^n, s_0) - \frac{\log_2 ||\mathcal{S}||}{n} \right) \]
over all joint distributions \( p(u^n, x^n) \) can be carried out while the cardinality of the auxiliary random variable \( U^n \) is bounded by
\[ || \times_{i=1}^n U_i || \leq \min \{ \|\mathcal{X}^n\|, \|\mathcal{Y}^n\|, \|\mathcal{Z}^n\| \} \]
\[ = \min \{ \|\mathcal{X}\|^n, \|\mathcal{Y}\|^n, \|\mathcal{Z}\|^n \} \]
\[ = \min \{ \|\mathcal{X}\|, \|\mathcal{Y}\|, \|\mathcal{Z}\| \}^n . \] (C.31)
This is true for any $s_0 \in \mathcal{S}$. As follows from (19) and (20) in Section III-D, the achievable region can be written as the intersection $\bigcap_{s_0 \in \mathcal{S}} \mathcal{R}_n(s_0)$. Now, as for each $s_0 \in \mathcal{S}$ and $0 \leq \lambda \leq 1$ we have the same cardinality bound, then this bound is also valid for maximizing the intersection of the regions for these $s_0$'s (this intersection is a convex set). Therefore, it is enough to consider only RV's $U^n$ with bounded cardinality according to (C.31), to obtain a complete characterization of the achievable region.

**APPENDIX D**

**PROOF OF LEMMA C.3 (CONVERGENCE OF $\lim_{n \to \infty} C^n(\lambda)$)**

In this section we prove Lemma C.3. This lemma implies that

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

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

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

exists for the physically degraded FSBC when $\lambda$ is fixed and finite.

**Details of the Proof**

Recall the definition of $F_n(\lambda)$ in (15):

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

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

$$
nF_n(\lambda) \geq lF_l(\lambda) + mF_m(\lambda).$$
Let \( q(u^n, x^n) = q_1(u^n, x^n) q_2(u^n_{i+1}, x^n_{i+1}) \). Then, by definition,
\[
nF_n(\lambda) \geq [I(U^n; Z^n|s^0(n)) + \lambda I(X^n; Y^n|U^n, s^0(n))]_{q, q_2} - (1 + \lambda) \log_2 ||S||
\]
\[
= I(U^n_1; Z^n_1|s^0_0(n))_q + I(U^n_{i+1}; Z^n_{i+1}|s^0_0(n))_{q_2} + I(U^n_{i+1}; U^n_{i+1}; Z^n_{i+1}|Z^n_{i+1}, s^0_0(n))_{q_2}
\]
\[
+ \lambda \left( I(X^n_1; Y^n_1|U^n_1, Z^n_{i+1}|s^0_0(n))_{q_2} + I(X^n_{i+1}; Y^n_{i+1}|X^n_1, U^n_1, U^n_{i+1}, Y^n_{i+1}, s^0_0(n))_{q_1, q_2} \right)
\]
\[
+ I(X^n_1, X^n_{i+1}; Y^n_{i+1}|U^n_1, U^n_{i+1}, Y^n_{i+1}, s^0_0(n))_{q_1, q_2}
\]
\[
- (1 + \lambda) \log_2 ||S||.
\]

Now, by construction \((U^n, X^n)\) is independent of the initial state. Thus,
\[
q(u^n_1, x^n_1, u^n_{i+1}, x^n_{i+1}|s^0_0(n)) = q(u^n_1, x^n_1, u^n_{i+1}, x^n_{i+1})
\]
\[
= q_1(u^n_1, x^n_1) q_2(u^n_{i+1}, x^n_{i+1})
\]
\[
\Rightarrow q(u^n_1, x^n_1, u^n_{i+1}) = q_1(u^n_1, x^n_1) q_2(u^n_{i+1})
\]
\[
\Rightarrow q(x^n_1|u^n_1, u^n_{i+1}) = \frac{q_1(u^n_1, x^n_1) q_2(u^n_{i+1})}{q_1(u^n_1) q_2(u^n_{i+1})}
\]
\[
= q_1(x^n_1|u^n_1)
\]
\[
\Rightarrow q(x^n_1|u^n_1, s^0_0(n)) = q_1(x^n_1|u^n_1, s^0_0(n))
\]
\[
\Rightarrow I(X^n_1; Y^n_1|U^n_1, U^n_{i+1}, s^0_0(n))_{q_1, q_2} = H(X^n_1|U^n_1, s^0_0(n)) - H(X^n_1|Y^n_1, U^n_1, U^n_{i+1}, s^0_0(n))
\]
\[
\geq H(X^n_1|U^n_1, s^0_0(n)) - H(X^n_1|Y^n_1, U^n_1, s^0_0(n))
\]
\[
= I(X^n_1; Y^n_1|U^n_1, s^0_0(n))_{q_1, q_2}
\]
\[
\geq I(X^n_1; Y^n_1|U^n_1, s^0_0(l))_{q_1, q_2}.
\]

In conclusion,
\[
nF_n(\lambda) \geq I(U^n_1; Z^n_1|s^0_0(l))_{q_1} + I(U^n_{i+1}; Z^n_{i+1}|Z^n_{i+1}, s^0_0(n))_{q_1, q_2}
\]
\[
+ \lambda \left( I(X^n_1; Y^n_1|U^n_1, s^0_0(l))_{q_1} + I(X^n_1, X^n_{i+1}; Y^n_{i+1}|U^n_1, U^n_{i+1}, Y^n_{i+1}, s^0_0(n))_{q_1, q_2} \right)
\]
\[
+ I(X^n_{i+1}; Y^n_{i+1}|X^n_1, U^n_1, U^n_{i+1}, Y^n_{i+1}, s^0_0(n))_{q_1, q_2}
\]
\[
- (1 + \lambda) \log_2 ||S||.
\]
Now, note that
\[
p(u_{i+1}^n, z_1^l | s_i, s_0^*(n)) = q_2(u_{i+1}^n)p(z_1^l | s_i, s_0^*(n))
\]

\[
\Rightarrow p(u_{i+1}^n | z_1^l, s_i, s_0^*(n)) = q_2(u_{i+1}^n).
\]

Hence
\[
I(U_{i+1}^n; Z_{i+1}^n | Z_i^l, s_0^*(n))_{q_1 q_2} \overset{(a)}{\geq} I(U_{i+1}^n; Z_{i+1}^n | Z_i^l, s_0^*(n))_{q_1 q_2} - \log_2 ||S||
\]
\[
= H(U_{i+1}^n | Z_i^l, s_0^*(n)) - H(U_{i+1}^n | Z_i^l, Z_{i+1}^n, s_0^*(n)) - \log_2 ||S||
\]
\[
= H(U_{i+1}^n) - H(U_{i+1}^n | Z_i^l, Z_{i+1}^n, s_0^*(n)) - \log_2 ||S||
\]
\[
\geq H(U_{i+1}^n | Z_{i+1}^n, s_0^*(n) )_{q_2} - H(U_{i+1}^n | Z_{i+1}^n, Z_{i+1}^n, s_0^*(n) ) - \log_2 ||S||
\]
\[
= I(U_{i+1}^n; Z_{i+1}^n | s_i, s_0^*(n))_{q_2} - \log_2 ||S||
\]
\[
\overset{(b)}{\geq} I(U_{i+1}^n; Z_{i+1}^n | s_i = s_0^*(m))_{q_2} - \log_2 ||S||
\]

where (a) follows from Lemma A.1 and (b) is because \(s_0^*(m)\) is the minimizing state. We also have
\[
p(x_{i+1}^n, u_1^l, u_{i+1}^n, y_1^l | s_i, s_0^*(n)) = q_2(x_{i+1}^n, u_1^l, u_{i+1}^n | s_i, s_0^*(n))p(y_1^l | s_i, s_0^*(n))
\]
\[
= q_2(x_{i+1}^n, u_1^l, u_{i+1}^n | s_i, s_0^*(n))q_2(u_{i+1}^n | s_i, s_0^*(n))p(y_1^l | s_i, s_0^*(n))
\]
\[
= q_2(x_{i+1}^n, u_1^l, u_{i+1}^n | s_i, s_0^*(n))q_2(u_{i+1}^n | s_i, s_0^*(n))
\]
\[
\Rightarrow p(x_{i+1}^n, u_1^l, u_{i+1}^n, y_1^l | s_i, s_0^*(n)) = q_2(x_{i+1}^n, u_1^l, u_{i+1}^n, s_i, s_0^*(n)).
\]

Therefore,
\[
I(X_{i+1}^n; Y_{i+1}^n | U_1^l, U_{i+1}^n, Y_1^l, s_0^*(n))_{q_1 q_2} \overset{(a)}{\geq} I(X_{i+1}^n; Y_{i+1}^n | U_1^l, U_{i+1}^n, Y_1^l, s_0^*(n)) - \log_2 ||S||
\]
\[
= H(X_{i+1}^n | U_1^l, U_{i+1}^n, Y_1^l, S_i, s_0^*(n)) - H(X_{i+1}^n | Y_{i+1}^n, U_1^l, U_{i+1}^n, Y_1^l, S_i, s_0^*(n))
\]
\[
- \log_2 ||S||
\]
\[
= H(X_{i+1}^n | U_1^l, U_{i+1}^n, S_i, s_0^*(n)) - H(X_{i+1}^n | Y_{i+1}^n, U_1^l, U_{i+1}^n, Y_1^l, S_i, s_0^*(n))
\]
\[
- \log_2 ||S||
\]
\[
\geq H(X_{i+1}^n | U_1^l, U_{i+1}^n, S_i, s_0^*(n)) - H(X_{i+1}^n | Y_{i+1}^n, U_{i+1}^n, S_i, s_0^*(n)) - \log_2 ||S||
\]
\[
= I(X_{i+1}^n; Y_{i+1}^n | U_1^l, U_{i+1}^n, S_i, s_0^*(n))_{q_2} - \log_2 ||S||
\]
\[
\overset{(b)}{\geq} I(X_{i+1}^n; Y_{i+1}^n | U_{i+1}^n, s_i = s_0^*(m))_{q_2} - \log_2 ||S||.
\]
So, finally,

\[ nF_n(\lambda) \geq lF_l(\lambda) + I(U_{l+1}^n; Z_{l+1}^n | Z_l^1, s_0^l(n))_{q_2, q} + \lambda I(X_{l+1}^n; Y_{l+1}^n | U_1^l, U_{l+1}^n, Y_1^l, s_0^l(n))_{q_2, q_2} \]

\[ \geq lF_l(\lambda) + I(U_{l+1}^n; Z_{l+1}^n | s_l = s_0^l(m))_{q_2} + \lambda I(X_{l+1}^n; Y_{l+1}^n | U_{l+1}^n, s_l = s_0^l(m))_{q_2} - (1 + \lambda) \log_2 ||S|| \]

\[ = lF_l(\lambda) + mF_{m}(\lambda), \]

where we used the fact that, due to stationarity of the channel when conditioned on the initial state, we have

\[ mF_{m}(\lambda) = I(U_{l+1}^n; Z_{l+1}^n | s_l = s_0^l(m))_{q_2} + \lambda I(X_{l+1}^n; Y_{l+1}^n | U_{l+1}^n, s_l = s_0^l(m))_{q_2} - (1 + \lambda) \log_2 ||S||. \]

We note that this is the most critical point in this proof: 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 \( p(x_{l+1}^n, y_{l+1}^m | x_{l+1}^n, s_l = s_0) = p(x_{l+1}^n, y_{l+1}^m | x_{l+1}^n, s_0) \) (follows from Equation (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 both \( \frac{1}{n} I(U^n; Z^n | s_0) \) and \( \frac{1}{n} I(X^n; Y^n | U^n, s_0) \) are bounded, independent of \( n \). To see this for \( \frac{1}{n} I(U^n; Z^n | s_0) \) for example, consider the following set of inequalities:

\[ \frac{1}{n} I(U^n; Z^n | s_0) \leq \frac{1}{n} \sum_{i=1}^{n} H(Z_i | s_0) \leq \log_2 ||Z||, \]

since all the \( Z_i \)'s are defined over the same alphabet \( Z_i \equiv Z \). Similarly

\[ \frac{1}{n} I(X^n; Y^n | U^n, s_0) \leq \log_2 ||X||, \]

and thus

\[ F_n(\lambda) = \max_{p(u^n, x^n)} \left\{ \min_{s_0 \in S} \frac{1}{n} I(U^n; Z^n | s_0) + \lambda \min_{s_0 \in S} \frac{1}{n} I(X^n; Y^n | U^n, s_0) \right\} - (1 + \lambda) \frac{\log_2 ||S||}{n} \]

\[ \leq \log_2 ||Z|| + \lambda \log_2 ||X|| < \infty. \]

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 \( n^8 \):

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

**APPENDIX E**

**Proof of Theorem 2 (The Indecomposable FSBC)**

A. Preliminaries

First, let us define:

\[
\overline{F}_n(\lambda) = \max_{p(u^n, x^n)} \frac{1}{n} \left[ \max_{s_0 \in S} I(U^n; Z^n|s_0) + \lambda \max_{s'_0 \in S} I(X^n; Y^n|U^n, s'_0) \right]
\]

\[
F_n(\lambda) = \max_{p(u^n, x^n)} \frac{1}{n} \left[ \min_{s_0 \in S} I(U^n; Z^n|s_0) + \lambda \min_{s'_0 \in S} I(X^n; Y^n|U^n, s'_0) \right] \triangleq C_n(\lambda)
\]

\[
G_n(\lambda) = \max_{p(u^n, x^n)} \frac{1}{n} [I(U^n; Z^n) + \lambda I(X^n; Y^n|U^n)].
\]

We have with the following lemma:

**Lemma E.1.** The following relationships hold

\[
\overline{F}_n(\lambda) \geq G_n(\lambda)
\]

\[
F_n(\lambda) \leq G_n(\lambda) + \frac{\log_2 ||S||}{n}.
\]

Before proving Lemma E.1 we note that it implies that

\[
\lim_{n \to \infty} F_n(\lambda) \leq \lim_{n \to \infty} G_n(\lambda) \leq \lim_{n \to \infty} \overline{F}_n(\lambda),
\]

if the limits exist. In Appendix D we showed that the limit \( \lim_{n \to \infty} F_n(\lambda) \) exists. The fact that for the indecomposable channel \( \lim_{n \to \infty} F_n(\lambda) \) exists will follow from the proof of Theorem 2. We now provide the proof of Lemma E.1

**Proof:** Let \( p(s_0) \) denote the (unknown) distribution of the initial channel state. This distribution, naturally, depends on the past, but it does not matter for the sake of this proof whether it is stationary or not, only that it exists.

---

\( ^8 \)Here we use [5, Lemma 4A.1] which states that for a sequence \( \{a_n\}_{n \in \mathbb{N}} \), with \( \bar{a} = \sup_n a_n < \infty \), if for all \( n \geq 1 \) and all \( N > n \), \( Na_N \geq na_n + (N - n)a_{n-n} \), then \( \lim_{N \to \infty} a_N = \bar{a} \).

April 30, 2008
Now consider the first statement: let \( p_G(u^n, x^n) \) be the maximizing distribution for \( G_n(\lambda) \). Then,

\[
\nF_n(\lambda) \geq \max_{s_0} I(U^n; Z^n|s_0) + \lambda \max_{s_0} I(X^n; Y^n|U^n, s'_0)
\]

\[
= \left[ \max_{s_0} (H(U^n|s_0) - H(U^n|Z^n, s_0)) + \lambda \max_{s_0} (H(X^n|U^n, s'_0) - H(X^n|Y^n, U^n, s'_0)) \right]_{p_G}
\]

\[
= [H(U^n) - \min_{s_0} H(U^n|Z^n, s_0) + \lambda \left( H(X^n|U^n) - \min_{s_0} H(X^n|Y^n, U^n, s'_0) \right)]_{p_G}
\]

\[
(a) \quad [H(U^n) - \sum_{s_0 \in S} p(s_0) \min_{s_0} H(U^n|Z^n, s_0) + \lambda \left( H(X^n|U^n) - \sum_{s'_0 \in S} p(s'_0) H(X^n|Y^n, U^n, s'_0) \right)]_{p_G}
\]

\[
= [H(U^n) - H(U^n|Z^n, s_0) + \lambda (H(X^n|U^n) - H(X^n|Y^n, U^n, s_0))]_{p_G}
\]

\[
\geq [H(U^n) - H(U^n|Z^n) + \lambda (H(X^n|U^n) - H(X^n|Y^n, U^n))]_{p_G}
\]

\[
= [I(U^n; Z^n) + \lambda I(X^n; Y^n|U^n)]_{p_G}
\]

\[
\triangleq n G_n(\lambda),
\]

where in (a) we plugged the unknown \( p(s_0) \).

For the second statement, let \( p(u^n, x^n) \) denote the maximizing distribution for \( F_n(\lambda) \) and write:

\[
\nF_n(\lambda) = \left[ \min_{s_0} I(U^n; Z^n|s_0) + \lambda \min_{s'_0} I(X^n; Y^n|U^n, s'_0) \right]_{\mathcal{P}}
\]

\[
= \left[ H(U^n) - \sum_{s_0 \in S} p(s_0) \max_{s_0} H(U^n|Z^n, s_0) + \lambda \left( H(X^n|U^n) - \sum_{s'_0 \in S} p(s'_0) \max_{s'_0} H(X^n|Y^n, U^n, s'_0) \right) \right]_{\mathcal{P}}
\]

\[
\leq \left[ H(U^n) - \sum_{s_0 \in S} p(s_0) H(U^n|Z^n, s_0) + \lambda \left( H(X^n|U^n) - \sum_{s'_0 \in S} p(s'_0) H(X^n|Y^n, U^n, s'_0) \right) \right]_{\mathcal{P}}
\]

\[
= [H(U^n|S_0) - H(U^n|Z^n, s_0) + \lambda (H(X^n|U^n, S_0) - H(X^n|Y^n, s_0))]_{\mathcal{P}}
\]

\[
= [I(U^n; Z^n|S_0) + \lambda I(X^n; Y^n|U^n, S_0)]_{\mathcal{P}}
\]

\[
(a) \quad [I(U^n; Z^n) + \lambda I(X^n; Y^n|U^n)]_{\mathcal{P}} + (1 + \lambda) \log_2 ||S||
\]

\[
\leq n G_n(\lambda) + (1 + \lambda) \log_2 ||S||.
\]

where (a) is due to Lemma A.1.
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 \( F_n(\lambda) \). Let \( E_n(\lambda, p(u^n, x^n)) \) be the expression of \( F_n(\lambda) \) when the distribution is specified (i.e. the maximization over \( p(u^n, x^n) \) is dropped) and let \( s_0 \) and \( s_0' \) be the states that minimize, respectively, the first and second mutual information expressions in \( E_n(\lambda, \tilde{p}(u^n, x^n)) \) (recall that \( \tilde{p}(u^n, x^n) \) is the p.m.f. that maximizes \( F_n(\lambda) \)). Then,

\[
F_n(\lambda) \geq E_n(\lambda, \tilde{p}(u^n, x^n)).
\]

Let \( m_1 \) be fixed, and in addition denote

\[
m_2 = n - m_1
\]

\[
u_1 \triangleq u^{m_1} \quad u_2 \triangleq u_{m_1+1}^n
\]

\[
x_1 \triangleq x^{m_1} \quad x_2 \triangleq x_{m_1+1}^n
\]

\[
y_1 \triangleq y^{m_1} \quad y_2 \triangleq y_{m_1+1}^n
\]

\[
z_1 \triangleq z^{m_1} \quad z_2 \triangleq z_{m_1+1}^n.
\]

First consider

\[
p(z_1, s_{m_1}, u_1, u_2, \tilde{s}_0) = \sum_{x_{m_1}, y_{m_1}, z_{m_1}^{-1}} p(x_1, y_1, s_1, z_1 | u_1, u_2, \tilde{s}_0) = \sum_{x_{m_1}, y_{m_1}, z_{m_1}^{-1}} p(y_1, s_1, z_1 | x_1, u_1, u_2, \tilde{s}_0)p(x_1 | u_1, u_2, \tilde{s}_0)
\]

\[
= \sum_{x_{m_1}, y_{m_1}, z_{m_1}^{-1}} p(y_1, s_1, z_1 | x_1, \tilde{s}_0)p(x_1 | u_1, u_2, \tilde{s}_0)
\]

\[
= \sum_{x_{m_1}} p(z_1, s_{m_1}, x_1, \tilde{s}_0)p(x_1 | u_1, u_2, \tilde{s}_0),
\]

where (a) follows from Equation (2). Therefore, \( p(z_1, s_{m_1}, u_1, u_2, \tilde{s}_0) \) is not independent of \( U_2 \), since \( U_2 \) can be correlated with \( X_1 \). This is in contrast to 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:

\[
p(z_2|z_1, x_1, u_1, u_2, s_{m_1}, \tilde{s}_0) = \sum_{x_2} p(z_2, x_2|z_1, x_1, u_1, u_2, s_{m_1}, \tilde{s}_0)
\]

\[
= \sum_{x_2} p(x_2|z_1, x_1, u_1, u_2, s_{m_1}, \tilde{s}_0)p(z_2|z_1, x_1, x_2, u_1, u_2, s_{m_1}, \tilde{s}_0)
\]

\[
\overset{(a)}{=} \sum_{x_2} p(x_2|z_1, x_1, u_1, u_2)p(z_2|z_1, x_1, x_2, s_{m_1}, \tilde{s}_0)
\]

\[
\overset{(b)}{=} \sum_{x_2} p(x_2|z_1, x_1, u_1, u_2)p(z_2|x_2, s_{m_1})
\]

\[
= \sum_{x_2} p(x_2|z_1, x_1, u_1, u_2, s_{m_1})p(z_2|z_1, x_1, x_2, s_{m_1}, u_1, u_2)
\]

\[
= p(z_2|z_1, x_1, u_2, u_1),
\]

(E.1)

where (a) is because there is no feedback and (b) is because given \(S_{m_1}, Z_2\) is independent of the past.

Now expand \(I(U_1, U_2; Z_1, Z_2|\tilde{s}_0)\) in \(F_n(\lambda)\) as:

\[
I(U_1, U_2; Z_1, Z_2|\tilde{s}_0) = I(U_1, U_2; Z_1|\tilde{s}_0) + I(U_1, U_2; Z_2|Z_1, \tilde{s}_0)
\]

\[
\leq m_1 \log_2 ||Z|| + I(U_1, U_2; Z_2|Z_1, \tilde{s}_0)
\]

\[
\overset{(a)}{\leq} m_1 \log_2 ||Z|| + \log_2 ||S|| + m_1 \log_2 ||X|| + I(U_1, U_2; Z_2|Z_1, s_{m_1}, X_1, Z_1, \tilde{s}_0)
\]

where (a) is due to Lemma A.1. Similarly we obtain

\[
I(U^n; Z^n|s_0) = I(U_1, U_2; Z_1, Z_2|s_0)
\]

\[
\geq - \log_2 ||S|| - m_1 \log_2 ||X|| + I(U_1, U_2; Z_2|s_{m_1}, X_1, Z_1, s_0).
\]
Therefore,

\[
\frac{1}{n} (I(U^n; Z^n|\tilde{s}_0) - I(U^n; Z^n|s_0)) \\
\leq \frac{1}{n} \left( m_1 \log_2 ||Z|| + 2 \log_2 ||S|| + 2m_1 \log_2 ||X|| \\
+ I(U_1, U_2; Z_2|S_{m_1}, X_1, Z_1, \tilde{s}_0) - I(U_1, U_2; Z_2|S_{m_1}, X_1, Z_1, s_0) \right) \\
\overset{(a)}{=} \frac{1}{n} \left( m_1 \log_2 ||Z|| + 2 \log_2 ||S|| + 2m_1 \log_2 ||X|| \\
+ \sum_{X^{m_1}} \left( p(s_{m_1}, x^{m_1}|\tilde{s}_0) - p(s_{m_1}, x^{m_1}|s_0) \right) I(U_1, U_2; Z_2|x^{m_1}, s_{m_1}) \right) \\
\leq \frac{1}{n} \left( m_1 \log_2 ||Z|| + 2 \log_2 ||S|| + 2m_1 \log_2 ||X|| \\
+ \sum_{X^{m_1}} p(x^{m_1}) \sum_S \left| p(s_{m_1}, x^{m_1}|\tilde{s}_0) - p(s_{m_1}, x^{m_1}|s_0) \right| I(U_1, U_2; Z_2|x^{m_1}, s_{m_1}) \right) \\
\overset{(b)}{=} \frac{1}{n} \left( m_1 \log_2 ||Z|| + 2 \log_2 ||S|| + 2m_1 \log_2 ||X|| + \sum_{X^{m_1}} p(x^{m_1}) \epsilon||S||(n - m_1) \log_2 ||Z|| \right) \\
\overset{n \to \infty}{=} \epsilon||S|| \log_2 ||Z||, \\
\]  

(E.2)

where (a) follows from (E.1) and (b) is due to Definition 2. Now, for \( I(X_1, X_2; Y_1, Y_2|U_1, U_2, \tilde{s}_0) \) we have:

\[
I(X_1, X_2; Y_1, Y_2|U_1, U_2, s'_0) \\
= I(X_1; Y_1, Y_2|U_1, U_2, s'_0) + I(X_2; Y_1, Y_2|X_1, U_1, U_2, s'_0) \\
\leq \log_2 ||X^{m_1}|| + I(X_2; Y_1|X_1, U_1, U_2, s'_0) + I(X_2; Y_2|X_1, U_1, U_2, s'_0) \\
\leq m_1 \log_2 ||X|| + m_1 \log_2 ||Y|| + I(X_2; Y_2|Y_1, X_1, U_1, U_2, s'_0) \\
\leq m_1 \log_2 ||X|| + m_1 \log_2 ||Y|| + \log_2 ||S|| + I(X_2; Y_2|S_{m_1}, Y_1, X_1, U_1, U_2, s'_0). \\
\]

Similarly

\[
I(X_1, X_2; Y_1, Y_2|U_1, U_2, s'_0) \geq - \log_2 ||S|| + I(X_2; Y_2|S_{m_1}, Y_1, X_1, U_1, U_2, s'_0). \\
\]
Thus,
\[
\frac{1}{n} (I(X^n; Y^n|U^n, s'_0) - I(X^n; Y^n|U^n, s_0)) - I(X^n; Y^n|U^n, s'_0) \]
\[
\leq \frac{1}{n} \left( m_1 \log_2 ||\mathcal{X}|| + m_1 \log_2 ||\mathcal{Y}|| + 2 \log_2 ||S|| + I(X_2; Y_2|S_{m_1}, Y_1, X_1, U_1, U_2, s'_0) - I(X_2; Y_2|S_{m_1}, Y_1, X_1, U_1, U_2, s_0) \right) \]
\[
= \frac{1}{n} \left( m_1 \log_2 ||\mathcal{X}|| + m_1 \log_2 ||\mathcal{Y}|| + 2 \log_2 ||S||
\right)
\[
+ \sum_{\mathcal{X}^m_1} p(x^{m_1}) \sum_{s_{m_1} \in \mathcal{S}} \left( p(s_{m_1}|x^{m_1}, s'_0) - p(s_{m_1}|x^{m_1}, s_0) \right) I(X_2; Y_2|U_2, U_1, s_{m_1}, x^{m_1}) \]
\[
\leq \frac{1}{n} \left( m_1 \log_2 ||\mathcal{X}|| + m_1 \log_2 ||\mathcal{Y}|| + 2 \log_2 ||S||
\right)
\[
+ \sum_{\mathcal{X}^m_1} p(x^{m_1}) \sum_{s_{m_1} \in \mathcal{S}} \left| p(s_{m_1}|x^{m_1}, s'_0) - p(s_{m_1}|x^{m_1}, s_0) \right| I(X_2; Y_2|U_2, U_1, s_{m_1}, x^{m_1}) \]
\[
\leq \frac{1}{n} \left( m_1 \log_2 ||\mathcal{X}|| + m_1 \log_2 ||\mathcal{Y}|| + 2 \log_2 ||S|| + \sum_{\mathcal{X}^m_1} p(x^{m_1}) \epsilon ||S|| (n - m_1) \log_2 ||\mathcal{X}|| \right)
\[
\overset{(b)}{\longrightarrow} \epsilon ||S|| \log_2 ||\mathcal{X}||,
\]
where (c) is because there is no feedback. Finally, (b) follows from the definition of the indecomposable channel (see Definition 2).

Combining (E.2) with (E.3) we obtain that if the limits exist then

\[
\lim_{n \to \infty} \left( F_n(\lambda) - \bar{F}_n(\lambda, \bar{p}(u^n, x^n)) \right) \leq \epsilon \|S\| (\log_2 \|Z\| + \lambda \log_2 \|\mathcal{X}\|)
\]

as \( F_n(\lambda) \geq F_n(\lambda, \bar{p}(u^n, x^n)) \) then, by the sandwich theorem also

\[
\lim_{n \to \infty} \left( F_n(\lambda) - \bar{F}_n(\lambda) \right) \leq \epsilon \|S\| (\log_2 \|Z\| + \lambda \log_2 \|\mathcal{X}\|).
\]

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

\[
\lim_{n \to \infty} F_n(\lambda) = \lim_{n \to \infty} \bar{F}_n(\lambda).
\]

As the right-hand-side exists and finite, this also proves that \( \lim_{n \to \infty} F_n(\lambda) \) exists and finite.

Together with Lemma E.1 this means that the capacity region of Theorem 1 does not depend on the initial state when the channel is indecomposable.

APPENDIX F

AN ALTERNATIVE APPROACH TO THE DEFINITION OF THE AVERAGE PROBABILITY OF ERROR

Consider the following definitions for the average probability of error and the corresponding achievable region:

**Definition F.1.** The average probability of error of a code for the FSBC is defined as

\[
P_e^{(n)} = \Pr \left( g_Y(Y^n) \neq (M_0, M_1) \text{ or } g_Z(Z^n) \neq (M_0, M_2) \right),
\]

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

**Definition F.2.** 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 \( P_e^{(n)} \leq \epsilon \).

For these definitions we no longer have a characterization of the capacity region for the general (i.e. non-indecomposable) FSBC, but only lower and upper bounds. This is summarized in Theorem F.1 below:

**Theorem F.1.** For the physically degraded FSBC \( \mathcal{X} \times \mathcal{S}, p(y, z, s|x, s'), \mathcal{Y} \times \mathcal{Z} \times \mathcal{S} \) defined in Definition 7, subject to Definition F.1, we have that
1) Every rate triplet \((R_0, R_1, R_2)\) that belongs to the set

\[
\mathcal{C} = \lim_{n \to \infty} \mathcal{R}_n,
\]

is achievable, and the limit exists.

2) Every achievable rate triplet \((R_0, R_1, R_2)\) must belong to the set

\[
\mathcal{C} = \liminf_{n \to \infty} \mathcal{R}_n,
\]

and the limit exists.

**Proof Outline:** The achievability is the same as for Theorem 1. The converse is proved in Appendix F-A.

For the stochastically degraded channel we obtain the following corollary:

**Corollary F.1.** For the stochastically degraded FSBC of Definition 8, the capacity region subject to Definition F.1 is bounded from below by \(\mathcal{C}\) defined in (F.1) and from above by \(\mathcal{C}\) defined in (F.2), where \(p(y, z, s|x, s')\) is replaced by \(p(y, s|x, s')\tilde{p}(z|y)\), such that Equation (10) is satisfied.

For the indecomposable FSBC, Theorem 2 holds also under Definition F.1.

**A. A Converse for the General FSBC Using Definition F.1**

The converse for Definition F.1 is more complicated than in Appendix B since we cannot show that for

\[
\Delta^n(\lambda) \triangleq \max_{p(u^n, x^n)} \left\{ \frac{1}{n} I(U^n; Z^n) + \lambda \frac{1}{n} I(X^n; Y^n|U^n) \right\},
\]

\[
\lim_{n \to \infty} \Delta^n(\lambda)
\]

exists for the general FSBC. This is because without conditioning on the initial state, the non-indecomposable channel is non-stationary. We therefore have to follow a different approach. To that aim, we first recall that a rate pair \((R_1, R_2)\) is achievable by Definition F.1 if for any \(\epsilon > 0\), \(\delta > 0\), there exists \(n(\epsilon, \delta) \in \mathbb{N}\) such that for every \(n > n(\epsilon, \delta)\), an \((R_1 - \delta, R_2 - \delta, n)\) code with \(P_e^{(n)} \leq \epsilon\) can be constructed. It follows that these codes satisfy \(\sup_{n > n(\epsilon, \delta)} P_e^{(n)} \leq \epsilon\). Since we can pick \(n(\epsilon, \delta)\) arbitrarily large, then we conclude that for a rate pair to be achievable it must hold that there exists a sequence of \((R_1 - \delta, R_2 - \delta, n)\) codes such that \(\lim_{n(\epsilon, \delta) \to \infty} \sup_{n > n(\epsilon, \delta)} P_e^{(n)} \leq \epsilon\). By definition this can also be written as

\[
\limsup_{n \to \infty} P_e^{(n)} \leq \epsilon.
\]
If there does not exist a sequence of codes for which (F.4) holds, it is not possible to find an \( n(\epsilon, \delta) \) that satisfies Definition F.1 and the rate pair is not achievable. Using the condition (F.4), the statement of the converse for Definition F.1 is given in the following lemma:

**Lemma F.1.** If for some \( \lambda > 0 \),

\[
R_2 + \lambda R_1 > \liminf_{n \to \infty} \Delta^n(\lambda) + \epsilon,
\]

then

\[
\limsup_{n \to \infty} P_{e_2}^{(n)} R_2 + \lambda \limsup_{n \to \infty} P_{e_1}^{(n)} R_1 > \epsilon. \tag{F.5}
\]

Note that \( \liminf_{n \to \infty} \Delta^n(\lambda) \) exists and is finite.

**Proof:** Recall that \( P_{e_2}^{(n)} \) and \( P_{e_1}^{(n)} \) denote the probabilities of error when the initial state is not available at the receivers and transmitter. From Fano’s inequality (see [5, Equation 4.6.16]) we have

\[
H(M_2|Z^n) \leq P_{e_2}^{(n)} nR_2 + 1 \tag{F.6a}
\]

\[
H(M_1|Y^n) \leq P_{e_1}^{(n)} nR_1 + 1. \tag{F.6b}
\]

Now, note that

\[
nR_2 = H(M_2)
\]

\[
= H(M_2) - H(M_2|Z^n) + H(M_2|Z^n)
\]

\[
nR_2 - H(M_2|Z^n) = I(M_2; Z^n) \tag{F.7}
\]

\[
= I(U^n; Z^n)
\]

\[
nR_1 - H(M_1|Y^n) = H(M_1) - H(M_1|Y^n)
\]

\[
\leq H(M_1|M_2) - H(M_1|Y^n, M_2)
\]

\[
= I(M_1; Y^n|M_2)
\]

\[
= H(Y^n|M_2) - H(Y^n|M_1, M_2)
\]

\[
\leq H(Y^n|U^n) - H(Y^n|X^n, U^n)
\]

\[
= I(X^n; Y^n|U^n), \tag{F.8}
\]

where in (a) we defined \( U_i = M_2, \ i = 1, 2, ..., n \). This definition of \( U^n \) satisfies the Markov relationship
\( U^n | s'_0 \leftrightarrow X^n | s'_0 \leftrightarrow Y^n | s'_0 \). Combining both derivations we have that for our choice of \( U^n \):

\[
\begin{align*}
nR_2 - H(M_2 | Z^n) + \lambda (nR_1 - H(M_1 | Y^n)) \\
\leq I(U^n; Z^n) + \lambda I(X^n; Y^n | U^n) \\
\overset{(a)}{\leq} n\Delta^n(\lambda),
\end{align*}
\]

where (a) is because \( \Delta^n(\lambda) \) is obtained by maximizing over all joint distributions \( p(u^n, x^n) \) subject to the cardinality constraint (12) (we show in Appendix C-H that this is enough to achieve the maximum), hence we obtain

\[
H(M_2 | Z^n) + \lambda H(M_1 | Y^n) \geq n (R_2 + \lambda R_1 - \Delta^n(\lambda)).
\]

Combined with Fano’s inequalities (F.6), we have

\[
P_{e_2}^{(n)} nR_2 + 1 + \lambda \left( P_{e_1}^{(n)} nR_1 + 1 \right) \geq n \left( R_2 + \lambda R_1 - \Delta^n(\lambda) \right)
\]

\[
\Rightarrow P_{e_2}^{(n)} R_2 + \lambda P_{e_1}^{(n)} R_1 \geq R_2 + \lambda R_1 - \Delta^n(\lambda) - (1 + \lambda) \frac{1}{n}
\]

\[
\Rightarrow \sup_{n > n(\delta, \epsilon)} \left( P_{e_2}^{(n)} R_2 + \lambda P_{e_1}^{(n)} R_1 \right) \geq R_2 + \lambda R_1 - \inf_{n > n(\delta, \epsilon)} \left( \Delta^n(\lambda) + (1 + \lambda) \frac{1}{n} \right)
\]

\[
\Rightarrow \lim_{n \to \infty} \sup_{n > n(\delta, \epsilon)} \left( P_{e_2}^{(n)} R_2 + \lambda P_{e_1}^{(n)} R_1 \right) \geq R_2 + \lambda R_1 - \lim_{n \to \infty} \inf_{n > n(\delta, \epsilon)} \left( \Delta^n(\lambda) + (1 + \lambda) \frac{1}{n} \right)
\]

\[
\Rightarrow \limsup_{n \to \infty} P_{e_2}^{(n)} R_2 + \lambda \limsup_{n \to \infty} P_{e_1}^{(n)} R_1 \geq R_2 + \lambda R_1 - \liminf_{n \to \infty} \Delta^n(\lambda)
\]

\[
> \epsilon,
\]

which means that there will always be large values of \( n \) for which at least one of the receivers has a probability of error that is bounded away from zero, and therefore \((R_1, R_2)\) is not achievable. □

**References**


