Sequential Random Coding Error Exponents for Degraded Broadcast Channels

Cheng Chang and Anant Sahai
Dept. of EECS, University of California, Berkeley, CA, 94720
Email: cchang@eecs.berkeley.edu, sahai@eecs.berkeley.edu

Abstract

Discrete-time memoryless broadcast channels are an abstraction of the down-link for many centralized wireless systems. They capture some of the issues involved with many different users wanting to simultaneously receive information from a single site. The traditional analysis is in the context of block codes with the messages being known in advance by the encoder. Instead, we consider a sequential setting in which each user's message evolves in real-time as bits stream in to the encoder. In this context, we look at the probability of error not at the block level, but at the bit-level. Furthermore, in place of block-length, we look at the delay between when the bit arrives at the encoder and when it is decoded by the decoders.

The sequential random coding error exponent is studied and shown to be positive in the whole capacity region for degraded broadcast channels. Furthermore, we show that this exponent can be achieved in a delay-universal or "anytime" fashion in that the encoder does not have to specify the target delay. The choice of decoding delay is left up to the decoders which are free to vary this on an application specific basis — the longer it is willing to wait, the lower the probability of bit error will be.

I. Introduction

This paper treats degraded broadcast channels in a sequential setting. As such, it is a sister paper to our discussion of sequential codes for multiple-access channels [1] and to our discussion of sequential source coding in the Slepian-Wolf context.[2] The broad considerations presented there are identical to those present here, but we repeat a brief review here for the reader's convenience.

In the point-to-point communication scenario, there are many types of codes. The block-coding paradigm has a semi-infinite sequence of messages, each of which is quite large. These messages is assumed to be known to the encoders at the beginning of the epoch, and the decoder is assumed to produce an estimate for it by the end of an epoch. The next message is considered in the next epoch. The delay in this context is determined by the epoch size or block-length. The sequential-coding paradigm also has a semi-infinite sequence of messages, but each of these is assumed to be quite small relative to the delays of interest. These small messages become available to the encoder as time evolves, and are used to generate channel input symbols causally. There is no *a priori* choice of an epoch or block-length. Instead, the decoder decodes estimates of the messages as time goes on, but does so with some delay.

Convolutional and tree codes represent the most well known cases of sequential codes, though these techniques can also be used to construct block-codes. In [3], Forney reviews how the probability of error varies with delay in the point-to-point settings as the decoder is forced to make a bit-decision at a particular delay. The relevant exponent turns out to be the random block-coding error exponent, and in the high-rate regime, [4] tells us that no code can achieve a higher exponent with delay. In [5], Sahai further identifies the delay-universal or "anytime" variation on sequential codes in which the choice of the delay is left entirely up to the decoder. In [6], anytime communication problems are shown to be intimately connected to problems of automatic control. Because of this connection, multiuser anytime theory has the potential to help us answer architectural questions in the design of systems — in particular, the question of whether or not delay-sensitive applications like control need to be segregated from rate-sensitive applications.

Broadcast channels [7] and [8] are interesting for distributed wireless communication systems, and yet, the study of such channels has focused almost entirely on the block-coding case. The capacity region and random coding error exponents for block coding on degraded broadcast channels are explained in [9] and [10] respectively. Bounded delay decoding had not been considered. As pointed out in [11], the capacity region is still unknown for general broadcast channels.

In this paper, we study the sequential communication problem for degraded broadcast channels and bound the random coding error exponents for the problem. These exponents are considered in the context of delay-universal, or "anytime," reliability. The probability of error is required to go to zero exponentially with delay, where the delay is chosen entirely at the decoders. As such, the results here have implications in terms of sufficient conditions for the distributed stabilization of unstable linear systems which have a single sensor/observer but distributed actuators acting on different dimensions of the unstable system, but communicating to the observer through a broadcast channel [12].

In Section II we will first describe the model of degraded broadcast channels and sequential channel coding. Then in Section III-A we will derive the random coding bound on sequential coding for degraded broadcast channels. We close with a numeric example in Section IV.

II. Problem Setup

A. Broadcast channel

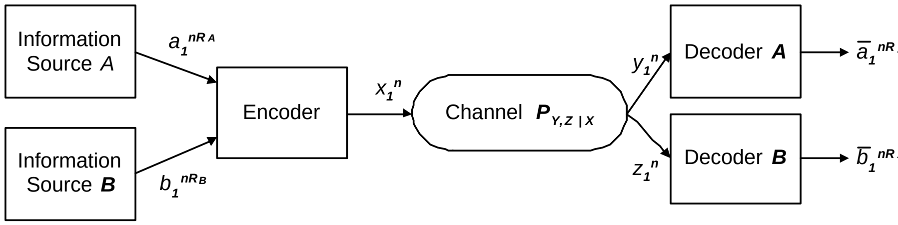

Fig. 1. Model of general broadcast channels

As shown in Fig 1, there are two independent information sources for two distinct decoders A and B. This setup in which the two information sources are independent is denoted by (K, II) in [13]. The broadcast channel is characterized by a transition probability matrix $P_{Y,Z|X}$ where $x \in \mathcal{X}$ is the channel input, $y \in \mathcal{Y}$ is the channel output to decoder A, and $z \in \mathcal{Z}$ is the channel output to decoder B. The capacity region (R_A, R_B) of general broadcast channels is still unknown [11].

B. Degraded Broadcast channel

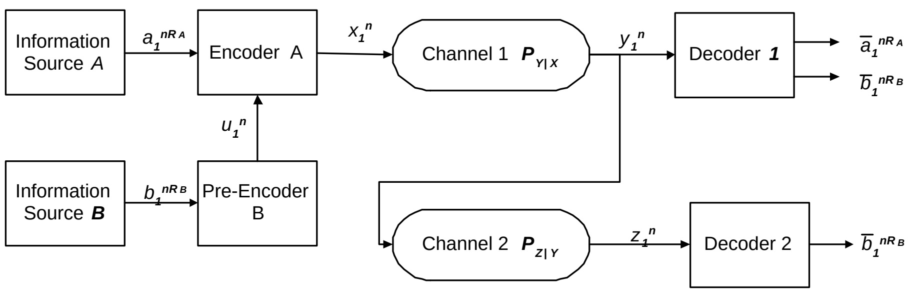

Fig. 2. Model of degraded broadcast channels and how to code over them. Since the two users can be ordered in terms of quality, we denote them 1 and 2 instead of A and B.

As shown in Fig 2, decoder 1 has a strictly better channel than decoder 2 in the sense that $P_{Y,Z|X}(y, z|x) = P_{Y|X}(y|x)P_{Z|Y}(z|y)$. Thus the message for decoder 2 can be decoded by decoder 1. The information source $\mathcal{A}$ and $\mathcal{B}$ remain independent.

Theorem 1: Capacity region for the degraded broadcast channel [9],[10],[7]:The capacity region $\mathcal{R}$ is the convex hull of the set of rate pairs (R_A, R_B), s.t. $0 \le R_B \le I(Z; U)$, $0 \le R_A \le I(X; Y|U)$, for some joint distribution $Q_U(u)Q_{X|U}(x|u)P_{Y,Z|X}(y, z|x)$, where the auxiliary random variable $|\mathcal{U}| \le \min\{|\mathcal{X}|, |\mathcal{Y}|, |\mathcal{Z}|\}$.

For block coding, the decoding error probability can be arbitrarily close to 0 in the capacity region. The converse is proved by Gallager in [10]. A *statistical degraded*[7] broadcast channel has the same marginal transition probabilities as a degraded broadcast channel, thus having the same capacity region and furthermore the same communication reliability. We will focus on the physically degraded broadcast channels in this paper, although all the results carry over directly to the statistically degraded case.

C. Random Block Coding Error Exponents

The random block-coding error exponent of the degraded broadcast channel is studied in [10]. Consider a random ensemble of $(n, 2^{nR_A}, 2^{nR_B})$ codes where encoder has 2^{nR_A} equal-probable messages from information source $\mathcal{A}$ and 2^{nR_B} equal-probable messages from information source $\mathcal{B}$, all with n channel uses. The inputs (length-n codewords) to the channel for message $(i_A, i_B) \in \{1, 2, \dots, 2^{nR_A}\} \times \{1, 2, \dots, 2^{nR_B}\}$, $\overrightarrow{x_{i_A,i_B}}$ is chosen according to the following *superposition rule*.

First choose 2^{nR_B} pre-codewords (*cloud centers*) $\overrightarrow{u_1}, \dots, \overrightarrow{u_{2^{nR_B}}}$ independently according to distribution Q_U, which has a product form in the case of memoryless channels:

$$P(\overrightarrow{u_i} = U_1^n) = \prod_{j=1}^{n} Q_U(U_j)$$

The codewords themselves are independently chosen according to a conditional distribution $Q_{X|U}$.

$$P(\overrightarrow{x_{i_A,i_B}} = X_1^n) = \prod_{j=1}^{n} Q_{X|U}(X_j|\overrightarrow{u_i}(j))$$

The decoders run ML decoding. Suppose that the two independent messages are i_A and i_B, Gallager showed the following coding theorem for degraded broadcast channels[10], except that we have changed units from nats to bits:

Theorem 2: [10] Expected error probability over all messages and all channel realizations. $\forall Q = (Q_U, Q_{X|U})$, for decoder 2:

$$P_2(\hat{i_B} \neq i_B) \leq 2^{-nE_{X12_B}(Q,R_B)}$$
$$E_{X12_B}(Q,R_B) = \max_{0\leq\rho\leq 1}[-\rho R_B - \log_2 \sum_z (\sum_u Q_U(u) P_{X12}^{\frac{1}{1+\rho}}(z|u))^{1+\rho}] \quad (1)$$

Where $P_{X12}(z|u)) = \sum_z \sum_y Q_{X|U}(x|u) P_{Y|X}(y|x) P_{Z|Y}(z|y)$ is the effective *channel* between pre-encoder B and decoder 2.

For decoder 1:

$$P_1(\hat{i_B} \neq i_B) \leq 2^{-nE_{X1_B}(Q,R_B)}$$
$$E_{X1_B}(Q,R_B) = \max_{0\leq\rho\leq 1}[-\rho R_B - \log_2 \sum_y (\sum_u Q_U(u) P_{X1}^{\frac{1}{1+\rho}}(y|u))^{1+\rho}] \quad (2)$$
$$P_1(\hat{i_A} \neq i_A) \leq 2^{-nE_{X1B}(Q,R_B)} + 2^{-nE_{1_A}(Q,R_A)}$$
$$E_{1_A}(Q,R_A) = \max_{0\leq\rho\leq 1}[-\rho R_A - \log_2 \sum_u Q_U(u) \sum_z (\sum_y Q_{X|U}(x|u) P_{Y|X}^{\frac{1}{1+\rho}}(y|x))^{1+\rho}] \quad (3)$$

Where $P_{X1}(y|u)) = \sum_x Q_{X|U}(x|u) P_{Y|X}(y|x)$ is the effective *channel* between pre-encoder B and decoder 1.

For rate pair (R_A, R_B) inside the capacity region: $R_A < I(X:Y|U), R_B < I(Z;U)$, the error exponents are larger than 0. As shown in [10], the proof of the above theorem closely parallels Chap.5 of [14].

D. Sequential Channel Coding

The model of sequential channel coding is as follows. For rate 1, the information bit $b_i \in \{0,1\}$ arrives at the encoder at time i, $i = 1, 2, \ldots, n, \ldots$. The encoder must send out an encoded symbol right away, that is only allowed to depend on the past bits b_1^i in a causal fashion. In order to achieve any rational rate $R = \frac{B}{C}$, we encode B information bits into C channel uses. The decoder can choose a decision time and hence delay for a certain information bit, and the expected probability of decoding error must be bounded by an exponential function of the delay.

Sequential encoders in this context can be thought of as labeled trees whose branching structure is determined by the desired rate and whose labels represent the channel inputs. A random sequential code is one in which the labels are chosen to be iid according to some distribution.

Definition 1: $(B, C, Q_U, \mathcal{U})$ random sequential coding scheme: For a distribution Q_U defined on the alphabet $\mathcal{U}$, and integers $B, C \in \mathcal{N}$. The encoder $\mathcal{E}$ is a sequence of maps $\{\mathcal{E}_j\}, j = 1, 2, \ldots$. The outputs of $\mathcal{E}_j$ are the outputs of the encoder from time $(j-1)C+1$ to jC. The rate of the encoder in bits per channel use is $R_B = \frac{B}{C}$.

$$\mathcal{E}_j : \{0,1\}^{jB} \times [0,1]^C \longrightarrow \mathcal{U}^C$$
$$\mathcal{E}_j(b_1^{jB}, \lambda_1^C(j, b_1^{jB})) = u_{(j-1)C+1}^{jC}$$

Where the $\lambda_i(j, b_1^{jB}), i = 1, 2, .., C; b_1^{jB} = 0, 1, \ldots, 2^{jB}-1; j = 1, 2, \ldots$ are iid continuous random variables uniformly distributed in $[0,1]$ and independent of the data bits. These represent the randomness used to generate the code and they are used to select the channel inputs according to:

$$u_{(j-1)C+i} = \mathcal{U}_k, i = 1, 2, \ldots, C$$

whenever $\sum_{l=1}^{k-1} Q_U(\mathcal{U}_l) < \lambda_i(j, b_1^{jB}) \leq \sum_{l=1}^{k} Q_U(\mathcal{U}_l)$

The random $\{\lambda_i(j, b_1^{jB})\}$ are considered common randomness known at both the encoder and decoder. By the construction of the codebook, it is clear that the encoder is *casual*, i.e. for any two information sequences b_1^{nB} and $\tilde{b}_1^{mB}$, if for some $k \leq \min(n, m)$ s.t. $b_1^{kB} = \tilde{b}_1^{kB}$, then the first kC outputs of the encoder for both information sequences are the same.

Forney showed in [3] that the point-to-point random coding error exponent defined in [14] can be achieved in the sequential setup by forcing the decoder to give its best decision with a constrained delay. The probability of ML decoding error of the i'th information source block at time nC is

$$Pe_n(i) \leq K 2^{-(n-i)CE_r(R)} \quad (4)$$

Where K is a constant and $E_r(R)$ is the random coding bound defined in [14]. Jelinek also showed that the same bound is achievable using a computationally tractable approximation to ML known as sequential decoding [15].

E. Superposition of Sequential Codes

For sequential coding over degraded broadcast channels, we will use a sequential *superposition codes* to achieve positive anytime reliabilities inside the capacity region. To motivate what is needed, we first revisit the block coding setting.

Recall that in the block coding setting, the independence properties that appear in the proof are:

- Two codewords are independent of each other if they have different B messages, regardless of whether the A messages are the same or not.
- Two codewords are conditionally independent of each other if they have different A messages, but the same B message.

The reason that this is called a superposition code is that it can be thought of as two distinct random codebooks that are then combined using a special letter-by-letter superposition operation. Explicitly, let the first random codebook (for the cloud centers) be a random rate R_B block code with letters drawn over $\mathcal{U} \times [0,1]$ according to $Q_U \times \text{Unif}[0,1]$. Thus, in addition to the traditional auxiliary random variable U, we have added a special continuous uniform random variable V. The second codebook is a random rate R_A block code with letters drawn over $[0,1]$ according to $\text{Unif}[0,1]$ alone. We can consider this an auxiliary random variable W. The letter by letter combining operation is given as follows:

$$X_i = G(V_i + W_i \bmod 1, U_i) \tag{5}$$

where

$$G(s,u) = \mathcal{X}_k \text{ if } \sum_{l=1}^{k-1} Q_{X|U}(\mathcal{X}_l|u) < s < \sum_{l=1}^{k} Q_{X|U}(\mathcal{X}_l|u)$$

Notice that G is exactly the function we would use to simulate something drawn from $Q_{X|U}$ given u and the realization of an independent uniform random variable s. This is just a matter of using the inverse conditional CDF. Moreover, notice that V_i, W_i are both independent $\text{Unif}[0,1]$ random variables and so $V_i + W_i \bmod 1$ is also a $\text{Unif}[0,1]$ random variable with the special property that if either of them is replaced with an independent uniform, the result will also be independent.[1] Thus, this letter by letter superposition of two different codebooks results in a broadcast codebook with the desired independence properties. The codewords with different cloud centers are independent, while those within a given cloud are conditionally independent.

At this point, it should be clear what needs to be done in the sequential setting. The pre-code for source B is the privileged one, and hence is a rate R_B random sequential code labeled with independent auxiliary random variables U and V for each letter, with distributions Q_U and $\text{Unif}[0,1]$ respectively. We use another sequential pre-code for source A at rate R_A that is labeled with independent auxiliary random variables W drawn from $\text{Unif}[0,1]$. These two sequential pre-codes are combined on a letter by letter basis according to (5).

By construction, it is obvious that the encoder is *casual*, i.e. for any two information sequences (a_1^{nA}, b_1^{nB}) and $(\tilde{a}_1^{mA}, \tilde{b}_1^{nB})$, if for some $k \leq \min(n,m)$ s.t. $(a_1^{kA}, , b_1^{kB}) = (\tilde{a}_1^{kA}, \tilde{b}_1^{kB}))$, then the first kC outputs of the encoder for both information sequence are the same. Furthermore, it is easily verified that the following properties hold:

- If $\tilde{b}_1^{kB} \neq b_1^{kB}$ at any point, then all hypothesized channel inputs $\tilde{X}_l$ are pairwise independent to the true X_l for all $l > kC$.
- If $\tilde{a}_1^{kB} \neq a_1^{kB}$ at any point, then the hypothesized channel inputs $\tilde{X}_l$ are pairwise conditionally independent to the true X_l for all $l > kC$, as long as we condition on U_l.

These pairwise independence and conditional pairwise independence properties will enable us to compute a bound on the random coding anytime exponent.

F. Encoding and Decoding

The encoder and decoders work as follows. Encoder $\mathcal{E}_A$ uses an $(A, B, C, Q_{X|U}, Q_U, \mathcal{U}, \mathcal{X})$ random sequential superposition scheme. This implicitly uses an $(B, C, Q_U \times \text{Unif}[0,1], \mathcal{U} \times [0,1])$ pre-coder $\mathcal{E}_B$ to generate a rate R_B random sequential code book, as well as an $(A, C, \text{Unif}[0,1], [0,1])$ pre-coder $\mathcal{E}'_A$ to generate a rate R_A random sequential code book, which are then superposed using (5).

At time nC, decoder 1 has access to a sequence $y_1, \ldots y_{nC}$ while decoder 2 has a sequence $z_1, \ldots z_{nC}$. Decoder 1 runs maximum-likelihood (ML) decoding to estimate both the information sequences. The decoded information sequence pair is denoted $(\hat{a}_1^{nA}, \hat{b}_1^{nB})$ and represents the ML sequence based on the channel outputs y seen so far. Decoder 2 on the other hand, is only interested in source $\mathcal{B}$ and so does ML decoding only at the level of the pre-code, and treats the superposition step as the additional noisy memoryless channel $Q_{X|U}$. This parallels what Gallager did in the error exponent derivation for block broadcast coding[10]. Mathematically:

[1] Of course, using continuous uniform random variables V and W is overkill. All we need is enough randomness.

$$\tilde{b}_1^{nB} = \arg\max_{t_1^{nB}} P_{X12}(z_1^{nC}|\mathcal{E}_B(t_1^{nB})) \tag{6}$$

Where P_{X12} is the effective *channel* between decoder 2 and information source $\mathcal{B}$.

$$P_{X12}(z|u)) = \sum_y \sum_x Q_{X|U}(x|u) P_{Y|X}(y|x) P_{Z|Y}(z|y)$$

.

Decoder 1 estimates both $\mathcal{A}$ and $\mathcal{A}$, for information source $\mathcal{B}$, it is tempting to again can treat encoder $\mathcal{E}_A$ as a channel. Thus:

$$\hat{b}_1^{nB} = \arg\max_{t_1^{nB}} P_{X1}(y_1^{nC}|\mathcal{E}_B(t_1^{nB}))$$

Where P_{X1} is the effective *channel* between decoder 1 and information source $\mathcal{B}$. $P_{X1}(y|u)) = \sum_x Q_{X|U}(x|u) P_{Y|X}(y|x)$. While for block coding we can pick the estimation $\hat{b}_1^{nB}$ as the *cloud center*[10] and estimate information source $\mathcal{A}$ around that center, we can not do the same in the sequential setting since the error probabilities for different information bits are not the same since they are exponential with delay. The estimates of late information bits are not at all reliable. Thus we have to apply ML decoding on both information sources together.

$$(\hat{a}_1^{nA}, \hat{b}_1^{nB}) = \arg\max_{(s_1^{nA}, t_1^{nB})} P_{Y|X}(y_1^{nC}|\mathcal{E}_A(s_1^{nA}, t_1^{nB}))$$

Where $s_1^{nA} \in \{0,1\}^{nA}, t_1^{nB} \in \{0,1\}^{nB}$ are two binary information sequences, and $\mathcal{E}_A(s_1^{nA}, t_1^{nB}) \in \mathcal{X}^{nC}, \mathcal{E}_B(t_1^{nB}) \in \mathcal{U}^{nC}$ are the outputs of the encoder and the pre-code encoder. Notice that at each time, the decoders re-estimate all the information bits. In the next section, we show that the expected probability of decoding error of the information bits decays exponentially with delay.

III. Random Coding Bound

In this section we derive a bound on the error probability for the sequential degraded broadcast channels using the randomized encoders in Definition 1 and the superposition strategy given earlier. It is clear that the encoders do not have any target delay specified, and so we are interested in the error probability of an information bit given some arbitrary decoding delay. We first state the main result of this paper in Theorem 3 and then discuss the proof.

Theorem 3: After nC channel uses, we have the ML decoded information bits $\tilde{b}_1^{nB}$ at decoder 2 and $(\hat{a}_1^{nA}, \hat{b}_1^{nB})$ at decoder 1. Write the error probabilities of decoding the j'th block of source $\mathcal{B}$ at decoder 2 and the j'th block of source A, B at decoder 1 as $P_n(\{\tilde{b}_{jB+1}^{(j+1)B} \neq b_{jB+1}^{(j+1)B}\})$, $P_n(\{\hat{a}_{jA+1}^{(j+1)A} \neq a_{jA+1}^{(j+1)A}\})$, $P_n(\{\hat{b}_{jB+1}^{(j+1)B} \neq b_{jB+1}^{(j+1)B}\})$ respectively. Then for every $0 \leq j \leq n$:

$$P_n(\{\tilde{b}_{jB+1}^{(j+1)B} \neq b_{jB+1}^{(j+1)B}\}) \leq K_1 2^{-dCE_{X12_B}(Q,R_B)} \tag{7}$$

$$P_n(\{\hat{b}_{jB+1}^{(j+1)B} \neq b_{jB+1}^{(j+1)B}\}) \leq K_2 2^{-dCE_{X1_B}(Q,R_B)} \tag{8}$$

$$P_n(\{\hat{a}_{jA+1}^{(j+1)A} \neq a_{jA+1}^{(j+1)A}\}) \leq (K_3 d + K_4) 2^{-dCE_1(Q,R_A,R_B)} \tag{9}$$

Where K_1, K_2, K_3 and K_4 are constants not depending on n, j and $dC = (n+1-j)C$ is the decoding delay. The first two error exponents $E_{X12_B}(Q,R_B)$ and $E_{X1_B}(Q,R_B)$ are defined in Eqn(1) and Eqn(2) respectively. The third error exponent $E_1(Q,R_A,R_B)$ is defined as follows:

$$E_1(Q,R_A,R_B) = \inf_{\alpha \in [0,1]} \{ \sup_{\rho \in [0,1]} \{ \alpha(-\rho R_A + E_{0_A}(Q,\rho)) + (1-\alpha)(-\rho(R_A+R_B) + E_{0_{AB}}(Q,\rho)) \} \}$$

$$E_{0_A}(Q,\rho) = -\log_2 \sum_u Q_U(u) \sum_y (\sum_x Q_{X|U}(x|u) P_{Y|X}^{\frac{1}{1+\rho}}(y|x))^{1+\rho} \tag{10}$$

$$E_{0_{AB}}(Q,\rho)) = -\log_2 \sum_y (\sum_x Q_X(x) P_{Y|X}^{\frac{1}{1+\rho}}(y|x))^{1+\rho} \tag{11}$$

Where $Q = (Q_U, Q_{X|U})$ are the distributions used to construct the sequential code. These error exponents can be made positive in the interior of the capacity region by properly choosing Q_U and $Q_{X|U}$. By letting $\alpha = 0, 1$, we can easily see that the sequential error exponents are not larger than those block coding error exponents defined in Theorem 2.

A. Error Probability for Information Source $\mathcal{B}$

The situation at decoder 2 is simple. By treating the output of the sequential pre-code for source A as a sequence of iid Uniform random variables, we can view the superposition operation as just another memoryless noisy channel. The optimal decoder will do at least as well as a decoder that ignores the structure of pre-code A. Verifying Eqn.7 and Eqn.8 is therefore simple since this is just a point-to-point problem in which we want to show that the random block coding error exponent is achieved by a random sequential code with bounded delay ML decoding. See [3], [5] for detailed proofs of this. It turns out that the pairwise independence property between the true suffix and any false suffix is all that is required since it allows the classical Gallager analysis of the random block coding error exponent to apply to the suffix.

B. Error Probability for Information Source $\mathcal{A}$

Information source $\mathcal{A}$ is only of interest to decoder 1. When we examine the situation from the perspective of decoder 1, we notice something interesting. The superposition operation of Eqn(5) is deterministic and works letter-by-letter. It is tempting therefore to combine the superposition operation with the memoryless, but random, noisy channel to give us a single memoryless multiple-access channel (MAC) connecting the outputs of the two sequential pre-codes to the channel output y. The construction of these two pre-codes was also done in a manner that is independent of each other and so the analysis of this part seems like it should reduce to a sequential MAC. The challenge that we face in applying the sequential MAC analysis of [1] is that the analysis there was given in terms of discrete input alphabets and the output of our pre-codes includes real-valued quantities. As a result, in this subsection we prove Eqn(9) directly using similar arguments.

When decoder 1 estimates information source $\mathcal{A}$, it has to do the joint ML decoding by estimating information source $\mathcal{A}$ and $\mathcal{B}$ simultaneously because of the sequential coding. After nC channel uses, depending on the position of the first wrongly decoded block of source $\mathcal{A}$ and $\mathcal{B}$, we have $(n+1)^2$ possible disjoint error events. We bound the probability of each error event using a random coding argument and bound the probability of decoding error of a particular block by a summing up the appropriate error probabilities.

Proposition 1: Partition of $\{0,1\}^{nA}\times\{0,1\}^{nB}$

Given the information sequence pair (a_1^{nA}, b_1^{nB}) we can partition the set $\{0,1\}^{nA}\times\{0,1\}^{nB}$ into $(n+1)^2$ subsets.

For $1\le j,k\le n$

$$F_n(j,k,(a_1^{nA},b_1^{nB})) =$$
$$\{(s_1^{nA},t_1^{nB})\in\{0,1\}^{nA}\times\{0,1\}^{nB}|s_1^{(j-1)A}=a_1^{(j-1)A}, s_{(j-1)A+1}^{jA}\neq a_{(j-1)A+1}^{jA}, t_1^{(k-1)B}=b_1^{(k-1)B}, t_{(k-1)B+1}^{kB}\neq b_{(k-1)B+1}^{kB}\}$$

For $1\le j\le n$

$$F_n(j,n+1,(a_1^{nA},b_1^{nB})) = \{(s_1^{nA},t_1^{nB})\in\{0,1\}^{nA}\times\{0,1\}^{nB}|s_1^{(j-1)A}=a_1^{(j-1)A}, s_{(j-1)A+1}^{jA}\neq a_{(j-1)A+1}^{jA}, t_1^{nB}=b_1^{nB}\}$$

For $1\le k\le n$

$$F_n(n+1,k,(a_1^{nA},b_1^{nB})) = \{(s_1^{nA},t_1^{nB})\in\{0,1\}^{nA}\times\{0,1\}^{nB}|s_1^{nA}=a_1^{nA} t_1^{(k-1)B}=b_1^{(k-1)B}, t_{(k-1)B+1}^{kB}\neq b_{(k-1)B+1}^{kB}\}$$

And finally $F_n(n+1,n+1,(a_1^{nA},b_1^{nB})) = \{(a_1^{nA},b_1^{nB})\}$.

We use the convention that if $i_1<i_2$, $a_{i_2}^{i_1}$ is an empty sequence. F is clearly a partition of $\{0,1\}^{nA}\times\{0,1\}^{nB}$.

Definition 2: Error Event $E_n(j,k,(a_1^{nA},b_1^{nB}))$: $E_n(j,k,(a_1^{nA},b_1^{nB})) = \{(\hat{a}_1^{nA},\hat{b}_1^{nB})\in F_n(j,k,(a_1^{nA},b_1^{nB}))\}$.

We call $E_n(j,k,(a_1^{nA},b_1^{nB}))$ the (j,k)'th error event that assumes that (a_1^{nA},b_1^{nB}) was actually the information sequence pair up to this time and captures the event that the decoded information sequence pair $(\hat{a}_1^{nA},\hat{b}_1^{nB})$ is in $F_n(j,k,(a_1^{nA},b_1^{nB}))$.

In words, $E_n(j,k,(a_1^{nA},b_1^{nB})), 1\le j,k\le n$, captures the error that the first decoding error for information source $\mathcal{A}$ is at block j while the first decoding error for information source $\mathcal{B}$ is at block k. If the decoder does not make any decoding errors for $\mathcal{A}$, but the first decoding error for $\mathcal{B}$ is at the $k'th$ block, then the error event is $E_n(n+1,k,(a_1^{nA},b_1^{nB}))$ and similarly for $E_n(j,n+1,(a_1^{nA},b_1^{nB}))$ with the roles of A and B reversed. Finally the event of making no decoding errors after nC channel uses is denoted $E_n(n+1,n+1,(a_1^{nA},b_1^{nB}))$.

Now we use the random coding bound argument to give an upper bound on the probability of $E_n(j,k,(a_1^{nA},b_1^{nB})), j\le k$.

Lemma 1: Random coding bound on the $(j,k)'th$ error event. Encoder $\mathcal{E}_A$ uses an $(A,B,C,Q_{X|U},Q_U,\mathcal{U},\mathcal{X})$ random sequential coding scheme, $\forall$ information sequence pairs (a_1^{nA},b_1^{nB}), $1\le j\le k\le n+1$ and $\forall 0\le\rho\le 1$:

$$P\left(E_n(j,k,\left(a_1^{nA},b_1^{nB}\right))\right)\le 2^{-(k-j)C(-\rho R_A+E_{0_A}(Q,\rho))-(n-k+1)C(-\rho(R_A+R_B)+E_{0_{AB}}(Q,\rho))} \quad (12)$$

Where $Q=(Q_U,Q_{X|U})$, $E_{0_A}(Q,\rho)$ and $E_{0_{AB}}(Q,\rho)$ are defined in Eqn(10) and Eqn(10) respectively.

Proof: : The proof here is similar to the derivation of the random coding bound on the block coding error probability in [14]. The probability of $E_n(j,k,(a_1^{nA},b_1^{nB}))$ is upper bounded by the probability there exists a sequence pair $(s_1^{nA},t_1^{nB})\in F_n(j,k,a_1^{nA},b_1^{nB})$, s.t.

$$P_1(y_1^{nC}|\mathcal{E}_A(s_1^{nA},t_1^{nB}))\ge P_1(y_1^{nC}|\mathcal{E}_A(a_1^{nA},b_1^{nB}))$$

Write $\mathcal{E}_A(a_1^{nA}, b_1^{nB})$ as x_1^{nC}, $\mathcal{E}_B(b_1^{nB})$ as u_1^{nC}, $\mathcal{E}_A(s_1^{nA}, t_1^{nB})$ as $\overline{x}_1^{nC}$ and $\mathcal{E}_B(t_1^{nB})$ as $\overline{u}_1^{nC}$. Since $(s_1^{nA}, t_1^{nB}) \in F_n(j, k, a_1^{nA}, b_1^{nB})$, we have $\overline{x}_1^{(j-1)C} = x_1^{(j-1)C}$, $\overline{u}_1^{(k-1)C} = u_1^{(k-1)C}$. So

$$\begin{aligned} P\left(E_n(j,k,\left(a_1^{nA}, b_1^{nB}\right))\right) &\leq P\left(\exists (s_1^{nA}, t_1^{nB}) \in F_n(j,k,a_1^{nA}, b_1^{nB}), s.t. P_1(y_1^{nC}|\overline{x}_1^{nC}) \geq P_1(y_1^{nC}|x_1^{nC})\right) \\ &= \sum_{x_1^{nC}} \sum_{u_1^{nC}} \sum_{y_1^{nC}} Q_U(u_1^{nC}) Q_{X|U}(x_1^{nC}|u_1^{nC}) P_1(y_1^{nC}|x_1^{nC}) \\ & \quad P\left(\exists (s_1^{nA}, t_1^{nB}) \in F_n(j,k,a_1^{nA}, b_1^{nB}), s.t. P_1(y_1^{nC}|\overline{x}_1^{nC}) \geq P_1(y_1^{nC}|x_1^{nC}) | x_1^{nC}, u_1^{nC}, y_1^{nC}\right) \end{aligned} \tag{13}$$

First we bound the conditional probability

$$\begin{aligned} & P(\exists (s_1^{nA}, t_1^{nB}) \in F_n(j,k,a_1^{nA}, b_1^{nB}), s.t. P_1(y_1^{nC}|\overline{x}_1^{nC}) \geq P_1(y_1^{nC}|x_1^{nC}) | x_1^{nC}, u_1^{nC}, y_1^{nC}) \\ & \leq [\sum_{(s_1^{nA}, t_1^{nB}) \in F_n(j,k,a_1^{nA}, b_1^{nB})} P(P_1(y_1^{nC}|\overline{x}_1^{nC}) \geq P_1(y_1^{nC}|x_1^{nC}) | x_1^{nC}, u_1^{nC}, y_1^{nC}))]^\rho, \forall \rho \in [0,1] \end{aligned} \tag{14}$$

$\forall (s_1^{nA}, t_1^{nB}) \in F_n(j,k,a_1^{nA}, b_1^{nB})$, by noticing that $\overline{x}_1^{(j-1)C} = x_1^{(j-1)C}$, $\overline{u}_1^{(k-1)C} = u_1^{(k-1)C}$ and the memorylessness of the channel. We have

$$P_1(y_1^{nC}|\overline{x}_1^{nC}) = P_1(y_1^{(j-1)C}|x_1^{(j-1)C}) P_1(y_{(j-1)C+1}^{(k-1)C}|\overline{x}_{(j-1)C+1}^{(k-1)C}) P_1(y_{(k-1)C+1}^{nC}|\overline{x}_{(k-1)C+1}^{nC}) \tag{15}$$

Now $\forall s > 0$:

$$\begin{aligned} & P(P_1(y_1^{nC}|\overline{x}_1^{nC}) \geq P_1(y_1^{nC}|x_1^{nC}) | x_1^{nC}, u_1^{nC}, y_1^{nC})) \\ & = \sum_{\overline{u}_1^{nC}, \overline{x}_1^{nC} : P_1(y_1^{nC}|\overline{x}_1^{nC}) \geq P_1(y_1^{nC}|x_1^{nC})} Q_{X|U}(\overline{x}_{(j-1)C+1}^{nC}|\overline{u}_{(j-1)C+1}^{nC}) Q_U(\overline{u}_{(k-1)C+1}^{nC}) \\ & \leq \sum_{\overline{x}_{(j-1)C+1}^{nC}, \overline{u}_{(j-1)C+1}^{nC}} Q_{X|U}(\overline{x}_{(j-1)C+1}^{nC}|\overline{u}_{(j-1)C+1}^{nC}) Q_U(\overline{u}_{(k-1)C+1}^{nC}) \frac{P_1(y_1^{nC}|\overline{x}_1^{nC})^s}{P_1(y_1^{nC}|x_1^{nC})^s} \\ & = \sum_{\overline{x}_{(j-1)C+1}^{nC}, \overline{u}_{(j-1)C+1}^{nC}} Q_{X|U}(\overline{x}_{(j-1)C+1}^{nC}|\overline{u}_{(j-1)C+1}^{nC}) Q_U(\overline{u}_{(k-1)C+1}^{nC}) \frac{P_1(y_{(j-1)C+1}^{nC}|\overline{x}_{(j-1)C+1}^{n)C})^s}{P_1(y_{(j-1)C+1}^{nC}|x_{(j-1)C+1}^{nC})^s} \end{aligned} \tag{16}$$

.

The size of $F_n(j,k,a_1^{nA}, b_1^{nB})$ can be bounded using

$$|F_n(j,k,a_1^{nA}, b_1^{nB})| \leq 2^{(n-k+1)(A+B)} 2^{(k-j)A} = 2^{(n-k+1)C(R_A+R_B)} 2^{(k-j)CR_A} = M \tag{17}$$

Substituting Eqn. 16 into Eqn. 14, and using the union bound argument, we get $\forall \rho \leq 1$:

$$\begin{aligned} & P(\exists (s_1^{nA}, t_1^{nB}) \in F_n(j,k,a_1^{nA}, b_1^{nB}), s.t. P_1(y_1^{nC}|\overline{x}_1^{nC}) \geq P_1(y_1^{nC}|x_1^{nC}) | x_1^{nC}, u_1^{nC}, y_1^{nC}) \\ & \leq [M \sum_{(\overline{x}_{(j-1)C+1}^{nC}, \overline{u}_{(j-1)C+1}^{nC})} Q_{X|U}(\overline{x}_{(j-1)C+1}^{nC}|\overline{u}_{(j-1)C+1}^{nC}) Q_U(\overline{u}_{(k-1)C+1}^{nC}) \frac{P_1(y_{(j-1)C+1}^{nC}|\overline{x}_{(j-1)C+1}^{n)C})^s}{P_1(y_{(j-1)C+1}^{nC}|x_{(j-1)C+1}^{nC})^s}]^\rho \end{aligned} \tag{18}$$

Substituting Eqn. 18 into Eqn. 13. We have:

$$\begin{aligned} P\left(E_n(j,k,\left(a_1^{nA}, b_1^{nB}\right))\right) \leq & \sum_{x_{(j-1)C+1}^{nC}} \sum_{u_{(j-1)C+1)}^{nC}} \sum_{y_{(j-1)C+1}^{nC}} Q_U(u_{(j-1)C+1}^{nC}) Q_{X|U}(x_{(j-1)C+1}^{nC}|u_{(j-1)C+1}^{nC}) P_1(y_{(j-1)C+1}^{nC}|x_{(j-1)C+1}^{nC}) \\ & [M \sum_{\overline{x}_{(j-1)C+1}^{nC}, \overline{u}_{(j-1)C+1}^{nC}} Q_{X|U}(\overline{x}_{(j-1)C+1}^{nC}|\overline{u}_{(j-1)C+1}^{nC}) Q_U(\overline{u}_{(k-1)C+1}^{nC}) \frac{P_1(y_{(j-1)C+1}^{nC}|\overline{x}_{(j-1)C+1}^{n)C})^s}{P_1(y_{(j-1)C+1}^{nC}|x_{(j-1)C+1}^{nC})^s}]^\rho \end{aligned}$$

By letting $s = \frac{1}{1+\rho}$, and noticing the fact that $\overline{x}$ and $\overline{u}$ are dummy variables, and $u_{(j-1)C+1}^k C = \overline{u}_{(j-1)C+1}^k C$.

$$\begin{aligned} & P\left(E_n(j,k,\left(a_1^{nA}, b_1^{nB}\right))\right) \\ & \leq M^\rho \sum_{y_{(k-1)C+1}^{nC}} \{ \sum_{x_{(k-1)C+1}^{nC}} \sum_{u_{(k-1)C+1}^{nC}} Q_{X|U}(x_{(k-1)C+1}^{nC}|u_{(k-1)C+1}^{nC}) Q_U(u_{(k-1)C+1}^{nC}) P_1(y_{(k-1)C+1}^{nC}|x_{(k-1)C+1}^{nC})^{\frac{1}{1+\rho}} \}^{1+\rho} \\ & \sum_{y_{(j-1)C+1}^{(k-1)C}} \sum_{u_{(j-1)C+1)}^{(k-1)C}} Q_U(u_{(j-1)C+1)}^{(k-1)C}) \{ \sum_{x_{(j-1)C+1}^{(k-1)C}} Q_{X|U}(x_{(j-1)C+1}^{(k-1)C}|u_{(j-1)C+1)}^{(k-1)C}) P_1(y_{(j-1)C+1}^{(k-1)C}|x_{(j-1)C+1}^{(k-1)C})^{\frac{1}{1+\rho}} \}^{1+\rho} \\ & = 2^{-(k-j)C(-\rho R_A + E_{0_A}(Q,\rho)) - (n-k+1)C(-\rho(R_A+R_B) + E_{0_{AB}}(Q,\rho))} \end{aligned}$$

The last equality is true because $Q_A(x_{i_1}^{i_2}) = \prod_{i=i_1}^{i=i_2} Q_A(x_i)$, etc. $\square$

By the definition of $E_1(Q, R_A, R_B)$ we have

$$P\left(E_n(j, k, \left(a_1^{nA}, b_1^{nB}\right))\right) \leq 2^{-(n-j+1)CE_1(Q,R_A,R_B)} \tag{19}$$

since it is the minimum over all the possible choices.

Similarly if $1 \leq k \leq j \leq n+1$, we use a similar technique to bound the error probability. The only difference is that if $b_{(k-1)B+1}^{kC} \neq t_{(k-1)B+1}^{kC}$, then $x_{(k-1)C+1}^{nC}$ and $\overline{x}_{(k-1)C+1}^{nC}$ are independently distributed $\sim Q_X$. Thus we have for $1 \leq k \leq j \leq n+1$:

$$\begin{aligned} P\left(E_n(j, k, \left(a_1^{nA}, b_1^{nB}\right))\right) &\leq 2^{-(j-k)C(-\rho R_A + E_{0_{AB}}(Q,\rho)) - (n+1-j)C(-\rho(R_A+R_B) + E_{0_{AB}}(Q,\rho))} \\ &\leq 2^{-(n-k+1)C(-\rho(R_A+R_B)+E_{0_{AB}}(Q,\rho))} \\ &\leq 2^{-(n-k+1)CE_1(Q,R_A,R_B)} \end{aligned} \tag{20}$$

The last inequality is true by the definition of $E_1(Q, R_A, R_B)$. Now combining Eqn(19) and Eqn(20), we have:

$$P\left(E_n(j, k, \left(a_1^{nA}, b_1^{nB}\right))\right) \leq 2^{-(n-\min(j,k)+1)CE_1(Q,R_A,R_B)} \tag{21}$$

Using Eqn(21), we can bound the error probabilities $P_n(\{\hat{a}_{jA+1}^{(j+1)A} \neq a_{jA+1}^{(j+1)A}\})$. The probability of making a decoding error at time nC (after nC channel uses) on the j'th information block $a_{jA+1}^{(j+1)A}$ is upper bounded by making a decoding error on *any* block with block number not larger than j.

$$\begin{aligned} P_n(\{\hat{a}_{jA+1}^{(j+1)A} \neq a_{jA+1}^{(j+1)A}\}) &\leq P(\bigcup_{i=1}^{j}\{\hat{a}_{iA+1}^{(i+1)A} \neq a_{iA+1}^{(i+1)A}) \leq \sum_{i=1}^{j}\sum_{h=1}^{n+1} P(E_n(i,h,(a_1^{nA}, b_1^{nB}))) \\ &= \sum_{i=1}^{j}\sum_{h=1}^{j} P(E_n(i,h,(a_1^{nA}, b_1^{nB}))) + \sum_{i=1}^{j}\sum_{h=j+1}^{n+1} P(E_n(i,h,(a_1^{nA}, b_1^{nB}))) \\ &\leq \sum_{i=1}^{j}\sum_{h=1}^{j} 2^{-(n-\min(i,h)+1)CE_1(Q,R_A,R_B)} + \sum_{i=1}^{j}\sum_{h=j+1}^{n+1} 2^{-(n-i+1)CE_1(Q,R_A,R_B)} \\ &\leq 2\sum_{i=1}^{j}\sum_{h=1}^{i} 2^{-(n-h+1)CE_1(Q,R_A,R_B)} + \sum_{i=1}^{j}\sum_{h=j+1}^{n+1} 2^{-(n-i+1)CE_1(Q,R_A,R_B)} \\ &\leq \sum_{i=1}^{j} \frac{2}{1-2^{-CE_1(Q,R_A,R_B)}} 2^{-(n-i+1)CE_1(Q,R_A,R_B)} + (n+1-j)\sum_{i=1}^{j} 2^{-(n-i+1)CE_1(Q,R_A,R_B)} \\ &\leq \frac{2}{(1-2^{-CE_1(Q,R_A,R_B)})^2} 2^{-(n-j+1)CE_1(Q,R_A,R_B)} + \frac{(n+1-j)}{1-2^{-CE_1(Q,R_A,R_B)}} 2^{-(n-j+1)CE_1(Q,R_A,R_B)} \\ &= \left(\frac{(n+1-j)}{1-2^{-CE_1(Q,R_A,R_B)}} + \frac{2}{(1-2^{-CE_1(Q,R_A,R_B)})^2}\right) 2^{-(n-j+1)CE_1(Q,R_A,R_B)} \\ &= (K_3 d + K_4) 2^{-dCE_1(Q,R_A,R_B)} \end{aligned} \tag{22}$$

Where $dC = (n+1-j)C$ is the decoding delay. $\square$

Using the same technique we can bound the error probability for information source $\mathcal{B}$ if decoder 1 uses a joint ML decoding rule. The error exponent is :

$$E_2(Q, R_A, R_B) = \inf_{\alpha\in[0,1]}\{\frac{1}{1-\alpha}\sup_{\rho\in[0,1]}\{\alpha(-\rho R_A + E_{0_A}(Q,\rho)) + (1-\alpha)(-\rho(R_A+R_B) + E_{0_{AB}}(Q,\rho))\}\} \tag{23}$$

Where $E_{0_A}(Q,\rho)$ and $E_{0_{AB}}(Q,\rho))$ are defined in Eqn(10). The $\frac{1}{1-\alpha}$ term is from the fact that for information source $\mathcal{B}$, the dominant error event is such that the decoding for information source $\mathcal{A}$ makes an error before $\mathcal{B}$. Decoder 1 can choose point-to-point ML decoding or joint decoding whichever[2] has the bigger error exponent to decode information source $\mathcal{B}$.

[2]In reality, we want to just get a MAP estimate for the sources. While MAP and ML are the same in the point-to-point case, we suspect that in multiterminal problems, there can be significant differences between MAP and ML decoding. We are actively exploring this now.

C. Discussion

In the previous two subsections, we proved the main part of Theorem 3. The theorem is true for all possible $Q = (Q_U, Q_{X|U})$, although for some Q the error exponents can be non-positive and hence useless. In this subsection, we sketch the proof of why the error exponents $E_{X12_B}(Q, R_B)$, $E_{X1_B}(Q, R_B)$ and $E_1(Q, R_A, R_B)$ can be positive everywhere in the interior of the achievable region of Theorem 2 with properly chosen Q.

First, we know $\exists Q = (Q_U, Q_{X|U})$, s.t. $0 \leq R_A < I(X;Y|U)$ and $0 \leq R_B < I(Z;U)$. Given this, we will only sketch the proof of why $E_1(Q, R_A, R_B) > 0$, the proof for the positiveness of $E_{X12_B}(Q, R_B)$ and $E_{X1_B}(Q, R_B)$ are similar.

$$\begin{aligned} &-\rho(R_A + R_B) + E_{0_{AB}}(Q, \rho))|_{\rho=0} = 0 \\ &\frac{\partial}{\partial \rho}(-\rho(R_A + R_B) + E_{0_{AB}}(Q, \rho)))|_{\rho=0} = -(R_A + R_B) + I(X;Y) > 0 \end{aligned} \tag{24}$$

The last inequality is true because $U \rightarrow X \rightarrow Y \rightarrow Z$, thus:

$$R_A + R_B < I(X;Y|U) + I(Z;U) \leq I(X;Y|U) + I(Y;U) = H(Y|U) - H(Y|X,U) + H(Y) - H(Y|U) = H(Y) - H(Y|X) = I(X;Y)$$

Similarly we have:

$$\begin{aligned} &-\rho R_A + E_{0_A}(Q, \rho))|_{\rho=0} = 0 \\ &\frac{\partial}{\partial \rho}(-\rho R_A + E_{0_A}(Q, \rho))|_{\rho=0} = -R_A - I(X;Y|U) > 0 \end{aligned} \tag{25}$$

Thus $\exists \rho^* \in [0,1]$,s.t. $-\rho^* R_A + E_{0_A}(Q, \rho^*))$ and $-\rho^*(R_A + R_B) + E_{0_{AB}}(Q, \rho^*))$ are both positive.

$$\begin{aligned} E_1(Q, R_A, R_B) &= \inf_{\alpha \in [0,1]} \{ \sup_{\rho \in [0,1]} \{\alpha(-\rho R_A + E_{0_A}(Q, \rho)) + (1-\alpha)(-\rho(R_A + R_B) + E_{0_{AB}}(Q, \rho))\}\} \\ &\geq \inf_{\alpha \in [0,1]} \{\alpha(-\rho^* R_A + E_{0_A}(Q, \rho^*)) + (1-\alpha)(-\rho^*(R_A + R_B) + E_{0_{AB}}(Q, \rho^*))\} > 0 \end{aligned}$$

This resolves the cases where no convex hull operation is needed. Otherwise, $\forall (R_A, R_B) \in \mathcal{R}$, where $\mathcal{R}$ is the capacity region of the degraded broadcast channel, $\exists Q = (Q_U, Q_{X|U})$, $Q' = (Q_{U'}, Q_{X|U'})$ and $\beta \in [0,1]$, s.t

$$\begin{aligned} &0 \leq R_A \leq \beta I(X;Y|U) + (1-\beta) I(X;Y|U') \\ &0 \leq R_B \leq \beta I(Z;U) + (1-\beta) I(Z;U') \end{aligned}$$

For the sake of simplicity, we assume $\beta = \frac{J}{K}, J, H \in \mathcal{N}, J < H$. Now, instead of using an $(A, B, C, Q_{X|U}, Q_U, \mathcal{U}, \mathcal{X})$ random sequential superposition scheme, we use a more complicated $(A, B, C, Q_{X|U}, Q_U, \mathcal{U}, Q_{X|U'}, Q_{U'}, \mathcal{U}', \mathcal{X}, J, H)$ random sequential superposition scheme. Basically, the $(A, B, C, Q_{X|U}, Q_U, \mathcal{U}, Q_{X|U'}, Q_{U'}, \mathcal{U}', \mathcal{X}, J, H)$ coding scheme is a "time sharing" version of $(A, B, C, Q_{X|U}, Q_U, \mathcal{U}, \mathcal{X})$ and $(A, B, C, Q_{X|U'}, Q_{U'}, \mathcal{U}', \mathcal{X})$ with the first one operating at time $kHC+1$ to time $(kH+J)C$, the second one operating at time $(kH+J)C+1$ to time $(k+1)HC$. Write $Q = (Q_U, Q_{X|U})$ and $Q' = (Q_{U'}, Q_{X|U'})$, it can be shown that the error exponent for decoding message $\mathcal{A}$ at decoder 1 can be written as:

$$\begin{aligned} E_1(Q, Q', \beta, R_A, R_B) &= \inf_{\alpha \in [0,1]} \{ \sup_{\rho \in [0,1]} \{\alpha(-\rho R_A + \beta E_{0_A}(Q, \rho) + (1-\beta) E_{0_A}(Q', \rho)) \\ &\quad +(1-\alpha)(-\rho(R_A + R_B) + \beta E_{0_{AB}}(Q, \rho) + (1-\beta) E_{0_{AB}}(Q', \rho)\}\} \end{aligned} \tag{26}$$

Following the proof in the first case, we have: $\exists \rho^* \in [0,1]$,s.t. $-\rho^* R_A + \beta E_{0_A}(Q, \rho^*) + (1-\beta) E_{0_A}(Q', \rho^*))$ and $(R_A + R_B) + \beta E_{0_{AB}}(Q, \rho^*) + (1-\beta) E_{0_{AB}}(Q', \rho^*)$ are both positive. Thus

$$\begin{aligned} E_1(Q, Q', \beta, R_A, R_B) &= \inf_{\alpha \in [0,1]} \{ \sup_{\rho \in [0,1]} \{\alpha(-\rho R_A + \beta E_{0_A}(Q, \rho) + (1-\beta) E_{0_A}(Q', \rho)) \\ &\quad +(1-\alpha)(-\rho(R_A + R_B) + \beta E_{0_{AB}}(Q, \rho) + (1-\beta) E_{0_{AB}}(Q', \rho)\}\} \\ &\geq \inf_{\alpha \in [0,1]} \{\alpha(-\rho^* R_A + \beta E_{0_A}(Q, \rho^*) + (1-\beta) E_{0_A}(Q', \rho^*)) \\ &\quad +(1-\alpha)(-\rho^*(R_A + R_B) + \beta E_{0_{AB}}(Q, \rho^*) + (1-\beta) E_{0_{AB}}(Q', \rho^*)\} > 0 \end{aligned} \tag{27}$$

Similarly, using the $(A, B, C, Q_{X|U}, Q_U, \mathcal{U}, Q_{X|U'}, Q_{U'}, \mathcal{U}', \mathcal{X}, J, H)$ sequential random coding scheme, we can achieve positive error exponents for decoding message $\mathcal{B}$ at both decoder 1 and 2.

IV. An Example

Consider a binary symmetric degraded broadcast channel consisting of two binary symmetric channels as shown in Fig. 3. We plot E^* in Fig. 3, where

$$E^* = \sup_{Q=(Q_U,Q_{X|U})} \{\min\{E_{X12_B}(Q,R_B), E_{X1_B}(Q,R_B), E_{1_A}(Q,R_A,R_B)\}\}$$

Where the error exponents are defined at Eqn(7), (8) and (9). The sequential error exponents are positive in the whole capacity region.

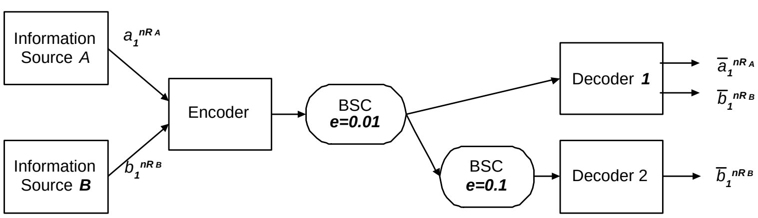

Fig. 3. A degraded broadcast channel

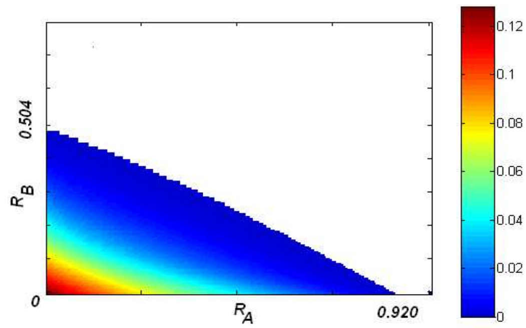

Fig. 4. Minimum sequential error exponents of the channel in Fig. 3

V. Conclusions and Future Work

We studied sequential channel coding for degraded broadcast channels and showed how superposition coding could be done in this setting. By applying a variation of Gallager's random coding analysis, we achieved a positive sequential random coding anytime exponent within the whole degraded broadcast channel rate region. This exponent measured the probability of bit-error with bit-delay rather than block-error with block-length. Furthermore, the code was "anytime" or delay universal in that the decoder could decide on a delay without telling the encoder what it is. The decisions become exponentially more reliable with increasing delay, and thus all message bits are eventually known perfectly. The focus here was on showing positive anytime exponents, and we believe that it may be possible to improve on the bounds given here.

The analyzed random coding scheme consists of a superposition of two sequential encoders designed for the degraded broadcast channel and two ML decoders. Since ML decoding is computationally prohibitive, a practical approximate decoding scheme is desirable. In addition, the complexity of the encoding grows as time goes on. To address this, we believe it should be possible to accept a certain maximum tolerable delay (or equivalently, a certain small enough error probability) beyond which we are no longer interested in correcting errors with additional waiting. This can then be realized using a (possibly time-varying) convolutional code with a long enough constraint length. [15] might also be extended to such contexts which would give implementable sequential decoders. Furthermore, similar sequential random coding bounds can be achieved for other multi-terminal problems such as the two-way and interference channels.

References

[1] C. Chang and A. Sahai, "Sequential random coding error exponents for multiple access channels," *WirelessCom 05 Symposium on Information Theory*, 2005.

[2] S. Draper, C. Chang, and A. Sahai, "Sequential random binning for streaming distributed source coding," *ISIT*, 2005.

[3] G. Forney, "Convolutional codes ii. maximum-likelihood decoding," *Inform. and Control*, vol. 25, pp. 222–266, 1974.

[4] M. Pinsker, "Bounds of the probability and of the number of correctable errors for nonblock codes," *Translation from Problemy Peredachi Informatsii*, vol. 3, pp. 44–55, 1967.

[5] A. Sahai, *Anytime Information Theory, PhD thesis*. Massachusetts Institute Technology, 2001.

[6] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link: Part i: scalar systems," *IEEE Transactions on Information Theory*, March 2005, submitted for publication.

[7] T. M. Cover and J. A. Thomas, *Elements of Inofrmation Theory*. New York: John Wiley and Sons Inc., 1991.

[8] T. M. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, pp. 2–14, 1972.

[9] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Transactions on Information Theory*, vol. 19, pp. 97 – 20, 1973.

[10] R. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, vol. 10(3), pp. 3–14, 1974.

[11] T. M. Cover, "Comments on broadcast channels," *IEEE Transactions on Information Theory*, vol. 44, pp. 2524–2530, 1998.

[12] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link part ii: vector systems," *IEEE Transactions on Information Theory*, March 2005, submitted for publication.

[13] E. C. van der Meulen, "Random coding theorems for the general discrete memoryless broadcast channel," *IEEE Transactions on Information Theory*, vol. 21, pp. 180 – 190, 1975.

[14] R. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons.

[15] F. Jelinek, "Upper bounds on sequential decoding performance parameters," *IEEE Transactions on Information Theory*, vol. 20, pp. 227 – 239, March 1974.