



*IRWIN AND JOAN JACOBS  
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES*

# **Exact Random Coding Exponents and Universal Decoders for the Asymmetric Broadcast Channel**

**Ran Averbuch and Neri Merhav**

**CCIT Report #909  
February 2017**

 Electronics  
Computers  
Communications

*THE ANDREW & ERNA VITERBI FACULTY OF ELECTRICAL ENGINEERING  
TECHNION—ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 3200003, ISRAEL*



# Exact Random Coding Exponents and Universal Decoders for the Asymmetric Broadcast Channel

Ran Averbuch and Neri Merhav

February 23, 2017

The Andrew & Erna Viterbi Faculty of Electrical Engineering  
Technion - Israel Institute of Technology  
Technion City, Haifa 3200004, ISRAEL  
{rans@campus, merhav@ee}.technion.ac.il

## Abstract

This work contains two main contributions concerning the asymmetric broadcast channel. The first is an analysis of the exact random coding error exponents for both users, and the second is the derivation of universal decoders for both users. These universal decoders are certain variants of the maximum mutual information (MMI) universal decoder, which achieve the corresponding random coding exponents of optimal decoding. In addition, we introduce some lower bounds, which involve optimization over very few parameters, unlike the original, exact exponents, which involve minimizations over auxiliary probability distributions. Numerical results for the binary symmetric broadcast channel show improvements over previously derived error exponents for the same model.

**Index Terms:** Error exponent, asymmetric broadcast channel, universal decoding, MMI.

# 1 Introduction

One of the most elementary system configuration models in multi-user information theory is the broadcast channel (BC). It has been introduced in the early seventies of the twentieth century by Cover [1], and since then, a vast amount of papers and books, studying different topics of the broadcast problem, have been published. Generally speaking, the BC is a communication model, where a single transmitter wishes to communicate different messages to two or more receivers. The various messages may be private (i.e., aimed to one receiver only) or common (i.e., aimed to two or more receivers).

Although the characterization of the capacity region of the general BC is still an open problem, some special cases have been solved, most notably, the degraded BC (DBC), first presented in [1]. The capacity region of the DBC, conjectured by Cover, was first proved to be achievable by Bergmans [2], and the converse was established by Bergmans [3] and Gallager [4]. Another special case, which is somewhat more general than the DBC and which was first introduced and solved by Körner and Marton [5], is the broadcast channel with degraded message sets, also known as the asymmetric broadcast channel (ABC). The direct part of their coding theorem relies on Bergmans' scheme, which suggested the use of an hierarchical random code: First generate "cloud centers", which designate messages intended to both the receiver with the relatively high channel quality, henceforth referred to as the *strong user*, and the receiver with the relatively low channel quality, henceforth referred to as the *weak user*. Then, in the second step, "around" each cloud center, generate a codeword for each message that is intended to the strong user only. The transmitter sends a codeword pertaining to one of the clouds. The strong decoder fully decodes both the common message and his private message, whereas the weak decoder decodes the common message only. Other channels in which one receiver is superior to another and channels with nested information were studied by Csiszár and Körner [6] and by El Gamal [7], to name a few.

Multi-user information theory is, first and foremost, driven by the quest to characterize capacity regions, i.e., the region of all sets of rates that allow reliable communication (a.k.a. achievable rates). A somewhat sharper performance metric concerns the exponential decay rate (the error exponent) of the probability of error for each user, as a function of the coding rates within the interior of the capacity region. On top of that, an interesting question concerns

the trade-off between the error exponent of the strong user and the one of the weak user, or equivalently, the achievable region in the plane of error exponents for a given set of coding rates.

While the capacity regions of the DBC and the ABC have been known for many years, only little has been known about their reliability functions. Earlier works on error exponents for the general DBC and ABC include those of Gallager [4] and Körner and Sgarro [8], respectively. In both works, the coding scheme of [2] was adopted, but the decoder was sub-optimal. More recently, Kaspi and Merhav [9] have derived some tighter lower bounds to the reliability functions of both users by analyzing random coding error exponents of their optimal decoders. While their derivation was exponentially tight at most of the steps, there were still some steps in [9] where exponential tightness might have been compromised. Moreover, Kaspi and Merhav have analyzed ensembles of i.i.d. codes, which are not as good as ensembles of fixed composition codes [12, Section 7.3]. These two points give rise to the thought that there is room for improvement upon the results of [9], and indeed, such an improvement is one of the contributions of this work. In fact, the exponential error bounds, derived in this paper, both for the strong user and the weak one, are tight in the sense that they provide the exact random coding exponents for the ensemble of fixed composition codes. Moreover, the resulting expressions are much simpler and easier to calculate than those of the best exponential bounds of Kaspi and Merhav (see, in particular, the second part of [9]).

Interestingly, one of the ingredients that contributes significantly to this simplification in the error exponent expressions, is the derivation of *universal decoders* for both users, and this simplification is achieved thanks to a simple sandwich argument, asserting that a lower bound to the error exponent of the universal decoder cannot be larger than an upper bound to the error exponent of the optimal decoder, but on the other hand, the latter turns out to be mathematically smaller than or equal to the former, and so, by contrasting the two exponential error bounds, which must therefore be equivalent, the expressions are considerably simplified. In other words, beyond this simplification of the error exponent bounds, there is an additional bonus, which is in obtaining universal decoders for both users. These decoders achieve the same random coding error exponents as the corresponding optimal decoders of the two users. Both universal decoders are certain variants of the maximum mutual information (MMI) decoder [11, Theorem 5.2], but they are different from the earlier proposed MMI-like universal decoders for the ABC, due to Körner and Sgarro [8]. For one thing, our universal decoder for the weak user

depends explicitly on the entire code, unlike the one in [8], which depends on the cloud centers only.

Since we rely heavily on the method of types, our exponential error bounds have the flavor of those of Csiszár and Körner [11]. While exponentially tight, their shortcoming is that they are not easy to calculate since they involve minimizations over auxiliary channels, and these might be computationally painful especially for large alphabets. To alleviate this difficulty, we also propose Gallager–style bounds [12], which require optimizations over very few (one or two) parameters, but the caveat is that exponential tightness might be sacrificed. Moreover, the Gallager–style bounds lend themselves to better intuitive understanding on the behavior of the error exponents for both of the users. Specifically, we derive a *phase diagram* for the weak user, which fully describes the functional behavior of the bound in different regions of the plane of rates. We also demonstrate our results numerically for an example of the binary symmetric BC, and compare our results to those in earlier works, showing explicitly the improvement.

The remaining part of the paper is organized as follows. In Section 2, we establish notation conventions, formalize the model and the problem, and finally, review some preliminaries. In Section 3, we summarize the main theoretical results of this paper, and give some numerical results for the binary symmetric BC. Section 4 provides the proofs concerning the strong user in the ABC (the exact random coding error exponent and the universal decoder), and Section 5 contains a similar treatment for the weak user. In Section 6, we derive lower bounds on the exact random coding error exponents, and in Section 7 we study them.

## 2 Notation Conventions and Problem Formulation

### 2.1 Notation Conventions

Throughout the paper, random variables will be denoted by capital letters, specific values they may take will be denoted by the corresponding lower case letters, and their alphabets will be denoted by calligraphic letters. Random vectors and their realizations will be denoted, respectively, by capital letters and the corresponding lower case letters, both in the bold face font. Their alphabets will be superscripted by their dimensions. For example, the random vector  $\mathbf{X} = (X_1, \dots, X_n)$ , ( $n$  - positive integer) may take a specific vector value  $\mathbf{x} = (x_1, \dots, x_n)$  in  $\mathcal{X}^n$ , the  $n$ -th order Cartesian power of  $\mathcal{X}$ , which is the alphabet of each component of

this vector. Sources and channels will be subscripted by the names of the relevant random variables/vectors and their conditionings, whenever applicable, following the standard notation conventions, e.g.,  $Q_X$ ,  $Q_{Y|X}$ , and so on. When there is no room for ambiguity, these subscripts will be omitted. For a generic joint distribution  $Q_{XY} = \{Q_{XY}(x, y), x \in \mathcal{X}, y \in \mathcal{Y}\}$ , which will often be abbreviated by  $Q$ , information measures will be denoted in the conventional manner, but with a subscript  $Q$ , that is,  $H_Q(X)$  is the marginal entropy of  $X$ ,  $H_Q(X|Y)$  is the conditional entropy of  $X$  given  $Y$ ,  $I_Q(X; Y) = H_Q(X) - H_Q(X|Y)$  is the mutual information between  $X$  and  $Y$ , and so on. The weighted divergence between two conditional distributions (channels), say,  $Q_{Z|X}$  and  $W = \{W(z|x), x \in \mathcal{X}, z \in \mathcal{Z}\}$ , with weighting  $Q_X$  is defined as

$$D(Q_{Z|X}||W|Q_X) = \sum_{x \in \mathcal{X}} Q_X(x) \sum_{z \in \mathcal{Z}} Q_{Z|X}(z|x) \log \frac{Q_{Z|X}(z|x)}{W(z|x)}, \quad (1)$$

where logarithms, here and throughout the sequel, are taken to the natural base. The probability of an event  $\mathcal{E}$  will be denoted by  $\Pr\{\mathcal{E}\}$ , and the expectation operator with respect to (w.r.t.) a probability distribution  $P$  will be denoted by  $\mathbb{E}\{\cdot\}$ . For two positive sequences  $a_n$  and  $b_n$ , the notation  $a_n \doteq b_n$  will stand for equality in the exponential scale, that is,  $\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$ . The indicator function of an event  $\mathcal{E}$  will be denoted by  $\mathcal{I}\{\mathcal{E}\}$ . The notation  $[x]_+$  will stand for  $\max\{0, x\}$ .

The empirical distribution of a sequence  $\mathbf{x} \in \mathcal{X}^n$ , which will be denoted by  $\hat{P}_{\mathbf{x}}$ , is the vector of relative frequencies,  $\hat{P}_{\mathbf{x}}(x)$ , of each symbol  $x \in \mathcal{X}$  in  $\mathbf{x}$ . The type class of  $\mathbf{x} \in \mathcal{X}^n$ , denoted  $\mathcal{T}(\mathbf{x})$ , is the set of all vectors  $\mathbf{x}'$  with  $\hat{P}_{\mathbf{x}'} = \hat{P}_{\mathbf{x}}$ . When we wish to emphasize the dependence of the type class on the empirical distribution  $\hat{P}$ , we will denote it by  $\mathcal{T}(\hat{P})$ . Information measures associated with empirical distributions will be denoted with 'hats' and will be subscripted by the sequences from which they are induced. For example, the entropy associated with  $\hat{P}_{\mathbf{x}}$ , which is the empirical entropy of  $\mathbf{x}$ , will be denoted by  $\hat{H}_{\mathbf{x}}(X)$ . Similar conventions will apply to the joint empirical distribution, the joint type class, the conditional empirical distributions and the conditional type classes associated with pairs (and multiples) of sequences of length  $n$ . Accordingly,  $\hat{P}_{\mathbf{x}\mathbf{y}}$  would be the joint empirical distribution of  $(\mathbf{x}, \mathbf{y}) = \{(x_i, y_i)\}_{i=1}^n$ ,  $\mathcal{T}(\mathbf{x}, \mathbf{y})$  or  $\mathcal{T}(\hat{P}_{\mathbf{x}\mathbf{y}})$  will denote the joint type class of  $(\mathbf{x}, \mathbf{y})$ ,  $\mathcal{T}(\mathbf{x}|\mathbf{y})$  will stand for the conditional type class of  $\mathbf{x}$  given  $\mathbf{y}$ ,  $\hat{H}_{\mathbf{x}\mathbf{y}}(X, Y)$  will designate the empirical joint entropy of  $\mathbf{x}$  and  $\mathbf{y}$ ,  $\hat{H}_{\mathbf{x}\mathbf{y}}(X|Y)$  will be the empirical conditional entropy,  $\hat{I}_{\mathbf{x}\mathbf{y}}(X; Y)$  will denote the empirical mutual information, and so on. When we wish to emphasize the dependence of  $\mathcal{T}(\mathbf{x}|\mathbf{y})$  upon  $\mathbf{y}$

and the relevant empirical conditional distribution,  $Q_{X|Y} = \hat{P}_{\mathbf{x}|\mathbf{y}}$ , we denote it by  $\mathcal{T}(Q_{X|Y}|\mathbf{y})$ . Similar conventions will apply to triples of sequences, say,  $\{(\mathbf{x}, \mathbf{y}, \mathbf{z})\}$ , etc. Likewise, when we wish to emphasize the dependence of empirical information measures upon a given empirical distribution given by  $Q$ , we denote them using the subscript  $Q$ , as described above.

## 2.2 Problem Formulation

We consider a memoryless ABC with a finite input alphabet  $\mathcal{X}$  and finite output alphabets  $\mathcal{Y}$  and  $\mathcal{Z}$ . Let  $W_1 = \{W_1(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$  and  $W_2 = \{W_2(z|x), x \in \mathcal{X}, z \in \mathcal{Z}\}$  denote the single-letter input-output transition probability matrices, associated with the strong user and the weak user, respectively. When these channels are fed by an input vector  $\mathbf{x} \in \mathcal{X}^n$ , they produce the corresponding output vectors  $\mathbf{y} \in \mathcal{Y}^n$  and  $\mathbf{z} \in \mathcal{Z}^n$ , according to

$$W_1(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^n W_1(y_t|x_t), \quad (2)$$

$$W_2(\mathbf{z}|\mathbf{x}) = \prod_{t=1}^n W_2(z_t|x_t). \quad (3)$$

We are interested in sending one out of  $M_y M_z$  messages to the strong user, that observes  $\mathbf{y}$ , and one out of  $M_z$  messages to the weak user, that observes  $\mathbf{z}$ . Specifically, consider the following mechanism of random selection of an hierarchical code for the ABC. Let  $\mathcal{U}$  be a finite alphabet, let  $P_U$  be a given probability distribution on  $\mathcal{U}$ , and let  $P_{X|U}$  be a given matrix of conditional probabilities of  $X$  given  $U$ . We first select, independently at random,  $M_z = e^{nR_z}$   $n$ -vectors (“cloud centers”),  $\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{M_z-1}$ , all under the uniform distribution over the type class  $\mathcal{T}(P_U)$ . Next, for each  $i = 0, 1, \dots, M_z - 1$ , we select conditionally independently (given  $\mathbf{u}_i$ ),  $M_y = e^{nR_y}$  codewords,  $\mathbf{x}_{i,0}, \mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,(M_y-1)}$ , under the uniform distribution across the conditional type class  $\mathcal{T}(P_{X|U}|\mathbf{u}_i)$ . We denote the sub-code  $\mathcal{C}_i = \{\mathbf{x}_{i,0}, \mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,(M_y-1)}\}$ . Once selected, the entire codebook  $\mathcal{C} = \cup_{i=0}^{M_z-1} \mathcal{C}_i$ , together with the collection of all cloud centers,  $\{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{M_z-1}\}$ , are revealed to the encoder and to both decoders.

The optimal decoder for the strong user is given by

$$[\hat{i}(\mathbf{y}), \hat{j}(\mathbf{y})] = \arg \max_{0 \leq i \leq M_z-1, 0 \leq j \leq M_y-1} W_1(\mathbf{y}|\mathbf{x}_{i,j}), \quad (4)$$

while the optimal decoder for the weak user (the bin index decoder) is given by

$$\tilde{i}(\mathbf{z}) = \arg \max_{0 \leq i \leq M_z-1} W_2(\mathbf{z}|\mathcal{C}_i), \quad (5)$$

where

$$W_2(\mathbf{z}|\mathcal{C}_i) \triangleq \frac{1}{M_y} \sum_{\mathbf{x} \in \mathcal{C}_i} W_2(\mathbf{z}|\mathbf{x}) = \frac{1}{M_y} \sum_{j=0}^{M_y-1} W_2(\mathbf{z}|\mathbf{x}_{i,j}). \quad (6)$$

Let  $\mathbf{Y} \in \mathcal{Y}^n$  and  $\mathbf{Z} \in \mathcal{Z}^n$  be the channel outputs resulting from the transmission of  $\mathbf{X}_{i,j}$ . Define the average error probabilities of decoders (4) and (5) as

$$\bar{P}_{e,s}(R_y, R_z) = \frac{1}{M_y M_z} \sum_{i=0}^{M_z-1} \sum_{j=0}^{M_y-1} \Pr\left\{[\hat{i}(\mathbf{Y}), \hat{j}(\mathbf{Y})] \neq (i, j) \mid \mathbf{X}_{i,j} \text{ sent}\right\}, \quad (7)$$

and

$$\bar{P}_{e,w}(R_y, R_z) = \frac{1}{M_y M_z} \sum_{i=0}^{M_z-1} \sum_{j=0}^{M_y-1} \Pr\left\{\tilde{i}(\mathbf{Z}) \neq i \mid \mathbf{X}_{i,j} \text{ sent}\right\}, \quad (8)$$

where in both definitions,  $\Pr\{\cdot\}$  designates probabilities associated with the randomness of the codebook, as well as that of the channel outputs given its input. The corresponding random coding error exponents are defined as

$$E_s(R_y, R_z) = \lim_{n \rightarrow \infty} \left[ -\frac{\ln \bar{P}_{e,s}(R_y, R_z)}{n} \right], \quad (9)$$

and

$$E_w(R_y, R_z) = \lim_{n \rightarrow \infty} \left[ -\frac{\ln \bar{P}_{e,w}(R_y, R_z)}{n} \right], \quad (10)$$

provided that the limits exist. Our main objective is to obtain single-letter expressions for  $E_s(R_y, R_z)$  and  $E_w(R_y, R_z)$ . As for the universal decoders, consider first the weak user. We wish to find a function  $F(\mathbf{z}, \mathbf{u}_i, \mathcal{C}_i)$ , that is independent of the (unknown) parameters of the channel  $W_2$ , such that the following *universal decoder* for the weak user

$$\tilde{i}_U(\mathbf{z}) = \arg \max_{0 \leq i \leq M_z-1} F(\mathbf{z}, \mathbf{u}_i, \mathcal{C}_i) \quad (11)$$

achieves an average error probability whose exponent is  $E_w(R_y, R_z)$ . By the same token, we wish to find a universal decoder for the strong user, of the form

$$[\hat{i}_U(\mathbf{y}), \hat{j}_U(\mathbf{y})] = \arg \max_{0 \leq i \leq M_z-1, 0 \leq j \leq M_y-1} G(\mathbf{y}, \mathbf{u}_i, \mathbf{x}_{i,j}), \quad (12)$$

where the function  $G$  is independent of  $W_1$ , yet the decoder  $[\hat{i}_U(\mathbf{y}), \hat{j}_U(\mathbf{y})]$  achieves  $E_s(R_y, R_z)$ .



### 3 Main Results

#### 3.1 Exact Random Coding Error Exponents

Let  $Q_{UXY}$  and  $Q_{UXZ}$  denote two generic joint probability distributions of the random vectors  $(U, X, Y)$  and  $(U, X, Z)$ , whose  $(UX)$ -marginals are both identical to  $P_{UX}$ . Define

$$E_y(Q_{UXY}, R_y, R_z) = \min \left\{ [I_Q(U; Y) + [I_Q(X; Y|U) - R_y]_+ - R_z]_+, [I_Q(X; Y|U) - R_y]_+ \right\}, \quad (13)$$

and

$$E_z(Q_{UXZ}, R_y, R_z) = [I_Q(U; Z) + [I_Q(X; Z|U) - R_y]_+ - R_z]_+. \quad (14)$$

Our first main result is the following.

**Theorem 1.** Under the assumptions of Section 2, the limits (9) and (10) exist and are given by the following single-letter expressions:

$$E_s(R_y, R_z) = \min_{Q_{Y|UX}} \left\{ D(Q_{Y|UX} \| W_{Y|X} | P_{UX}) + E_y(Q_{UXY}, R_y, R_z) \right\}, \quad (15)$$

$$E_w(R_y, R_z) = \min_{Q_{Z|UX}} \left\{ D(Q_{Z|UX} \| W_{Z|X} | P_{UX}) + E_z(Q_{UXZ}, R_y, R_z) \right\}. \quad (16)$$

We prove the result concerning the strong user in Section 4 and the result concerning the weak user in Sections 5. Notice that both error exponents depend on both coding rates, in contrast to the error exponents given in the previous works [4] and [8].

Several remarks are now in order.

- An immediate byproduct of Theorem 1 is finding the set of rate pairs  $(R_y, R_z)$  for which both  $E_s(R_y, R_z) > 0$  and  $E_w(R_y, R_z) > 0$ . It is not difficult to show that this set is given by:

$$\mathcal{R} = \left\{ (R_y, R_z) \mid R_y < I(X; Y|U), R_y + R_z < I(X; Y), R_z < I(U; Z) \right\}, \quad (17)$$

evaluated with the distribution  $P_{UX} \times W_{Y|X} \times W_{Z|X}$ . The convex hull of the closure of the union over all code distributions  $\{P_{UX}\}$  gives the capacity region. We may also consider an *individual attainable region* for each user, i.e., the set of rate pairs for which the probability of error vanishes for one of the users, but without taking into account the other user. Later on, individual attainable regions will become relevant when we consider the phase diagrams. It is not difficult to show that the attainable region for the weak user, to be denoted by  $\mathcal{R}_w$ ,

is given by  $\mathcal{R}_w = \{(R_y, R_z) \mid R_y + R_z < I(X; Z)\} \cup \{(R_y, R_z) \mid R_z < I(U; Z)\}$ , evaluated with the distribution  $P_{UX} \times W_{Z|X}$ , while the attainable region for the strong user, to be denoted by  $\mathcal{R}_s$ , is given by  $\mathcal{R}_s = \{(R_y, R_z) \mid R_y + R_z < I(X; Y)\} \cap \{(R_y, R_z) \mid R_y < I(X; Y|U)\}$ , evaluated with the distribution  $P_{UX} \times W_{Y|X}$ . Notice that the attainable region of the weak user is not bounded, i.e., reliable bin index decoding may still be guaranteed for any satellites rate  $R_y$ , as long as  $R_z < I(U; Z)$ .

- The computation of the error exponents involves minimizations over auxiliary channels  $Q_{Y|UX}$  and  $Q_{Z|UX}$ . For large input and output alphabets, we are motivated to look for alternative expressions for the error exponents, whose optimization does not depend on the alphabet sizes, even at the expense of some loss in the exponential tightness. We will discuss such an alternative form in the sequel.

- Both error exponents depend on the input distribution. While in the single-user regime, we may maximize the final expression over the input distribution in order to maximize the error exponent, this is no longer the case for the ABC. Even in the simplest case of a binary symmetric BC, we see that the best code for the strong user is the worst one for the weak user, and vice versa. To see why is that true, let  $P_U = (\frac{1}{2}, \frac{1}{2})$  and let  $P_{X|U}$  be a BSC with a crossover probability  $\frac{1}{2}$ . In this case, the hierarchy of the codebook degenerates, i.e., the codebook has a constant composition, which is best for the strong user. In the other extreme,  $P_{X|U}$  is a BSC with a crossover probability 0. The error probability of the strong user is almost one, but the error exponent of the weak user is the largest and independent of  $R_y$ . Hence, the choice of the input distribution trades off between the error exponents of the two users.

- As can be seen from the minimum in eq. (13), there are two different kinds of error events for the strong user. Let  $Q^*$  denote the minimizer in (15). Now, if for some  $(R_y, R_z)$ , the inequality  $[I_{Q^*}(U; Y) + [I_{Q^*}(X; Y|U) - R_y]_+ - R_z]_+ > [I_{Q^*}(X; Y|U) - R_y]_+$  holds, then the dominant error event for the strong user is caused by competing codewords from the true cloud, otherwise, the dominant error event is caused by competitive clouds.

- In fact, the cardinality  $|\mathcal{U}|$  is a free parameter in our problem. As such, we may let  $|\mathcal{U}| \rightarrow \infty$ , and it is definitely not obvious that a finite  $|\mathcal{U}|$  is optimal. This is because we cannot see how to apply the usual cardinality bounding techniques based on the support lemma [11, page 310]. It must be clear that even if the optimal  $|\mathcal{U}|$  is finite, it may not be the same as the bound given in the converse theorem of the capacity region of the ABC ( $|\mathcal{U}| \leq |\mathcal{X}| + 2$ ) [5].

### 3.2 Universal Decoders

As mentioned in the Introduction, universal MMI decoders for both receivers were proposed in [8], where for the weak user, this decoder was defined by:

$$\tilde{i}_{\text{KS}}(\mathbf{z}) = \arg \max_{0 \leq i \leq M_z - 1} \hat{I}_{\mathbf{u}_i, \mathbf{z}}(U; Z). \quad (18)$$

The error exponent of such a decoder is inferior to the error exponent of the optimal (ML) decoder, because for one thing, it makes no use of  $\{\mathcal{C}_i\}$ , but only of the cloud centers. The universal decoder (18) achieves the following error exponent [8]

$$E_{\text{w,KS}}(R_z) = \min_{Q_{Z|UX}} \left\{ D(Q_{Z|UX} \| W_{Z|X} | P_{UX}) + [I_Q(U; Z) - R_z]_+ \right\}, \quad (19)$$

and by comparing it numerically to (16) in the case of the binary symmetric BC (see Subsection 3.4), it is evident that  $E_w(R_y, R_z)$  can be strictly higher than  $E_{\text{w,KS}}(R_z)$ , due to the additional term in (14). Hence, one may wonder whether a different universal decoder exists, whose error exponent is as large as  $E_w(R_y, R_z)$ . It turns out that the answer to this question is affirmative, and indeed, this universal decoder relies entirely on  $\mathcal{C}$  and  $\{\mathbf{u}_i\}$ . In Section 5, we prove the following theorem.

**Theorem 2.** Define the function

$$F(\mathbf{z}, \mathbf{u}_i, \mathcal{C}_i) = \max_{0 \leq j \leq M_y - 1} \left\{ \hat{I}_{\mathbf{u}_i, \mathbf{z}}(U; Z) + [\hat{I}_{\mathbf{u}_i, \mathbf{x}_{ij}, \mathbf{z}}(X; Z|U) - R_y]_+ \right\}. \quad (20)$$

The universal decoder

$$\tilde{i}_U(\mathbf{z}) = \arg \max_{0 \leq i \leq M_z - 1} F(\mathbf{z}, \mathbf{u}_i, \mathcal{C}_i) \quad (21)$$

achieves  $E_w(R_y, R_z)$ .

It turns out that there is also another universal decoder (with the same error exponent), whose structure is much more similar to the ML decoder of (5), in the sense that its metric is based on summation over  $\mathcal{C}_i$ , except that here, the unknown likelihood function is replaced by the exponentiated empirical mutual information. In the Appendix we prove the following theorem.

**Theorem 3.** The universal decoder

$$\tilde{i}_U(\mathbf{z}) = \arg \max_{0 \leq i \leq M_z - 1} \left\{ \sum_{j=0}^{M_y - 1} e^{n \hat{I}_{\mathbf{u}_i, \mathbf{x}_{ij}, \mathbf{z}}(UX; Z)} \right\} \quad (22)$$

achieves  $E_w(R_y, R_z)$ .

We next proceed to the strong user and present a universal decoder. It turns out that the MMI-like metric of the universal bin index decoder, as given in Theorem 2 (but with  $\mathbf{z}$  replaced by  $\mathbf{y}$ ), works well also for the strong user. The main difference between them is rooted in the way they use the metric. While the weak user first maximizes it within each cloud, and only then finds the cloud with the maximal value, the strong user maximizes it over both indices simultaneously. More precisely, we claim the following, which is proved in Section 4.

**Theorem 4.** Define the function

$$G(\mathbf{y}, \mathbf{u}_i, \mathbf{x}_{ij}) = \hat{I}_{\mathbf{u}_i \mathbf{y}}(U; Y) + [\hat{I}_{\mathbf{u}_i \mathbf{x}_{ij} \mathbf{y}}(X; Y|U) - R_y]_+. \quad (23)$$

The universal decoder

$$[\tilde{i}_U(\mathbf{y}), \tilde{j}_U(\mathbf{y})] = \arg \max_{0 \leq i \leq M_z - 1, 0 \leq j \leq M_y - 1} G(\mathbf{y}, \mathbf{u}_i, \mathbf{x}_{ij}) \quad (24)$$

achieves  $E_s(R_y, R_z)$ .

At this point, it is interesting to compare  $[\tilde{i}_U(\mathbf{y}), \tilde{j}_U(\mathbf{y})]$  to the universal decoder of the strong user in [8],

$$[\hat{i}_{\text{KS}}(\mathbf{y}), \hat{j}_{\text{KS}}(\mathbf{y})] = \arg \max_{0 \leq i \leq M_z - 1, 0 \leq j \leq M_y - 1} \hat{I}_{\mathbf{u}_i \mathbf{x}_{ij} \mathbf{y}}(UX; Y), \quad (25)$$

and whose random coding error exponent is given by [8]

$$E_{s, \text{KS}}(R_y, R_z) = \min_{Q_{Y|UX}} \left\{ D(Q_{Y|UX} \| W_{Y|X} | P_{UX}) + E_{y, \text{KS}}(Q_{UXY}, R_y, R_z) \right\}, \quad (26)$$

where

$$E_{y, \text{KS}}(Q_{UXY}, R_y, R_z) = \min \left\{ [I_Q(UX; Y) - (R_y + R_z)]_+, [I_Q(X; Y|U) - R_y]_+ \right\}. \quad (27)$$

By the identity  $I(UX; Y) = I(U; Y) + I(X; Y|U)$ , it is easy to see that  $E_{s, \text{KS}}(R_y, R_z) = E_s(R_y, R_z)$ , proving that (25) has an error exponent as that of (4), a fact that was not asserted in [8].

### 3.3 Gallager-Style Lower Bounds

As mentioned before, the calculations of (15) and (16) involve minimizations over auxiliary channels, which become painful when the input and output alphabets are large. For this

reason, we look for other forms of error exponent formulas, where the number of parameters to be optimized does not grow with the alphabet sizes, but the price of this might be some loss in the tightness of the bounds, i.e., we obtain *lower bounds* on the random coding error exponents. Even in the single user case, the random coding error exponent involves a minimization over an auxiliary channel, where Csiszár and Körner [11, Exercise 10.24] show that the exact error exponent is lower bounded by the following expression

$$E_G(R) = \max_{\rho \in [0,1]} \left\{ -\log \sum_y \left[ \sum_x P(x) W^{\frac{1}{1+\rho}}(y|x) \right]^{1+\rho} - \rho R \right\}, \quad (28)$$

where the subscript 'G' stands for "Gallager", who was the first to derive and analyze the error exponent in this form [12]. It is important to note that for the optimal code distribution, (28) is not only a lower bound, but the exact random coding error exponent [11]. It turns out that the exact random coding error exponents of the two users in the ABC can be lower bounded by the same methods as in [11]. In Section 6, we prove the following theorem.

**Theorem 5.** Define the functions

$$\Phi(u, y, s) = \sum_x P(x|u) [W_1(y|x)]^{\frac{1}{1+s}}, \quad (29)$$

$$\Psi(u, z, s) = \sum_x P(x|u) [W_2(z|x)]^{\frac{1}{1+s}}. \quad (30)$$

The exact random coding error exponent of the strong user is lower bounded by

$$E_s(R_y, R_z) \geq \min \left\{ E_{y,1}(R_y), E_{y,2}(R_y, R_z) \right\}, \quad (31)$$

where

$$E_{y,1}(R_y) = \max_{\rho \in [0,1]} \left\{ -\sum_u P(u) \log \left( \sum_y \Phi^{1+\rho}(u, y, \rho) \right) - \rho R_y \right\}, \quad (32)$$

$$E_{y,2}(R_y, R_z) = \max_{\mu \in [0,1]} \max_{\lambda \in [0,\mu]} \left\{ -\log \left[ \sum_y \left( \sum_u P(u) \Phi^{\frac{1+\lambda}{1+\mu}}(u, y, \lambda) \right)^{1+\mu} \right] - \lambda R_y - \mu R_z \right\}. \quad (33)$$

In addition, the random coding error exponent of the weak user is lower bounded by

$$E_w(R_y, R_z) \geq \max_{\mu \in [0,1]} \max_{\lambda \in [0,\mu]} \left\{ -\log \left[ \sum_z \left( \sum_u P(u) \Psi^{\frac{1+\lambda}{1+\mu}}(u, z, \lambda) \right)^{1+\mu} \right] - \lambda R_y - \mu R_z \right\}. \quad (34)$$

These lower bounds involve maximizations over one or two parameters only, in contrast to the

original error exponents, and so, they are much easier to evaluate. In Section 7, we study them and show how they behave in different regions of the plane of rates. In contrast to the single user case, both lower bounds of the two users depend on the code distribution, but now we are no longer able to optimize both of them simultaneously, for the reason we mentioned above, in subsection 3.1.

### 3.4 Numerical Results and Phase Diagrams

We next provide some numerical results, comparing our exponents to those of [8] and [9]. Let  $W_1$  and  $W_2$  be two binary symmetric channels (BSCs) with crossover parameters  $p_y$  and  $p_z$ , respectively ( $p_z > p_y$ ). Let  $\mathcal{U}$  be binary as well and let  $P_U$  be uniformly distributed over  $\{0, 1\}$ . Also, let  $P_{X|U}$  be a BSC with crossover parameter  $\beta \in [0, 1]$ .

The capacity region of our model is given by:

$$\begin{aligned} R_z &\leq \ln 2 - h(\beta * p_z) \\ R_y &\leq h(\beta * p_y) - h(p_y), \end{aligned} \quad (35)$$

where  $\beta * p = \beta(1 - p) + (1 - \beta)p$  and  $h(x)$  is the binary entropy function.

#### 3.4.1 Gallager-Style Lower Bounds

Using Theorem 5, we find that for the strong user,

$$E_s(R_y, R_z) \geq \min \left\{ E_{y,1}(R_y), E_{y,2}(R_y, R_z) \right\}, \quad (36)$$

where,

$$\begin{aligned} E_{y,1}(R_y) &= \max_{\rho \in [0,1]} \left\{ -\log \left\{ \left[ (1 - \beta)(1 - p_y)^{\frac{1}{1+\rho}} + \beta \cdot p_y^{\frac{1}{1+\rho}} \right]^{1+\rho} \right. \right. \\ &\quad \left. \left. + \left[ (1 - \beta) \cdot p_y^{\frac{1}{1+\rho}} + \beta \cdot (1 - p_y)^{\frac{1}{1+\rho}} \right]^{1+\rho} \right\} - \rho R_y \right\}, \end{aligned} \quad (37)$$

$$\begin{aligned} E_{y,2}(R_y, R_z) &= \max_{\mu \in [0,1]} \max_{\lambda \in [0,\mu]} \left\{ -\ln 2 - (1 + \mu) \cdot \log \left\{ \frac{1}{2} \cdot \left[ (1 - \beta)(1 - p_y)^{\frac{1}{1+\lambda}} + \beta \cdot p_y^{\frac{1}{1+\lambda}} \right]^{\frac{1+\lambda}{1+\mu}} \right. \right. \\ &\quad \left. \left. + \frac{1}{2} \cdot \left[ (1 - \beta) \cdot p_y^{\frac{1}{1+\lambda}} + \beta \cdot (1 - p_y)^{\frac{1}{1+\lambda}} \right]^{\frac{1+\lambda}{1+\mu}} \right\} - \lambda R_y - \mu R_z \right\}. \end{aligned} \quad (38)$$

For the weak user,

$$\begin{aligned}
& E_w(R_y, R_z) \\
& \geq \max_{\mu \in [0,1]} \max_{\lambda \in [0,\mu]} \left\{ -\ln 2 - (1 + \mu) \cdot \log \left\{ \frac{1}{2} \cdot \left[ (1 - \beta)(1 - p_z)^{\frac{1}{1+\lambda}} + \beta \cdot p_z^{\frac{1}{1+\lambda}} \right]^{\frac{1+\lambda}{1+\mu}} \right. \right. \\
& \quad \left. \left. + \frac{1}{2} \cdot \left[ (1 - \beta) \cdot p_z^{\frac{1}{1+\lambda}} + \beta \cdot (1 - p_z)^{\frac{1}{1+\lambda}} \right]^{\frac{1+\lambda}{1+\mu}} \right\} - \lambda R_y - \mu R_z \right\}. \quad (39)
\end{aligned}$$

We present the lower bounds by plotting families of curves, one for each exponent, as a function of one rate, while the other rate is kept fixed. Let us choose the channel probabilities to be  $p_y = 0.05$  and  $p_z = 0.1$ , and  $\beta = 0.25$ . In Fig. 1, we plot lower bounds to  $E_s(R_y, R_z)$  as a function of  $R_y$ , as given by (36), where  $R_z$  takes five different values. As long as  $R_z < 0.09$ , the

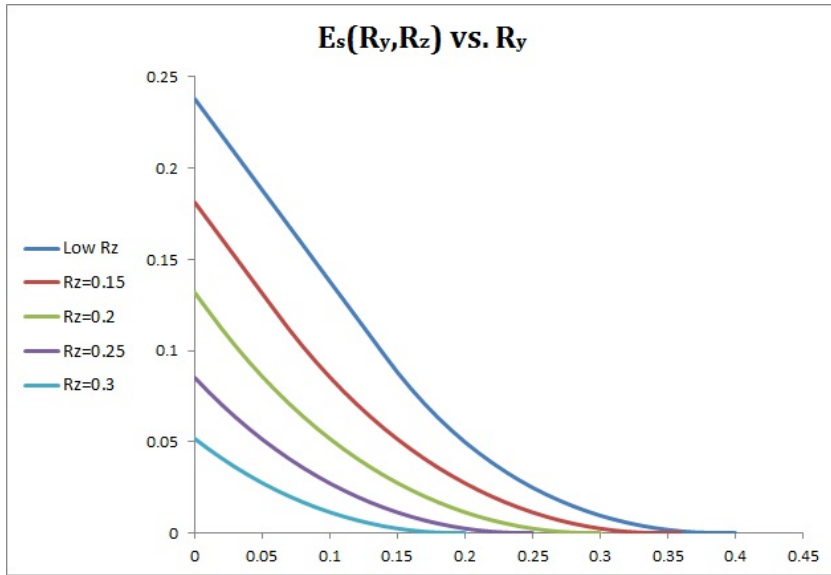


Figure 1: Error exponents for the strong user for different values of  $R_z$ .

dominant error event is caused by wrong codewords from the true cloud. In this case, the error exponent is independent of the number of clouds and is given by the dark blue curve. As  $R_z$  increases more, we find that above some critical rate, the error exponent begins to depend on the number of clouds, since the dominant error event is due to wrong codewords from competitive clouds. When the rate of the weak user is high, i.e., when the exponential number of clouds is higher than the capacity of the channel to the strong user ( $R_z > 0.49 \approx h(0.05)$ ), reliable communication is no longer possible.

In Fig. 2, we plot lower bounds to  $E_w(R_y, R_z)$  as a function of  $R_z$ , as given by (39), where

$R_y$  takes five different values. At  $R_y = 0$ , we should obtain the error exponent of a single user.

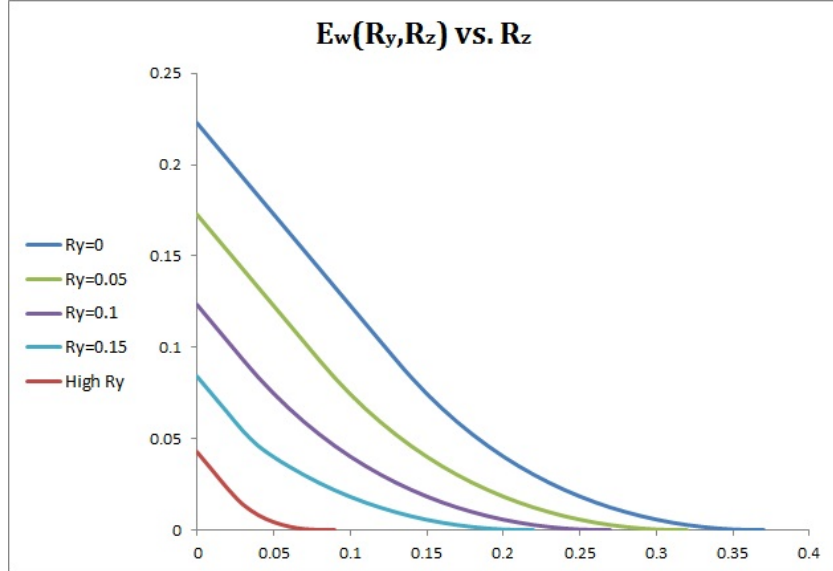


Figure 2: Error exponents for the weak user for different values of  $R_y$ .

In this case, the numerical value at zero-rate is given by  $E_w(0,0) = 0.22314$ , and  $E_w(0,R_z)$  vanishes at  $R_z \cong 0.36 \approx h(0.1)$ , which is the capacity of the channel to the weak user. For  $R_y > 0.32$ ,  $E_w(R_y, R_z)$  becomes independent of  $R_y$ , and is given by the red curve. In this case, we get a lower bound to the error exponent of the equivalent binary symmetric channel from the cloud center  $U$  to the channel output of the weak user  $Z$ .

### 3.4.2 Exact Exponents

As for the exact random coding error exponents, given by Theorem 1, the optimization problems require the minimization over the auxiliary channels  $Q_{Y|UX}$  and  $Q_{Z|UX}$ . Let us compare the Gallager-style lower bounds to the exact exponents. In Fig. 3, we see two pairs of curves of the exact exponents and their lower bounds, where  $R_z = 0.05$  and  $\beta$  takes two different values. The exact exponents are strictly better than the Gallager-style exponents. Similar results are obtained for the weak user as well (not shown here). It is important to note that in some regions in the  $R_y - R_z$  plane, the lower bounds are equal to the exact random coding error exponents.



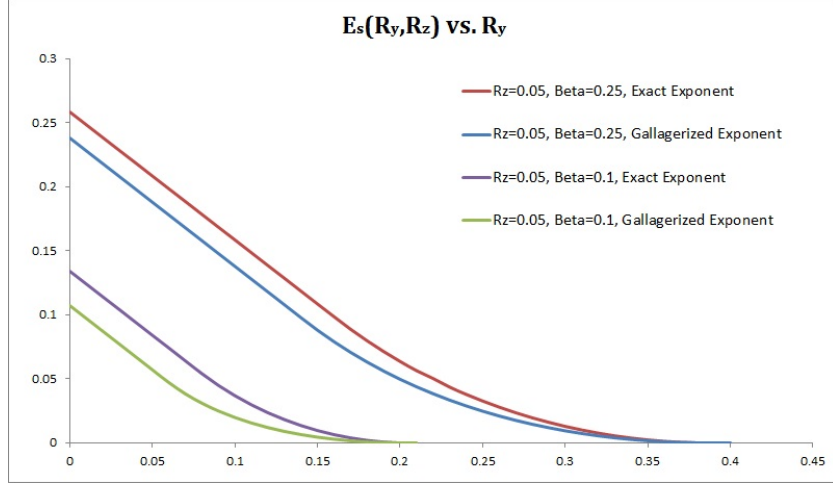


Figure 3: Exact exponents and lower bound for the strong user.

### 3.4.3 Comparison with Previous Works

As far as we know, no other works on universal decoding for the ABC exists, besides the one of [8]. Although the error exponent of the strong user given there is optimal w.r.t. the ML decoder, it is not the case for the weak user. The universal decoder of [8] for the weak user uses only the cloud centers and is independent of  $R_y$ , while the new universal decoder of Theorem 2 makes use of the entire codebook, which is the main reason for the resulted improvement. The difference between the error exponents is larger for lower values of  $R_y$ . As before, let  $p_z = 0.1$  and  $\beta = 0.25$ . Fig. 4 demonstrates the difference between the error exponents of the two universal decoders in the extreme case of  $R_y = 0$ .

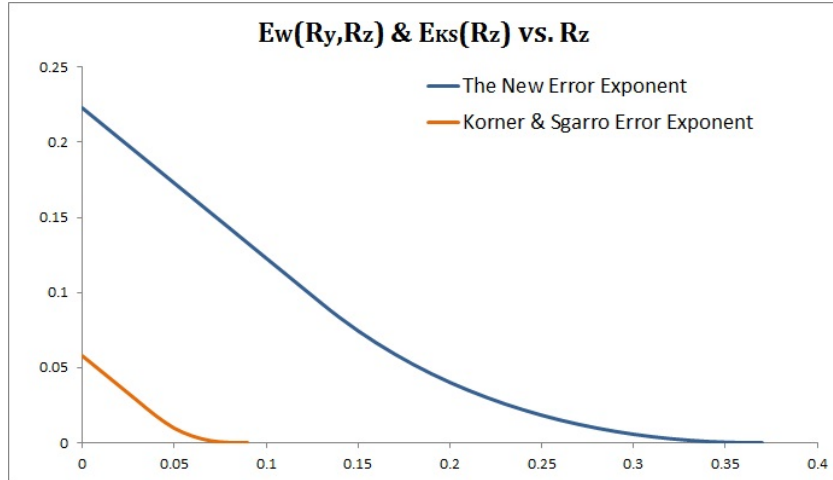


Figure 4: Comparison to [8] for the weak user.

To the best of our knowledge, the most up-to-date work on exponential lower bounds to the reliability functions of the ABC is [9], where random coding error exponents were derived using two different techniques. Each of those derivations includes at least one step that may not be exponentially tight. Also, in [9], the random codebooks are assumed to be drawn i.i.d.. We expect our proposed exact random coding error exponents to improve on [9], because of two reasons: first, our analysis is exponentially tight, and second, our ensemble is of the uniform distribution across types. This kind of random codes are known [12, Section 7.3] to be better than the i.i.d. ensembles. Our comparison here focuses on the error exponent of the weak user only. Again, let  $p_z = 0.1$ ,  $\beta = 0.25$  and  $R_y = 0.4$ . Fig. 5 compares the two error exponents, and shows that the new exponent is better.

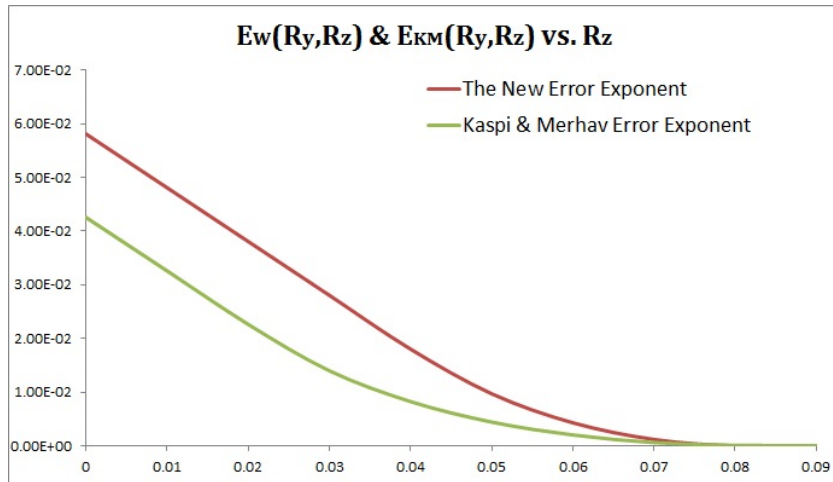


Figure 5: Comparison to [9] for the weak user.

### 3.4.4 Phase Diagrams

In the single user case, it is known that the error exponent behaves differently in different ranges of rates, i.e., it is affine at low rates and curvy at high rates. By the same token, for the ABC, the plane of rates can be divided into several different regions, where in each one of them, the error exponents behaves differently. This partition of the plane of rate pairs is of course, more involved than in the single-user case. We refer to it as a phase diagram, a term borrowed from physics. In order to study the various types of behavior of the lower bound of Theorem 5, let us invoke the following alternative and equivalent lower bound for the random coding error

Region	Dependence on $(R_y, R_z)$	$\mu^*$	$s^*$
<b>R-0</b>	affine in $R_y + R_z$	1	1
<b>R-1</b>	curvy in $R_y + R_z$	$[0,1)$	1
<b>R-2</b>	curvy in $R_y$ , affine in $R_z$	1	$[0,1)$
<b>R-3</b>	curvy in both $R_y$ and $R_z$	$[0,1)$	$[0,1)$
<b>R-4</b>	affine in $R_z$ , independent of $R_y$	1	0
<b>R-5</b>	curvy in $R_z$ , independent of $R_y$	$[0,1)$	0
<b>R-6</b>	vanishes for all $(R_z, R_y)$	0	0

Table 1: Dependence of (40) on  $R_y$  and  $R_z$  in various regions in the plane (see Fig. 6).

exponent of the weak user

$$E_w(R_y, R_z) \geq \max_{\mu \in [0,1]} \max_{s \in [0,1]} \left\{ -\log \left[ \sum_z \left( \sum_u P(u) \Psi^{\frac{1+s\mu}{1+\mu}}(u, z, s\mu) \right)^{1+\mu} \right] - s\mu R_y - \mu R_z \right\}. \quad (40)$$

Since the maximization region is now the unit square, this form is more convenient to analyze than that of (34). Fig. 6 displays a partition of the plane  $R_y - R_z$  to different regions for the Gallager-style lower bound of the weak user, where  $\beta = 0.1$ , and  $p_z = 0.1$ . Although not shown here, the phase diagrams of the exact exponents behave similarly. The study in Section 7

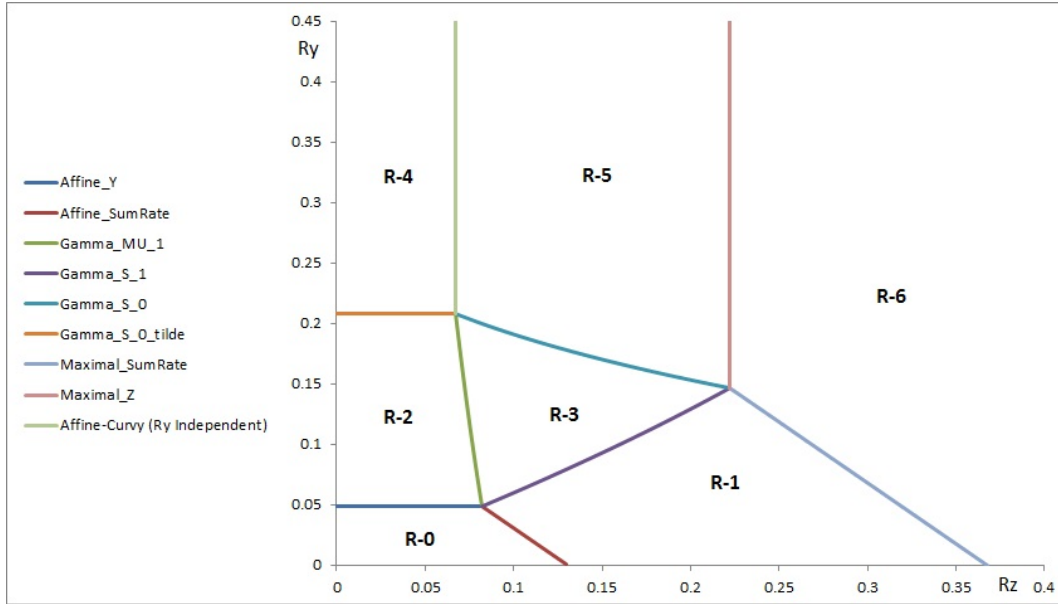


Figure 6: Phase diagram for the weak user ( $\beta = 0.1$ ).

provides a characterization of the different regions from the viewpoint of the type of dependence of the error exponent upon the rates and the maximizers  $s^*$  and  $\mu^*$  (see Table 1).

## 4 Universal decoding for the strong user

### 4.1 Analysis for a General Decoder

Let us first derive the exact random coding error exponent for a strong user that uses the following generic decoder

$$[\hat{i}(\mathbf{y}), \hat{j}(\mathbf{y})] = \arg \max_{0 \leq i \leq M_z - 1, 0 \leq j \leq M_y - 1} f(Q_{U_i X_{ij} Y}), \quad (41)$$

where from now on,  $Q_{UXY}$  will designate the joint empirical distribution induced by the three sequences  $\mathbf{u}$ ,  $\mathbf{x}$  and  $\mathbf{y}$ , i.e.,  $Q_{UXY} = \hat{P}_{\mathbf{u}\mathbf{x}\mathbf{y}}$ . The average error probability  $\bar{P}_e(R_y, R_z, n)$ , associated with (41) is

$$\bar{P}_e(R_y, R_z, n) \triangleq \frac{1}{M_y M_z} \sum_{i=0}^{M_z-1} \sum_{j=0}^{M_y-1} \Pr \left\{ \left\{ \bigcup_{k \neq j} \left\{ f(Q_{U_i X_{ik} Y}) \geq f(Q_{U_i X_{ij} Y}) \mid \mathbf{X}_{ij} \text{ sent} \right\} \right\} \right. \\ \left. \bigcup \left\{ \bigcup_{l \neq i} \bigcup_k \left\{ f(Q_{U_l X_{lk} Y}) \geq f(Q_{U_i X_{ij} Y}) \mid \mathbf{X}_{ij} \text{ sent} \right\} \right\} \right\} \quad (42)$$

where  $\Pr\{\cdot\}$  pertains to the randomness of the codebook as well as that of the channel output given its input. Without loss of generality, we assume throughout, that the transmitted codeword is  $\mathbf{X}_{00} = \mathbf{x}_{00}$ . We define

$$\mathcal{A} \triangleq \bigcup_{k=1}^{M_y-1} \mathcal{A}_k \triangleq \bigcup_{k=1}^{M_y-1} \left\{ f(Q_{U_0 X_{0k} Y}) \geq f(Q_{U_0 X_{00} Y}) \right\} \quad (43)$$

and

$$\mathcal{B} \triangleq \bigcup_{l=1}^{M_z-1} \mathcal{B}_l \triangleq \bigcup_{l=1}^{M_z-1} \bigcup_{k=0}^{M_y-1} \mathcal{B}_{lk} \triangleq \bigcup_{l=1}^{M_z-1} \bigcup_{k=0}^{M_y-1} \left\{ f(Q_{U_l X_{lk} Y}) \geq f(Q_{U_0 X_{00} Y}) \right\}. \quad (44)$$

Define the real number  $s$  as

$$s \triangleq f(Q_{U_0 X_{00} Y}). \quad (45)$$

The pairwise average error probability, conditioned on the center of the competitive cloud, is given by

$$\Pr(\mathcal{B}_{lk} | \mathbf{U}_l = \mathbf{u}') \triangleq \Pr \left\{ f(Q_{U_l X_{lk} Y}) \geq f(Q_{U_0 X_{00} Y}) \mid \mathbf{U}_l = \mathbf{u}' \right\} \quad (46)$$

$$= \sum_{\{\mathbf{x}' : f(Q_{U' X' Y}) \geq s\}} P(\mathbf{x}' | \mathbf{u}') \quad (47)$$

$$= \sum_{\{Q_{X'|U'Y} \in \mathcal{S}(Q_{U'Y}) : f(Q_{U' X' Y}) \geq s\}} \sum_{\tilde{\mathbf{x}} \in \mathcal{T}(Q_{X'|U'Y} | \mathbf{u}', \mathbf{y})} P(\tilde{\mathbf{x}} | \mathbf{u}') \quad (48)$$

$$= \sum_{\{Q_{X'|U'Y} \in \mathcal{S}(Q_{U'Y}): f(Q_{U'X'Y}) \geq s\}} P(\mathbf{x}'|\mathbf{u}') \cdot |\mathcal{T}(Q_{X'|U'Y}|\mathbf{u}', \mathbf{y})| \quad (49)$$

$$\doteq \sum_{\{Q_{X'|U'Y} \in \mathcal{S}(Q_{U'Y}): f(Q_{U'X'Y}) \geq s\}} \exp \left\{ -n \cdot I_Q(X; Y|U) \right\} \quad (50)$$

$$\doteq \max_{\{Q_{X'|U'Y} \in \mathcal{S}(Q_{U'Y}): f(Q_{U'X'Y}) \geq s\}} \exp \left\{ -n \cdot I_Q(X; Y|U) \right\} \quad (51)$$

$$= \exp \left\{ -n \cdot \min_{\{Q_{X'|U'Y} \in \mathcal{S}(Q_{U'Y}): f(Q_{U'X'Y}) \geq s\}} I_Q(X; Y|U) \right\} \quad (52)$$

$$\triangleq \exp \left\{ -n \cdot E_0(s, Q_{U'Y}) \right\}, \quad (53)$$

where  $\mathcal{S}(Q_{UY})$  denotes the set of conditional distributions  $\{Q_{X|UY}\}$  that are consistent with  $P_{UX}$ . For a given  $\mathbf{U}_l = \mathbf{u}'$ , the events  $\{\mathcal{B}_{lk}\}_k$  are all pairwise independent since we have assumed that the various codewords are pairwise conditional independent given the cloud center. Using the exponential tightness of the truncated union bound [13, Lemma A.2], we get

$$\Pr \left\{ \bigcup_{k=0}^{M_y-1} \mathcal{B}_{lk} \mid \mathbf{U}_l = \mathbf{u}' \right\} \doteq \min \left\{ 1, \sum_{k=0}^{M_y-1} \Pr(\mathcal{B}_{lk} \mid \mathbf{U}_l = \mathbf{u}') \right\} \quad (54)$$

$$= \min \left\{ 1, M_y \cdot \Pr(\mathcal{B}_{l,0} \mid \mathbf{U}_l = \mathbf{u}') \right\} \quad (55)$$

$$\doteq \min \left\{ 1, e^{nR_y} \cdot \exp \left[ -n \cdot E_0(s, Q_{U'Y}) \right] \right\} \quad (56)$$

$$\triangleq \exp \left\{ -n \cdot E_1(s, Q_{U'Y}) \right\}, \quad (57)$$

where

$$E_1(s, Q_{U'Y}) = \min_{Q_{X'|U'Y} \in \mathcal{S}(Q_{U'Y})} \left\{ [I_Q(X; Y|U) - R_y]_+ : f(Q_{U'X'Y}) \geq s \right\}. \quad (58)$$

Next, we obtain the probability of  $\mathcal{B}_l$  by calculating the expectation w.r.t. the randomness of  $\mathbf{U}_l$ :

$$\Pr \left\{ \mathcal{B}_l \right\} = \sum_{\mathbf{u}' \in \mathcal{T}(P_U)} P_U(\mathbf{u}') \cdot \Pr \left\{ \bigcup_{k=0}^{M_y-1} \mathcal{B}_{lk} \mid \mathbf{U}_l = \mathbf{u}' \right\} \quad (59)$$

$$\doteq \sum_{\mathbf{u}' \in \mathcal{T}(P_U)} P_U(\mathbf{u}') \cdot \exp \left\{ -n \cdot E_1(s, Q_{U'Y}) \right\} \quad (60)$$

$$= \sum_{\{Q_{U'Y} \in \mathcal{S}(Q_Y)\}} \sum_{\tilde{\mathbf{u}} \in \mathcal{T}(Q_{U'Y}|\mathbf{y})} P_U(\tilde{\mathbf{u}}) \cdot \exp \left\{ -n \cdot E_1(s, Q_{\tilde{U}Y}) \right\} \quad (61)$$

$$= \sum_{\{Q_{U'Y} \in \mathcal{S}(Q_Y)\}} \frac{|\mathcal{T}(Q_{U'Y}|\mathbf{y})|}{|\mathcal{T}(\mathbf{u}')|} \cdot \exp \left\{ -n \cdot E_1(s, Q_{U'Y}) \right\} \quad (62)$$

$$\doteq \sum_{\{Q_{U|Y} \in \mathcal{S}(Q_Y)\}} \exp \left\{ -n \cdot \left[ I_Q(U; Y) + E_1(s, Q_{U|Y}) \right] \right\} \quad (63)$$

$$\doteq \exp \left\{ -n \cdot \min_{\{Q_{U|Y} \in \mathcal{S}(Q_Y)\}} \left[ I_Q(U; Y) + E_1(s, Q_{U|Y}) \right] \right\} \quad (64)$$

$$\triangleq \exp \left\{ -n \cdot E_2(s, Q_Y) \right\}, \quad (65)$$

where  $\mathcal{S}(Q_Y)$  is the set of all  $\{Q_{U|Y}\}$  such that  $\sum_y Q_Y(y) Q_{U|Y}(u|y) = P_U(u)$  for every  $u \in \mathcal{U}$ . Next, we turn to calculate the probabilities of the events  $\mathcal{A}_k$ . One can easily check that the entire derivation of eqs. (46)-(53) holds in this case as well, except that now we condition on  $\mathbf{U}_0 = \mathbf{u}_0$ , such that the codewords are drawn from  $P(\cdot|\mathbf{u}_0)$ . We get

$$\Pr(\mathcal{A}_k | \mathbf{U}_0 = \mathbf{u}_0) \triangleq \Pr \left\{ f(Q_{U_0 X_{0k} Y}) \geq f(Q_{U_0 X_{00} Y}) \middle| \mathbf{U}_0 = \mathbf{u}_0 \right\} \quad (66)$$

$$\doteq \exp \left\{ -n \cdot E_0(s, Q_{U_0 Y}) \right\}. \quad (67)$$

Notice that, for a given  $\mathbf{U}_0 = \mathbf{u}_0$ ,  $\mathbf{X}_{00} = \mathbf{x}_{00}$  and  $\mathbf{Y} = \mathbf{y}$ , the events  $\{\mathcal{A}_k\}$  (errors caused by codewords from the correct cloud) and  $\{\mathcal{B}_l\}$  (errors caused by codewords from competitive clouds) are all pairwise independent. Thus, after taking the expectation w.r.t. the joint distribution of  $(\mathbf{U}_0, \mathbf{X}_{00}, \mathbf{Y})$ , we have

$$\bar{P}_e(R_y, R_z, n) = \mathbb{E} \left[ \Pr \left\{ \left\{ \bigcup_{k=1}^{M_y-1} \mathcal{A}_k \right\} \cup \left\{ \bigcup_{l=1}^{M_z-1} \mathcal{B}_l \right\} \middle| \mathbf{U}_0 = \mathbf{u}_0, \mathbf{X}_{00} = \mathbf{x}_{00}, \mathbf{Y} = \mathbf{y} \right\} \right] \quad (68)$$

$$\doteq \mathbb{E} \left[ \min \left\{ 1, \sum_{k=1}^{M_y-1} \Pr(\mathcal{A}_k | \mathbf{U}_0 = \mathbf{u}_0, \mathbf{X}_{00} = \mathbf{x}_{00}, \mathbf{Y} = \mathbf{y}) + \sum_{l=1}^{M_z-1} \Pr(\mathcal{B}_l | \mathbf{U}_0 = \mathbf{u}_0, \mathbf{X}_{00} = \mathbf{x}_{00}, \mathbf{Y} = \mathbf{y}) \right\} \right] \quad (69)$$

$$\doteq \mathbb{E} \left[ \min \left\{ 1, e^{nR_y} \exp \left\{ -n \cdot E_0(S, Q_{U_0 Y}) \right\} + e^{nR_z} \exp \left\{ -n \cdot E_2(S, Q_Y) \right\} \right\} \right] \quad (70)$$

$$\doteq \mathbb{E} \left[ \min \left\{ 1, \exp \left\{ -n \cdot \min \left\{ [E_0(S, Q_{U_0 Y}) - R_y], [E_2(S, Q_Y) - R_z] \right\} \right\} \right\} \right] \quad (71)$$

$$= \mathbb{E} \left[ \exp \left( -n \cdot \min \left\{ [E_0(S, Q_{U_0 Y}) - R_y]_+, [E_2(S, Q_Y) - R_z]_+ \right\} \right) \right] \quad (72)$$

$$\doteq \exp \left\{ -n \cdot \min_{Q_{Y|U_0 X_{00}}} \left[ D(Q_{Y|U_0 X_{00}} \| W_{Y|X_{00}} | P_{U_0 X_{00}}) + E_3(f(Q_{U_0 X_{00} Y}), Q_{U_0 Y}, R_y, R_z) \right] \right\}, \quad (73)$$

where we have defined

$$E_3(S, Q_{U_0 Y}, R_y, R_z) \triangleq \min \left\{ [E_0(S, Q_{U_0 Y}) - R_y]_+, [E_2(S, Q_Y) - R_z]_+ \right\}. \quad (74)$$

## 4.2 A Converse-Like<sup>1</sup> Result for the Strong User

We have the following:

**Lemma 1.** For every empirical distribution  $Q_{U_0 X_{00} Y}$ ,

$$E_3(f(Q_{U_0 X_{00} Y}), Q_{U_0 Y}, R_y, R_z) \leq \min \left\{ \left[ I_Q(U_0; Y) + [I_Q(X_{00}; Y|U_0) - R_y]_+ - R_z \right]_+, \right. \\ \left. [I_Q(X_{00}; Y|U_0) - R_y]_+ \right\}. \quad (75)$$

**Proof.** We start by recalling that the function  $E_3$  is defined as

$$E_3(f(Q_{U_0 X_{00} Y}), Q_{U_0 Y}, R_y, R_z) \triangleq \min \left\{ [E_0(f(Q_{U_0 X_{00} Y}), Q_{U_0 Y}) - R_y]_+, \right. \\ \left. [E_2(f(Q_{U_0 X_{00} Y}), Q_Y) - R_z]_+ \right\}, \quad (76)$$

and we separately upper bound each one of the terms. We can upper bound them by choosing any specific distribution, instead of minimizing over them. Let us start with the left term:

$$[E_0(f(Q_{U_0 X_{00} Y}), Q_{U_0 Y}) - R_y]_+ \quad (77)$$

$$= \min_{\{Q_{X|U_0 Y} \in \mathcal{S}(Q_{U_0 Y}) : f(Q_{U_0 X Y}) \geq f(Q_{U_0 X_{00} Y})\}} [I_Q(X; Y|U_0) - R_y]_+ \quad (78)$$

$$\leq [I_Q(X_{00}; Y|U_0) - R_y]_+. \quad (79)$$

---

<sup>1</sup>A converse result is usually w.r.t. both encoding and decoding. In our case, here and in Subsection 5.1, the converse results are w.r.t. the decoding only.

For the right term inside the minimum of (76), we have the following

$$[E_2(f(Q_{U_0X_0Y}), Q_Y) - R_z]_+ \quad (80)$$

$$= \min_{\{Q_{U'|Y} \in \mathcal{S}(Q_Y)\}} [I_Q(U'; Y) + E_1(f(Q_{U_0X_0Y}), Q_{U'Y}) - R_z]_+ \quad (81)$$

$$\leq [I_Q(U_0; Y) + E_1(f(Q_{U_0X_0Y}), Q_{U_0Y}) - R_z]_+ \quad (82)$$

$$= \left[ I_Q(U_0; Y) + \min_{\{Q_{X|U_0Y} \in \mathcal{S}(Q_{U_0Y}) : f(Q_{U_0XY}) \geq f(Q_{U_0X_0Y})\}} [I_Q(X; Y|U_0) - R_y]_+ - R_z \right]_+ \quad (83)$$

$$\leq [I_Q(U_0; Y) + [I_Q(X_0; Y|U_0) - R_y]_+ - R_z]_+ \quad (84)$$

Combining both upper bounds, we see that (75) holds, thus completing the proof.  $\square$

### 4.3 An Optimal Universal Decoder

Let us now select

$$f(Q_{UXY}) = I_Q(U; Y) + [I_Q(X; Y|U) - R_y]_+. \quad (85)$$

We show that (85) achieves the maximum value of  $E_3(f(Q_{U_0X_0Y}), Q_{U_0Y}, R_y, R_z)$ , as given by Lemma 1, and therefore, this decoder has the same error exponent as the one of the optimal (ML) decoder. As before, we start with the left term inside the minimum of (76), and get

$$[E_0(f(Q_{U_0X_0Y}), Q_{U_0Y}) - R_y]_+ \quad (86)$$

$$= \min_{\{Q_{X|U_0Y} \in \mathcal{S}(Q_{U_0Y})\}} \left\{ [I_Q(X; Y|U_0) - R_y]_+ : f(Q_{U_0XY}) \geq f(Q_{U_0X_0Y}) \right\} \quad (87)$$

$$= \min_{\{Q_{X|U_0Y} \in \mathcal{S}(Q_{U_0Y})\}} \left\{ [I_Q(X; Y|U_0) - R_y]_+ : \right. \\ \left. I_Q(U_0; Y) + [I_Q(X; Y|U_0) - R_y]_+ \geq I_Q(U_0; Y) + [I_Q(X_0; Y|U_0) - R_y]_+ \right\} \quad (88)$$

$$= \min_{\{Q_{X|U_0Y} \in \mathcal{S}(Q_{U_0Y})\}} \left\{ [I_Q(X; Y|U_0) - R_y]_+ : \right. \\ \left. [I_Q(X; Y|U_0) - R_y]_+ \geq [I_Q(X_0; Y|U_0) - R_y]_+ \right\} \quad (89)$$

$$= [I_Q(X_0; Y|U_0) - R_y]_+. \quad (90)$$

For the right term inside the minimum of (76),

$$[E_2(f(Q_{U_0X_0Y}), Q_Y) - R_z]_+ \quad (91)$$

$$= \min_{\{Q_{UX|Y} \in \mathcal{S}(Q_Y) : f(Q_{UXY}) \geq f(Q_{U_0X_0Y})\}} [I_Q(U; Y) + [I_Q(X; Y|U) - R_y]_+ - R_z]_+ \quad (92)$$



$$\begin{aligned}
&= \min_{\{Q_{U_X|Y} \in \mathcal{S}(Q_Y)\}} \left\{ \left[ I_Q(U; Y) + [I_Q(X; Y|U) - R_y]_+ - R_z \right]_+ : \right. \\
&I_Q(U; Y) + [I_Q(X; Y|U) - R_y]_+ \geq I_Q(U_0; Y) + [I_Q(X_{00}; Y|U_0) - R_y]_+ \left. \right\} \quad (93)
\end{aligned}$$

$$= \left[ I_Q(U_0; Y) + [I_Q(X_{00}; Y|U_0) - R_y]_+ - R_z \right]_+. \quad (94)$$

Finally, compare the minimum between (90) and (94) to the right hand side of (75).

## 5 Universal Bin Index Decoding for the Weak User

### 5.1 Analysis for a General Decoding Metric and a Converse-Like Result

Let us first derive the exact random coding error exponent of the following bin index decoder,

$$\hat{i}(\mathbf{z}) = \arg \max_{0 \leq i \leq M_z - 1} F(\mathbf{z}, \mathbf{u}_i, \mathcal{C}_i), \quad (95)$$

where

$$F(\mathbf{z}, \mathbf{u}_i, \mathcal{C}_i) \triangleq \frac{1}{M_y} \sum_{j=0}^{M_y-1} e^{nf(Q_{U_i X_{ij} Z})}, \quad (96)$$

and assume that  $f$  is upper bounded by a real number  $\Delta$ . Note that (95) includes the optimal ML decoder (5) as a special case.

To present the formula of  $E_z^*(R_y, R_z)$ , the error exponent of (95), we first need a few definitions. For a given generic joint distribution  $Q_{UXZ}$ , let  $I_Q(X; Z|U)$  denote the conditional mutual information between  $X$  and  $Z$  given  $U$ . For a given marginal  $Q_{UZ}$ , let  $\mathcal{S}(Q_{UZ})$  denote the set of conditional distributions  $\{Q_{X|UZ}\}$  such that  $\sum_z Q_{UZ}(u, z) Q_{X|UZ}(x|u, z) = P_{UX}(u, x)$  for every  $(u, x) \in \mathcal{U} \times \mathcal{X}$ , where  $P_{UX} = P_U \times P_{X|U}$ . We first define

$$E_1(s, Q_{UZ}) = \min_{Q_{X|UZ} \in \mathcal{S}(Q_{UZ})} \left\{ [I_Q(X; Z|U) - R_y]_+ : f(Q_{UXZ}) + [R_y - I_Q(X; Z|U)]_+ \geq s \right\}, \quad (97)$$

where  $s$  is an arbitrary real. Next, for a given marginal  $Q_Z$ , define

$$E_2(s, Q_Z) = \min_{Q_{U|Z} \in \mathcal{S}(Q_Z)} [I_Q(U; Z) + E_1(s, Q_{UZ})], \quad (98)$$

where the minimization is across all  $\{Q_{U|Z}\}$  such that  $\sum_z Q_Z(z) Q_{U|Z}(u|z) = P_U(u)$  for every  $u \in \mathcal{U}$ . Finally, for a given  $Q_{U_0 Z}$ , let

$$s_0(Q_{U_0 Z}) = R_y + \max_{\{Q_{X|U_0 Z} \in \mathcal{S}(Q_{U_0 Z}) : I_Q(X; Z|U_0) \leq R_y\}} [f(Q_{U_0 X Z}) - I_Q(X; Z|U_0)], \quad (99)$$

and

$$s_1(Q_{U_0 X_0 Z}) = \max \{s_0(Q_{U_0 Z}), f(Q_{U_0 X_0 Z})\}. \quad (100)$$

Now, the error exponent of the decoder (95) is given in the following lemma.

**Lemma 2.** Under the assumptions of Section 2,

$$E_z^*(R_y, R_z) = \min_{Q_{Z|U_0 X_0}} \left\{ D(Q_{Z|U_0 X_0} \| W_{Z|X_0} | P_{U_0 X_0}) + \left[ E_2(s_1(Q_{U_0 X_0 Z}), Q_Z) - R_z \right]_+ \right\}, \quad (101)$$

where  $(U_0, X_{00})$  is a replica of  $(U, X)$ , i.e.,  $P_{U_0 X_{00}} = P_{UX}$ .

**Proof.** The average probability of error, associated with (95), is given by

$$P_e^* = \mathbb{E} \left[ \Pr \left\{ \bigcup_{i=1}^{M_z-1} \{F(\mathbf{Z}, \mathbf{U}_i, \mathcal{C}_i) \geq F(\mathbf{Z}, \mathbf{U}_0, \mathcal{C}_0)\} \right\} \right] \quad (102)$$

$$\doteq \mathbb{E} \left[ \min \left\{ 1, M_z \cdot \Pr \{F(\mathbf{Z}, \mathbf{U}_1, \mathcal{C}_1) \geq F(\mathbf{Z}, \mathbf{U}_0, \mathcal{C}_0)\} \right\} \right], \quad (103)$$

where the expectation is w.r.t. the randomness of  $\mathbf{U}_0, \mathcal{C}_0$  and  $\mathbf{Z}$ , where  $\mathbf{Z}$  is the channel output in response to the input  $\mathbf{X}_{00}$  (the transmitted codeword without loss of generality). The passage from (102) to (103) is due to the exponential tightness of the truncated union bound. Here, for a given  $\mathbf{z}$ ,  $\Pr \{F(\mathbf{z}, \mathbf{U}_1, \mathcal{C}_1) \geq F(\mathbf{z}, \mathbf{u}_0, \mathcal{C}_0)\}$  is calculated w.r.t. the randomness of  $\mathbf{U}_1$  and  $\mathcal{C}_1 = \{\mathbf{X}_{1,0}, \dots, \mathbf{X}_{1,(M_y-1)}\}$ , but for a given  $\mathbf{u}_0$  and  $\mathcal{C}_0$ .

Let  $N_1(Q_{U_1 X' Z})$  denote the number of codewords  $\mathbf{x}_{1,j} \in \mathcal{C}_1$ , such that the joint empirical distribution of  $\mathbf{x}_{1,j}$  with  $(\mathbf{u}_1, \mathbf{z})$  is  $Q_{U_1 X' Z}$ , that is

$$N_1(Q_{U_1 X' Z}) = \sum_{j=0}^{M_y-1} \mathcal{I} \{(\mathbf{u}_1, \mathbf{x}_{1,j}, \mathbf{z}) \in \mathcal{T}(Q_{U_1 X' Z})\}. \quad (104)$$

Defining

$$s \triangleq \frac{1}{n} \ln \left[ \sum_{j=0}^{M_y-1} e^{nf(Q_{U_0 X_0 j Z})} \right], \quad (105)$$

we have,

$$\Pr \{F(\mathbf{z}, \mathbf{u}_1, \mathcal{C}_1) \geq F(\mathbf{z}, \mathbf{u}_0, \mathcal{C}_0)\} = \Pr \{M_y \cdot F(\mathbf{z}, \mathbf{u}_1, \mathcal{C}_1) \geq e^{ns}\} \quad (106)$$

$$= \Pr \left\{ \sum_{j=0}^{M_y-1} e^{nf(Q_{U_1 X_{1j} Z})} \geq e^{ns} \right\} \quad (107)$$

$$= \Pr \left\{ \sum_{Q_{X'|U_1 Z} \in \mathcal{S}(Q_{U_1 Z})} N_1(Q_{U_1 X' Z}) e^{nf(Q_{U_1 X' Z})} \geq e^{ns} \right\} \quad (108)$$

$$\doteq \Pr \left\{ \max_{Q_{X'|U_1Z} \in \mathcal{S}(Q_{U_1Z})} N_1(Q_{U_1X'Z}) e^{nf(Q_{U_1X'Z})} \geq e^{ns} \right\} \quad (109)$$

$$= \Pr \left\{ \bigcup_{Q_{X'|U_1Z} \in \mathcal{S}(Q_{U_1Z})} \left\{ N_1(Q_{U_1X'Z}) e^{nf(Q_{U_1X'Z})} \geq e^{ns} \right\} \right\} \quad (110)$$

$$\doteq \sum_{Q_{X'|U_1Z} \in \mathcal{S}(Q_{U_1Z})} \Pr \left\{ N_1(Q_{U_1X'Z}) e^{nf(Q_{U_1X'Z})} \geq e^{ns} \right\} \quad (111)$$

$$\doteq \max_{Q_{X'|U_1Z} \in \mathcal{S}(Q_{U_1Z})} \Pr \left\{ N_1(Q_{U_1X'Z}) e^{nf(Q_{U_1X'Z})} \geq e^{ns} \right\}. \quad (112)$$

Now, for a given  $Q_{U_1X'Z}$ , designating the joint empirical distribution of a randomly chosen  $\mathbf{x}'$  (given  $\mathbf{u}_1$ ) together with  $(\mathbf{u}_1, \mathbf{z})$ , the binomial random variable  $N_1(Q_{U_1X'Z})$  has  $e^{nR_y}$  trials and probability of success which is of the exponential order of  $e^{-nI_Q(X;Z|U)}$ . Thus, a standard large deviations analysis (see, e.g., [14, pp. 167–169]) yields

$$\Pr \left\{ N_1(Q_{U_1X'Z}) \geq e^{n[s-f(Q_{U_1X'Z})]} \right\} \doteq e^{-nE_0(Q_{U_1X'Z})}, \quad (113)$$

where

$$E_0(Q_{U_1X'Z}) = \begin{cases} [I_Q(X;Z|U) - R_y]_+ & f(Q_{U_1X'Z}) \geq s - [R_y - I_Q(X;Z|U)]_+ \\ \infty & f(Q_{U_1X'Z}) < s - [R_y - I_Q(X;Z|U)]_+ \end{cases}. \quad (114)$$

Therefore,  $\max_{Q_{X'|U_1Z} \in \mathcal{S}(Q_{U_1Z})} \Pr \left\{ N_1(Q_{U_1X'Z}) \geq e^{n[s-f(Q_{U_1X'Z})]} \right\}$  decays according to

$$E_1(s, Q_{U_1Z}) = \min_{Q_{X'|U_1Z} \in \mathcal{S}(Q_{U_1Z})} E_0(Q_{U_1X'Z}), \quad (115)$$

which is given by (97).

The conditional pairwise error probability, given  $\mathbf{U}_1 = \mathbf{u}_1$ , is of the exponential order of  $e^{-nE_1(s, Q_{U_1Z})}$ . Averaging w.r.t. the randomness of  $\mathbf{U}_1$ , we get the exponential order of  $e^{-nE_2(s, Q_Z)}$ , where  $E_2(s, Q_Z)$  is defined as in (98). To see why this is true, consider the following:

$$\sum_{\mathbf{u}_1 \in \mathcal{T}(P_U)} P_U(\mathbf{u}_1) \cdot \Pr \left\{ F(\mathbf{z}, \mathbf{u}_1, \mathcal{C}_1) \geq F(\mathbf{z}, \mathbf{u}_0, \mathcal{C}_0) \right\} \quad (116)$$

$$\doteq \sum_{Q_{U_1|Z} \in \mathcal{S}(Q_Z)} \sum_{\mathbf{u}_1 \in \mathcal{T}(Q_{U_1|Z}|\mathbf{z})} P_U(\mathbf{u}_1) \cdot e^{-n \cdot E_1(s, Q_{U_1Z})} \quad (117)$$

$$\doteq \sum_{Q_{U_1|Z} \in \mathcal{S}(Q_Z)} e^{-n \cdot E_1(s, Q_{U_1Z})} \cdot e^{-n \cdot I_Q(U;Z)} \quad (118)$$

$$\doteq \max_{Q_{U_1|Z} \in \mathcal{S}(Q_Z)} e^{-n \cdot [I_Q(U;Z) + E_1(s, Q_{U_1Z})]} \quad (119)$$

$$= e^{-n \cdot E_2(s, Q_Z)}. \quad (120)$$

Finally, we have that

$$P_e^* \doteq \mathbb{E} \left[ \min \left\{ 1, M_z \cdot e^{-n \cdot E_2(S, Q_Z)} \right\} \right] \quad (121)$$

$$= \mathbb{E} \left\{ e^{-n \left[ E_2(S, Q_Z) - R_z \right]_+} \right\}, \quad (122)$$

where the expectation is w.r.t. the randomness of

$$S = \frac{1}{n} \ln \left[ \sum_{j=0}^{M_y-1} e^{nf(Q_{U_0 X_{0j} Z})} \right], \quad (123)$$

the randomness of  $Q_Z$ , the empirical distribution of  $\mathbf{Z}$ , and  $\mathbf{U}_0$ , the real cloud center.

This expectation will be taken in two steps, the first is over the randomness of  $\{\mathbf{X}_{0,1}, \dots, \mathbf{X}_{0,(M_y-1)}\}$ , while  $\mathbf{X}_{00} = \mathbf{x}_{00}$ ,  $\mathbf{U}_0 = \mathbf{u}_0$  and  $\mathbf{Z} = \mathbf{z}$  are held fixed, whereas in the second step, the expectation is over the randomness of  $\mathbf{X}_{00}$ ,  $\mathbf{U}_0$  and  $\mathbf{Z}$ . Let  $\mathbf{x}_{00}$ ,  $\mathbf{u}_0$  and  $\mathbf{z}$  be given and let  $\epsilon > 0$  be arbitrarily small. Then,

$$P_e^*(\mathbf{x}_{00}, \mathbf{u}_0, \mathbf{z}) \triangleq \mathbb{E} \left\{ e^{-n \left[ E_2(S, Q_Z) - R_z \right]_+} \middle| \mathbf{X}_{00} = \mathbf{x}_{00}, \mathbf{U}_0 = \mathbf{u}_0, \mathbf{Z} = \mathbf{z} \right\} \quad (124)$$

$$\leq \sum_i \Pr \left\{ i\epsilon \leq S < (i+1)\epsilon \middle| \mathbf{X}_{00} = \mathbf{x}_{00}, \mathbf{U}_0 = \mathbf{u}_0, \mathbf{Z} = \mathbf{z} \right\} \\ \times \exp \left\{ -n \left[ E_2(i\epsilon, Q_Z) - R_z \right]_+ \right\}, \quad (125)$$

where  $i$  ranges from  $\frac{1}{\epsilon} f(Q_{U_0 X_{00} Z})$  to  $(R_y + \Delta)/\epsilon$ . Now,

$$e^{nS} = e^{nf(Q_{U_0 X_{00} Z})} + \sum_{j=1}^{M_y-1} e^{nf(Q_{U_0 X_{0j} Z})} \quad (126)$$

$$= e^{nf(Q_{U_0 X_{00} Z})} + \sum_{Q_{X'|U_0 Z} \in \mathcal{S}(Q_{U_0 Z})} N_0(Q_{U_0 X' Z}) e^{nf(Q_{U_0 X' Z})}, \quad (127)$$

where  $N_0(Q_{U_0 X' Z})$  is the number of codewords in  $\mathcal{C}_0 \setminus \{\mathbf{x}_{00}\}$ , whose joint empirical distribution with  $(\mathbf{u}_0, \mathbf{z})$  is  $Q_{U_0 X' Z}$ . On the one hand, we have

$$\Pr \left\{ \sum_{Q_{X'|U_0 Z} \in \mathcal{S}(Q_{U_0 Z})} N_0(Q_{U_0 X' Z}) e^{nf(Q_{U_0 X' Z})} \geq e^{nt} \right\} \doteq e^{-n \cdot E_1(t, Q_{U_0 Z})}, \quad (128)$$

and on the other hand,

$$\Pr \left\{ \sum_{Q_{X'|U_0 Z} \in \mathcal{S}(Q_{U_0 Z})} N_0(Q_{U_0 X' Z}) e^{nf(Q_{U_0 X' Z})} \leq e^{nt} \right\} \\ \doteq \Pr \left\{ \bigcap_{Q_{X'|U_0 Z} \in \mathcal{S}(Q_{U_0 Z})} \left\{ N_0(Q_{U_0 X' Z}) \leq e^{n[t - f(Q_{U_0 X' Z})]} \right\} \right\}. \quad (129)$$

This probability behaves exponentially like an indicator function of the condition that for every  $Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z})$ , either  $I_Q(X; Z|U_0) \geq R_y$  or  $R_y - I_Q(X; Z|U_0) \leq t - f(Q_{U_0X'Z})$  [10]. I.e.,

$$\Pr \left\{ \sum_{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z})} N_0(Q_{U_0X'Z}) e^{nf(Q_{U_0X'Z})} \leq e^{nt} \right\} \\ \doteq \mathcal{I} \left\{ R_y \leq \min_{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z})} \left\{ I_Q(X; Z|U_0) + [t - f(Q_{U_0X'Z})]_+ \right\} \right\}. \quad (130)$$

Let us now find what is the minimum value of  $t$  for which the value of this indicator function is unity. The condition is equivalent to

$$\min_{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z})} \max_{0 \leq a \leq 1} \left\{ I_Q(X; Z|U_0) + a [t - f(Q_{U_0X'Z})] \right\} \geq R_y, \quad (131)$$

or

$$\forall Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z}) \exists a \in [0, 1] : I_Q(X; Z|U_0) + a [t - f(Q_{U_0X'Z})] \geq R_y, \quad (132)$$

or

$$\forall Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z}) \exists a \in [0, 1] : t \geq f(Q_{U_0X'Z}) + \frac{1}{a} (R_y - I_Q(X; Z|U_0)), \quad (133)$$

or, equivalently,

$$t \geq \max_{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z})} \min_{0 \leq a \leq 1} \left[ f(Q_{U_0X'Z}) + \frac{1}{a} (R_y - I_Q(X; Z|U_0)) \right] \quad (134)$$

$$= \max_{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z})} \left[ f(Q_{U_0X'Z}) + \begin{cases} R_y - I_Q(X; Z|U_0) & , R_y \geq I_Q(X; Z|U_0) \\ -\infty & , R_y < I_Q(X; Z|U_0) \end{cases} \right] \quad (135)$$

$$= R_y + \max_{\{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z}) : I_Q(X; Z|U_0) \leq R_y\}} [f(Q_{U_0X'Z}) - I_Q(X; Z|U_0)] \quad (136)$$

$$\triangleq s_0(Q_{U_0Z}). \quad (137)$$

Thus, in summary, we have

$$\Pr \left\{ e^{nt} \leq \sum_{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z})} N_0(Q_{U_0X'Z}) e^{nf(Q_{U_0X'Z})} \leq e^{n(t+\epsilon)} \right\} \\ \doteq \begin{cases} 0 & t < s_0(Q_{U_0Z}) - \epsilon \\ e^{-n \cdot E_1(t, Q_{U_0Z})} & t \geq s_0(Q_{U_0Z}) \end{cases} \quad (138)$$

Therefore, we get the expected error probability

$$P_e^*(\mathbf{x}_{00}, \mathbf{u}_0, \mathbf{z}) \leq \sum_i \Pr \left\{ e^{ni\epsilon} \leq \sum_{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z})} N_0(Q_{U_0X'Z}) e^{nf(Q_{U_0X'Z})} \leq e^{n(i+1)\epsilon} \right\} \\ \times \exp \left\{ -n \left[ E_2 \left( \max \{i\epsilon, f(Q_{U_0X_{00}Z})\}, Q_Z \right) - R_z \right]_+ \right\} \quad (139)$$

$$\doteq \sum_{i \geq s_0(Q_{U_0Z})/\epsilon} \exp \left\{ -n E_1(i\epsilon, Q_{U_0Z}) \right\} \\ \times \exp \left\{ -n \left[ E_2 \left( \max \{i\epsilon, f(Q_{U_0X_{00}Z})\}, Q_Z \right) - R_z \right]_+ \right\}. \quad (140)$$

Since the dominant contribution to the sum over  $i$  is due to the term  $i = s_0(Q_{U_0Z})/\epsilon$  (by the non-decreasing monotonicity of the functions  $E_1(\cdot, Q_{U_0Z})$  and  $E_2(\cdot, Q_Z)$ ), we obtain

$$P_e^*(\mathbf{x}_{00}, \mathbf{u}_0, \mathbf{z}) \doteq \exp \left\{ -n \left[ E_2 \left( \max \{s_0(Q_{U_0Z}), f(Q_{U_0X_{00}Z})\}, Q_Z \right) - R_z \right]_+ \right\} \quad (141)$$

$$\triangleq \exp \left\{ -n \left[ E_2 \left( s_1(Q_{U_0X_{00}Z}), Q_Z \right) - R_z \right]_+ \right\}. \quad (142)$$

Now, after taking the expectation w.r.t. the joint distribution of  $(\mathbf{U}_0, \mathbf{X}_{00}, \mathbf{Z})$ , we get the exact random coding error exponent (101), and the proof of Lemma 2 is complete.  $\square$

Next, we introduce the following converse-like result for the weak user.

**Lemma 3.** For every empirical distribution  $Q_{U_0X_{00}Z}$ ,

$$E_2(s_1(Q_{U_0X_{00}Z}), Q_Z) \leq I_Q(U_0; Z) + [I_Q(X_{00}; Z|U_0) - R_y]_+. \quad (143)$$

**Proof.** By (98),

$$E_2(s_1(Q_{U_0X_{00}Z}), Q_Z) = \min_{Q_{U|Z} \in \mathcal{S}(Q_Z)} \left\{ I_Q(U; Z) + E_1(s_1(Q_{U_0X_{00}Z}), Q_{UZ}) \right\} \quad (144)$$

$$\leq I_Q(U_0; Z) + E_1(s_1(Q_{U_0X_{00}Z}), Q_{U_0Z}), \quad (145)$$

where

$$E_1(s_1(Q_{U_0X_{00}Z}), Q_{U_0Z}) \\ = \min_{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z}): f(Q_{U_0X'Z}) + [R_y - I_Q(X'; Z|U_0)]_+ \geq s_1(Q_{U_0X_{00}Z})} [I_Q(X'; Z|U_0) - R_y]_+. \quad (146)$$

Now, since  $s_1(Q_{U_0X_{00}Z})$  is given by the maximum

$$s_1(Q_{U_0X_{00}Z}) = \max \left\{ s_0(Q_{U_0Z}), f(Q_{U_0X_{00}Z}) \right\}, \quad (147)$$

we treat each case separately. First, if  $s_1(Q_{U_0X_{00}Z}) = f(Q_{U_0X_{00}Z})$ ,

$$\begin{aligned} & E_1(f(Q_{U_0X_{00}Z}), Q_{U_0Z}) \\ &= \min_{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z}): f(Q_{U_0X'Z}) + [R_y - I_Q(X'; Z|U_0)]_+ \geq f(Q_{U_0X_{00}Z})} [I_Q(X'; Z|U_0) - R_y]_+ \end{aligned} \quad (148)$$

$$\leq [I_Q(X_{00}; Z|U_0) - R_y]_+, \quad (149)$$

since the constraint is satisfied for  $Q_{X_{00}|U_0Z} \in \mathcal{S}(Q_{U_0Z})$ .

On the other hand, if  $s_1(Q_{U_0X_{00}Z}) = s_0(Q_{U_0Z})$ , which is given by (99), we have

$$s_0(Q_{U_0Z}) = R_y + f(Q_{U_0\tilde{X}Z}) - I_Q(\tilde{X}; Z|U_0), \quad (150)$$

where we have denoted the maximizer of (99) by  $Q_{\tilde{X}|U_0Z}$ , for which  $I_Q(\tilde{X}; Z|U_0) \leq R_y$  must be satisfied. Next, we upper bound the minimum defining  $E_1(s_0(Q_{U_0Z}), Q_{U_0Z})$  by using the same empirical distribution which is the maximizer of the right hand side of the constraint, for which the constraint becomes an exact equality:

$$E_1(s_0(Q_{U_0Z}), Q_{U_0Z}) \quad (151)$$

$$= \min_{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z}): f(Q_{U_0X'Z}) + [R_y - I_Q(X'; Z|U_0)]_+ \geq s_0(Q_{U_0Z})} [I_Q(X'; Z|U_0) - R_y]_+ \quad (152)$$

$$= \min_{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z}): f(Q_{U_0X'Z}) + [R_y - I_Q(X'; Z|U_0)]_+ \geq f(Q_{U_0\tilde{X}Z}) + R_y - I_Q(\tilde{X}; Z|U_0)} [I_Q(X'; Z|U_0) - R_y]_+ \quad (153)$$

$$\leq [I_Q(\tilde{X}; Z|U) - R_y]_+ = 0, \quad (154)$$

where the last equality is due to the constraint  $I_Q(\tilde{X}; Z|U) \leq R_y$ .

Combining the last two upper bounds, we get

$$E_1(s_1(Q_{U_0X_{00}Z}), Q_{U_0Z}) \leq [I_Q(X_{00}; Z|U_0) - R_y]_+, \quad (155)$$

and therefore

$$E_2(s_1(Q_{U_0X_{00}Z}), Q_Z) \leq I_Q(U_0; Z) + [I_Q(X_{00}; Z|U_0) - R_y]_+, \quad (156)$$

completing the proof of Lemma 3.  $\square$

## 5.2 Analysis for a General Suboptimal Decoding Metric and a Direct Part

Let us now derive the exact random coding error exponent of the following suboptimal bin index decoder:

$$\tilde{i}(\mathbf{z}) = \arg \max_{0 \leq i \leq M_z - 1} \left\{ \max_{0 \leq j \leq M_y - 1} f(Q_{U_i X_{ij} Z}) \right\}, \quad (157)$$

and assume, as before, that the function  $f$  is upper bounded by  $\Delta$ .

To present the formula of  $\tilde{E}_z(R_y, R_z)$ , the error exponent of (157), we first need a few new definitions. We first define

$$\tilde{E}_1(t, Q_{UZ}) = \min_{Q_{X|UZ} \in \mathcal{S}(Q_{UZ})} \left\{ [I_Q(X; Z|U) - R_y]_+ : f(Q_{UXZ}) \geq t \right\}, \quad (158)$$

where  $t$  is an arbitrary real number. Next, for a given  $Q_Z$ , define

$$\tilde{E}_2(t, Q_Z) = \min_{Q_{U|Z} \in \mathcal{S}(Q_Z)} [I_Q(U; Z) + \tilde{E}_1(t, Q_{UZ})]. \quad (159)$$

Finally, let

$$t_0(Q_{U_0 Z}) = \max_{\{Q_{X|U_0 Z} \in \mathcal{S}(Q_{U_0 Z}) : I_Q(X; Z|U_0) \leq R_y\}} f(Q_{U_0 X Z}), \quad (160)$$

and

$$t_1(Q_{U_0 X_0 Z}) = \max\{t_0(Q_{U_0 Z}), f(Q_{U_0 X_0 Z})\}. \quad (161)$$

The error exponent of (157) is given in the following lemma.

**Lemma 4.** Under the assumptions of Section 2,

$$\tilde{E}_z(R_y, R_z) = \min_{Q_{Z|U_0 X_0}} \left\{ D(Q_{Z|U_0 X_0} \| W_{Z|X_0} | P_{U_0 X_0}) + [\tilde{E}_2(t_1(Q_{U_0 X_0 Z}), Q_Z) - R_z]_+ \right\}, \quad (162)$$

where  $(U_0, X_0)$  is a replica of  $(U, X)$ , i.e.,  $P_{U_0 X_0} = P_{UX}$ .

**Proof.** The average probability of error, associated with (157), is given by

$$P_e^* \doteq \mathbb{E} \left[ \min \left\{ 1, M_z \cdot \mathbb{E} \left( \min \left[ 1, M_y \cdot \Pr \left\{ f(Q_{U_1 X_{10} Z}) \geq \max_{\mathbf{x} \in \mathcal{C}_0} f(Q_{U_0 X Z}) \right\} \right] \right) \right\} \right], \quad (163)$$

where the inner expectation is w.r.t. the randomness of  $\mathbf{U}_1$ , the outer expectation is w.r.t. the randomness of  $\mathbf{U}_0$ ,  $\mathcal{C}_0$  and  $\mathbf{Z}$ , the latter being the channel output in response to  $\mathbf{X}_{00}$ .

To see why this is true, observe that the average error probability,  $\bar{P}_e(R_y, R_z, n)$ , associated



with (157), is defined as

$$\begin{aligned} & \bar{P}_e(R_y, R_z, n) \\ & \triangleq \frac{1}{M_y M_z} \sum_{i=0}^{M_z-1} \sum_{j=0}^{M_y-1} \Pr \left\{ \bigcup_{l \neq i} \bigcup_k \left\{ f(Q_{U_l X_{lk} Z}) \geq \max_{\mathbf{x} \in \mathcal{C}_i} f(Q_{U_i X Z}) \mid \mathbf{X}_{ij} \text{ sent} \right\} \right\} \end{aligned} \quad (164)$$

where  $\Pr\{\cdot\}$  pertains to the randomness of the codebook as well as that of the channel output given its input. We define the following unions of events

$$\mathcal{G} \triangleq \bigcup_{l=1}^{M_z-1} \mathcal{G}_l \triangleq \bigcup_{l=1}^{M_z-1} \bigcup_{k=0}^{M_y-1} \mathcal{G}_{lk} \triangleq \bigcup_{l=1}^{M_z-1} \bigcup_{k=0}^{M_y-1} \left\{ f(Q_{U_l X_{lk} Z}) \geq \max_{\mathbf{x} \in \mathcal{C}_0} f(Q_{U_0 X Z}) \right\}. \quad (165)$$

Define

$$t \triangleq \max_{\mathbf{x} \in \mathcal{C}_0} f(Q_{U_0 X Z}). \quad (166)$$

The probability of  $\mathcal{G}_{lk}$ , conditioned on  $\mathbf{U}_l$ , is given by

$$\Pr(\mathcal{G}_{lk} \mid \mathbf{U}_l = \mathbf{u}') \triangleq \Pr \left\{ f(Q_{U_l X_{lk} Z}) \geq \max_{\mathbf{x} \in \mathcal{C}_0} f(Q_{U_0 X Z}) \mid \mathbf{U}_l = \mathbf{u}' \right\} \quad (167)$$

$$= \sum_{\{\mathbf{x}': f(Q_{U' X' Z}) \geq t\}} P(\mathbf{x}' \mid \mathbf{u}') \quad (168)$$

$$\doteq \max_{\{Q_{X'|U'Z} \in \mathcal{S}(Q_{U'Z}): f(Q_{U' X' Z}) \geq t\}} \exp \left\{ -n \cdot I_Q(X'; Z \mid U') \right\} \quad (169)$$

$$= \exp \left\{ -n \cdot \min_{\{Q_{X'|U'Z} \in \mathcal{S}(Q_{U'Z}): f(Q_{U' X' Z}) \geq t\}} I_Q(X'; Z \mid U') \right\} \quad (170)$$

$$\triangleq \exp \left\{ -n \cdot \tilde{E}_0(t, Q_{U'Z}) \right\}, \quad (171)$$

where the passage from (168) to (169) is due to (47)-(51). For a given  $\mathbf{U}_l = \mathbf{u}'$ , the events  $\{\mathcal{G}_{lk}\}_k$  are all pairwise independent since we have assumed that the various codewords are pairwise conditional independent given the cloud center. Using the exponential tightness of the truncated union bound, we get

$$\Pr \left\{ \bigcup_{k=0}^{M_y-1} \mathcal{G}_{lk} \mid \mathbf{U}_l = \mathbf{u}' \right\} \doteq \min \left\{ 1, \sum_{k=0}^{M_y-1} \Pr(\mathcal{G}_{lk} \mid \mathbf{U}_l = \mathbf{u}') \right\} \quad (172)$$

$$= \min \left\{ 1, M_y \cdot \Pr(\mathcal{G}_{l,0} \mid \mathbf{U}_l = \mathbf{u}') \right\} \quad (173)$$

$$\doteq \min \left\{ 1, e^{nR_y} \cdot \exp \left\{ -n \cdot \tilde{E}_0(t, Q_{U'Z}) \right\} \right\} \quad (174)$$

$$\triangleq \exp \left\{ -n \cdot \tilde{E}_1(t, Q_{U'Z}) \right\}, \quad (175)$$

where

$$\tilde{E}_1(t, Q_{U'Z}) = \min_{Q_{X'|U'Z} \in \mathcal{S}(Q_{U'Z})} \left\{ [I_Q(X'; Z|U') - R_y]_+ : f(Q_{U'X'Z}) \geq t \right\}. \quad (176)$$

Next, we obtain the probability of  $\mathcal{G}_l$  by calculating the expectation w.r.t. the randomness of  $\mathbf{U}_l$ :

$$\Pr\{\mathcal{G}_l\} = \sum_{\mathbf{u}' \in \mathcal{T}(P_U)} P_U(\mathbf{u}') \cdot \Pr \left\{ \bigcup_{k=0}^{M_y-1} \mathcal{G}_{lk} \mid \mathbf{U}_l = \mathbf{u}' \right\} \quad (177)$$

$$\doteq \sum_{\mathbf{u}' \in \mathcal{T}(P_U)} P_U(\mathbf{u}') \cdot \exp \left\{ -n \cdot \tilde{E}_1(t, Q_{U'Z}) \right\} \quad (178)$$

$$\doteq \exp \left\{ -n \cdot \min_{\{Q_{U'|Z} \in \mathcal{S}(Q_Z)\}} \left[ I_Q(U; Z) + \tilde{E}_1(t, Q_{U'Z}) \right] \right\} \quad (179)$$

$$\triangleq \exp \left\{ -n \cdot \tilde{E}_2(t, Q_Z) \right\}, \quad (180)$$

where the passage from (178) to (179) is due to (60)-(64). Conditioning on  $\mathbf{U}_0$ ,  $\mathcal{C}_0$  and  $\mathbf{Z}$ , the events  $\{\mathcal{G}_l\}$  are all pairwise independent since the various cloud centers are all independent.

We get

$$\Pr \left\{ \bigcup_{l=1}^{M_z-1} \mathcal{G}_l \mid \mathbf{U}_0, \mathcal{C}_0, \mathbf{Z} \right\} \doteq \min \left\{ 1, \sum_{l=1}^{M_z-1} \Pr(\mathcal{G}_l \mid \mathbf{U}_0, \mathcal{C}_0, \mathbf{Z}) \right\} \quad (181)$$

$$= \min \left\{ 1, (M_z - 1) \cdot \Pr(\mathcal{G}_1 \mid \mathbf{U}_0, \mathcal{C}_0, \mathbf{Z}) \right\} \quad (182)$$

$$\doteq \min \left\{ 1, e^{nR_z} \cdot \exp \left\{ -n \cdot \tilde{E}_2(t, Q_Z) \right\} \right\} \quad (183)$$

$$= \exp \left\{ -n \cdot [\tilde{E}_2(t, Q_Z) - R_z]_+ \right\}. \quad (184)$$

Finally, we have that

$$\tilde{P}_e = \mathbb{E} \left[ \exp \left\{ -n \cdot [\tilde{E}_2(T, Q_Z) - R_z]_+ \right\} \right], \quad (185)$$

where the expectation is taken w.r.t. the randomness of

$$T = \max_{\mathbf{x} \in \mathcal{C}_0} f(Q_{U_0XZ}), \quad (186)$$

and the randomness of  $Q_Z$  and  $U_0$ , the correct cloud center. This expectation will be taken in two steps, first, over the randomness of  $\{\mathbf{X}_{0,1}, \dots, \mathbf{X}_{0,(M_y-1)}\}$ , while  $\mathbf{X}_{00} = \mathbf{x}_{00}$ ,  $\mathbf{U}_0 = \mathbf{u}_0$  and  $\mathbf{Z} = \mathbf{z}$  are held fixed, and then - over the randomness of  $\mathbf{X}_{00}$ ,  $\mathbf{U}_0$  and  $\mathbf{Z}$ . Let  $\mathbf{x}_{00}$ ,  $\mathbf{u}_0$  and  $\mathbf{z}$  be

given and let  $\epsilon > 0$  be arbitrarily small. Then,

$$\tilde{P}_\epsilon(\mathbf{x}_{00}, \mathbf{u}_0, \mathbf{z}) \triangleq \mathbb{E} \left\{ e^{-n[\tilde{E}_2(T, Q_Z) - R_z]_+} \middle| \mathbf{X}_{00} = \mathbf{x}_{00}, \mathbf{U}_0 = \mathbf{u}_0, \mathbf{Z} = \mathbf{z} \right\} \quad (187)$$

$$\leq \sum_i \Pr \left\{ i\epsilon \leq T < (i+1)\epsilon \middle| \mathbf{X}_{00} = \mathbf{x}_{00}, \mathbf{U}_0 = \mathbf{u}_0, \mathbf{Z} = \mathbf{z} \right\} \\ \times \exp \left\{ -n[\tilde{E}_2(i\epsilon, Q_Z) - R_z]_+ \right\}, \quad (188)$$

where  $i$  ranges from  $\frac{1}{\epsilon}f(Q_{U_0 X_{00} Z})$  to  $\Delta/\epsilon$ . Now,

$$T = \max \left\{ f(Q_{U_0 X_{00} Z}), \max_{1 \leq j \leq M_y - 1} f(Q_{U_0 X_{0j} Z}) \right\}. \quad (189)$$

On the one hand, we have:

$$\Pr \left\{ \max_{1 \leq j \leq M_y - 1} f(Q_{U_0 X_{0j} Z}) \geq t \right\} = \Pr \left\{ \bigcup_{1 \leq j \leq M_y - 1} \left\{ f(Q_{U_0 X_{0j} Z}) \geq t \right\} \right\} \quad (190)$$

$$\doteq \min \left\{ 1, (M_y - 1) \cdot \Pr \left\{ f(Q_{U_0 X_{01} Z}) \geq t \right\} \right\} \quad (191)$$

$$\doteq \min \left\{ 1, e^{nR_y} \cdot \exp \left\{ -n \cdot \tilde{E}_0(t, Q_{U_0 Z}) \right\} \right\} \quad (192)$$

$$= \exp \left\{ -n[\tilde{E}_0(t, Q_{U_0 Z}) - R_y]_+ \right\} \quad (193)$$

$$= \exp \left\{ -n \cdot \tilde{E}_1(t, Q_{U_0 Z}) \right\}. \quad (194)$$

On the other hand,

$$\Pr \left\{ \max_{1 \leq j \leq M_y - 1} f(Q_{U_0 X_{0j} Z}) < t \right\} = \Pr \left\{ \bigcap_{1 \leq j \leq M_y - 1} \left\{ f(Q_{U_0 X_{0j} Z}) < t \right\} \right\} \quad (195)$$

$$= \left[ \Pr \left\{ f(Q_{U_0 X_{01} Z}) < t \right\} \right]^{M_y - 1} \quad (196)$$

$$\doteq \left[ 1 - e^{-n \cdot \tilde{E}_0(t, Q_{U_0 Z})} \right]^{e^{nR_y}} \quad (197)$$

$$= \exp \left\{ e^{nR_y} \cdot \ln \left[ 1 - e^{-n \cdot \tilde{E}_0(t, Q_{U_0 Z})} \right] \right\} \quad (198)$$

$$\doteq \exp \left\{ -e^{nR_y} [\tilde{E}_0(t, Q_{U_0 Z})] \right\} \quad (199)$$

$$\doteq \begin{cases} 0, & R_y > \tilde{E}_0(t, Q_{U_0 Z}) \\ 1, & R_y < \tilde{E}_0(t, Q_{U_0 Z}), \end{cases} \quad (200)$$

which can also be written as:

$$\Pr \left\{ \max_{1 \leq j \leq M_y - 1} f(Q_{U_0 X_{0j} Z}) < t \right\} \doteq \mathcal{I} \left\{ R_y < \tilde{E}_0(t, Q_{U_0 Z}) \right\}. \quad (201)$$

Let us now find the minimum  $t$  for which the value of this indicator function is unity. The condition is equivalent to

$$\min_{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z})} \max_{0 \leq a < \infty} \left\{ I_Q(X'; Z|U_0) + a \left[ t - f(Q_{U_0X'Z}) \right] \right\} \geq R_y, \quad (202)$$

or

$$\forall Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z}) \exists a \in [0, \infty) : I_Q(X'; Z|U_0) + a \left[ t - f(Q_{U_0X'Z}) \right] \geq R_y, \quad (203)$$

or

$$\forall Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z}) \exists a \in [0, \infty) : t \geq f(Q_{U_0X'Z}) + \frac{1}{a} \left( R_y - I_Q(X'; Z|U_0) \right), \quad (204)$$

or, equivalently,

$$t \geq \max_{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z})} \min_{0 \leq a < \infty} \left[ f(Q_{U_0X'Z}) + \frac{1}{a} \left( R_y - I_Q(X'; Z|U_0) \right) \right] \quad (205)$$

$$= \max_{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z})} \left[ f(Q_{U_0X'Z}) + \begin{cases} 0, & R_y \geq I_Q(X'; Z|U_0) \\ -\infty, & R_y < I_Q(X'; Z|U_0) \end{cases} \right] \quad (206)$$

$$= \max_{\{Q_{X'|U_0Z} \in \mathcal{S}(Q_{U_0Z}) : I_Q(X'; Z|U_0) \leq R_y\}} f(Q_{U_0X'Z}) \quad (207)$$

$$\triangleq t_0(Q_{U_0Z}). \quad (208)$$

Thus, in summary, we have

$$\Pr \left\{ t \leq \max_{1 \leq j \leq M_y - 1} f(Q_{U_0X_{0j}Z}) \leq t + \epsilon \right\} \doteq \begin{cases} 0 & , t < t_0(Q_{U_0Z}) - \epsilon \\ e^{-n \cdot \tilde{E}_1(t, Q_{U_0Z})} & , t \geq t_0(Q_{U_0Z}) \end{cases} \quad (209)$$

Then, the expected error probability w.r.t.  $\{\mathbf{X}_{0,1}, \dots, \mathbf{X}_{0,(M_y-1)}\}$  yields

$$\begin{aligned} \tilde{P}_e(\mathbf{x}_{00}, \mathbf{u}_0, \mathbf{z}) &\doteq \sum_i \Pr \left\{ i\epsilon \leq \max_{1 \leq j \leq M_y - 1} f(Q_{U_0X_{0j}Z}) < (i+1)\epsilon \right\} \\ &\quad \times \exp \left\{ -n \left[ \tilde{E}_2(\max\{f(Q_{U_0X_{00}Z}), i\epsilon\}, Q_Z) - R_z \right]_+ \right\} \end{aligned} \quad (210)$$

$$\begin{aligned} &\doteq \sum_{i \geq t_0(Q_{U_0Z})/\epsilon} \exp \left\{ -n \tilde{E}_1(i\epsilon, Q_{U_0Z}) \right\} \\ &\quad \times \exp \left\{ -n \left[ \tilde{E}_2(\max\{f(Q_{U_0X_{00}Z}), i\epsilon\}, Q_Z) - R_z \right]_+ \right\}. \end{aligned} \quad (211)$$

Since the dominant contribution to the sum over  $i$  is due to the term  $i = t_0(Q_{U_0Z})/\epsilon$  (by the non-decreasing monotonicity of the functions  $\tilde{E}_1(\cdot, Q_{U_0Z})$  and  $\tilde{E}_2(\cdot, Q_Z)$ ), we obtain

$$\tilde{P}_e(\mathbf{x}_{00}, \mathbf{u}_0, \mathbf{z}) \doteq \exp \left\{ -n \left[ \tilde{E}_2(\max\{t_0(Q_{U_0Z}), f(Q_{U_0X_{00}Z})\}, Q_Z) - R_z \right]_+ \right\} \quad (212)$$

$$\triangleq \exp \left\{ -n \left[ \tilde{E}_2(t_1(Q_{U_0X_{00}Z}), Q_Z) - R_z \right]_+ \right\}. \quad (213)$$

Now, after taking the expectation w.r.t. the joint distribution of  $(\mathbf{U}_0, \mathbf{X}_{00}, \mathbf{Z})$ , we get

$$\tilde{E}_z(R_y, R_z) = \min_{Q_{Z|U_0, X_{00}}} \left\{ D(Q_{Z|U_0, X_{00}} \| W_{Z|X_{00}} | P_{U_0, X_{00}}) + [\tilde{E}_2(t_1(Q_{U_0, X_{00}}, Z), Q_Z) - R_z]_+ \right\}, \quad (214)$$

and the proof of Lemma 4 is complete.  $\square$

Let us now select

$$f(Q_{UXZ}) = I_Q(U; Z) + [I_Q(X; Z|U) - R_y]_+. \quad (215)$$

We show that (215) achieves the maximum of  $E_2(s_1(Q_{U_0, X_{00}}, Z), Q_Z)$ , as given by Lemma 3, and therefore, this decoder has the same error exponent as that of the optimal decoder. First, the threshold  $t_0(Q_{U_0, Z})$  can be easily simplified as

$$t_0(Q_{U_0, Z}) = \max_{\{Q_{X|U_0, Z} \in \mathcal{S}(Q_{U_0, Z}): I_Q(X; Z|U_0) \leq R_y\}} f(Q_{U_0, XZ}) \quad (216)$$

$$= \max_{\{Q_{X|U_0, Z} \in \mathcal{S}(Q_{U_0, Z}): I_Q(X; Z|U_0) \leq R_y\}} \left\{ I_Q(U_0; Z) + [I_Q(X; Z|U_0) - R_y]_+ \right\} \quad (217)$$

$$= I_Q(U_0; Z) + \max_{\{Q_{X|U_0, Z} \in \mathcal{S}(Q_{U_0, Z}): I_Q(X; Z|U_0) \leq R_y\}} [I_Q(X; Z|U_0) - R_y]_+ \quad (218)$$

$$= I_Q(U_0; Z). \quad (219)$$

Now,  $t_1(Q_{U_0, X_{00}}, Z)$  is given by

$$t_1(Q_{U_0, X_{00}}, Z) = \max \left\{ I_Q(U_0; Z), I_Q(U_0; Z) + [I_Q(X_{00}; Z|U_0) - R_y]_+ \right\} \quad (220)$$

$$= I_Q(U_0; Z) + [I_Q(X_{00}; Z|U_0) - R_y]_+. \quad (221)$$

In general, the constraint of the inner minimization problem defining  $\tilde{E}_2(t_1(Q_{U_0, X_{00}}, Z), Q_Z)$  is given by

$$f(Q_{UXZ}) \geq t_1(Q_{U_0, X_{00}}, Z), \quad (222)$$

which can now be written as

$$I_Q(U; Z) + [I_Q(X; Z|U) - R_y]_+ \geq I_Q(U_0; Z) + [I_Q(X_{00}; Z|U_0) - R_y]_+, \quad (223)$$

or simply by  $f(Q_{UXZ}) \geq f(Q_{U_0, X_{00}}, Z)$ . Eventually, we have the following

$$\tilde{E}_2(t_1(Q_{U_0, X_{00}}, Z), Q_Z) = \min_{Q_{UX|Z} \in \mathcal{S}(Q_Z): f(Q_{UXZ}) \geq f(Q_{U_0, X_{00}}, Z)} f(Q_{UXZ}) \quad (224)$$

$$= I_Q(U_0; Z) + [I_Q(X_{00}; Z|U_0) - R_y]_+, \quad (225)$$

which is the same expression as on the right hand side of (143).

## 6 Gallager-Style Lower Bounds

In this section, we prove Theorem 5.

### 6.1 Derivation of eq. (32)

We start by changing the clipping operator to a maximization problem and using convexity properties to change the order of the maximization and the minimization:

$$\hat{E}_{y,1}(R_y) \triangleq \min_V \left\{ D(V\|W|P) + [I(X;Y|U) - R_y]_+ \right\} \quad (226)$$

$$= \min_V \left\{ D(V\|W|P) + \max_{\rho \in [0,1]} \left\{ \rho \cdot [I(X;Y|U) - R_y] \right\} \right\} \quad (227)$$

$$= \max_{\rho \in [0,1]} \left\{ -\rho R_y + \min_V \left\{ D(V\|W|P) + \rho \cdot I(X;Y|U) \right\} \right\}. \quad (228)$$

Next,

$$\begin{aligned} & \min_V \left\{ D(V\|W|P) + \rho \cdot I(X;Y|U) \right\} \\ &= \min_{V,Q} \left\{ \sum_{x,u} P(x,u) \sum_y V(y|x,u) \log \frac{V(y|x,u)}{W(y|x)} \right. \\ & \quad \left. + \rho \sum_{x,u,y} P(x,u)V(y|x,u) \log \frac{V(y|x,u)}{Q(y|u)} \right\} \end{aligned} \quad (229)$$

$$= \min_{V,Q} \left\{ \sum_{x,u,y} P(x,u)V(y|x,u) \left[ \log \frac{V(y|x,u)}{W(y|x)} + \rho \log \frac{V(y|x,u)}{Q(y|u)} \right] \right\} \quad (230)$$

$$= \min_{V,Q} \left\{ \sum_{x,u,y} P(x,u)V(y|x,u) \log \left[ \frac{V^{1+\rho}(y|x,u)}{W(y|x)Q^\rho(y|u)} \right] \right\}. \quad (231)$$

First, we minimize over the auxiliary channel  $V$ . Holding the auxiliary channel  $Q$  fixed, and differentiating w.r.t.  $V(y|x,u)$ , we find that the minimizing distribution is given by

$$V^*(y|x,u) = \frac{W^{\frac{1}{1+\rho}}(y|x)Q^{\frac{\rho}{1+\rho}}(y|u)}{\sum_{y'} W^{\frac{1}{1+\rho}}(y'|x)Q^{\frac{\rho}{1+\rho}}(y'|u)}. \quad (232)$$

Substituting it back into (231) and summing over  $y$ , we get that

$$\begin{aligned} & \min_Q \left\{ \sum_{x,u,y} P(x,u)V^*(y|x,u) \log \left[ \frac{[V^*(y|x,u)]^{1+\rho}}{W(y|x)Q^\rho(y|u)} \right] \right\} \\ &= \min_Q \left\{ -(1+\rho) \sum_{x,u} P(x,u) \log \left[ \sum_y W^{\frac{1}{1+\rho}}(y|x)Q^{\frac{\rho}{1+\rho}}(y|u) \right] \right\} \end{aligned} \quad (233)$$

$$= \min_Q \left\{ -(1 + \rho) \sum_u P(u) \sum_x P(x|u) \log \left[ \sum_y W^{\frac{1}{1+\rho}}(y|x) Q^{\frac{\rho}{1+\rho}}(y|u) \right] \right\} \quad (234)$$

$$\geq \min_Q \left\{ -(1 + \rho) \sum_u P(u) \log \left[ \sum_x P(x|u) \sum_y W^{\frac{1}{1+\rho}}(y|x) Q^{\frac{\rho}{1+\rho}}(y|u) \right] \right\}, \quad (235)$$

where inequality (235) is due to Jensen's inequality. Next, we minimize the lower bound over  $Q$ . Differentiating the last expression w.r.t.  $Q(y|u)$ , we find that the minimizing distribution is given by

$$Q^*(y|u) = \frac{[\Phi(u, y, \rho)]^{1+\rho}}{\sum_{y'} [\Phi(u, y', \rho)]^{1+\rho}}. \quad (236)$$

Substituting (236) into (235), we get

$$\begin{aligned} & -(1 + \rho) \sum_u P(u) \log \left[ \sum_x P(x|u) \sum_y W^{\frac{1}{1+\rho}}(y|x) [Q^*(y|u)]^{\frac{\rho}{1+\rho}} \right] \\ &= -(1 + \rho) \sum_u P(u) \log \left[ \sum_x P(x|u) \sum_y W^{\frac{1}{1+\rho}}(y|x) \frac{[\Phi(u, y, \rho)]^\rho}{\left\{ \sum_{y'} [\Phi(u, y', \rho)]^{1+\rho} \right\}^{\frac{\rho}{1+\rho}}} \right] \end{aligned} \quad (237)$$

$$= -(1 + \rho) \sum_u P(u) \log \left[ \frac{\sum_y \left( \Phi(u, y, \rho) \cdot [\Phi(u, y, \rho)]^\rho \right)}{\left\{ \sum_{y'} [\Phi(u, y', \rho)]^{1+\rho} \right\}^{\frac{\rho}{1+\rho}}} \right] \quad (238)$$

$$= -(1 + \rho) \sum_u P(u) \log \left[ \frac{\sum_y [\Phi(u, y, \rho)]^{1+\rho}}{\left\{ \sum_{y'} [\Phi(u, y', \rho)]^{1+\rho} \right\}^{\frac{\rho}{1+\rho}}} \right] \quad (239)$$

$$= -(1 + \rho) \sum_u P(u) \log \left[ \left\{ \sum_y [\Phi(u, y, \rho)]^{1+\rho} \right\}^{\frac{1}{1+\rho}} \right] \quad (240)$$

$$= - \sum_u P(u) \log \left\{ \sum_y [\Phi(u, y, \rho)]^{1+\rho} \right\}, \quad (241)$$

which completes the proof of eq. (32).  $\square$

## 6.2 Derivation of eq. (33)

Similarly as in (226)-(228),

$$\hat{E}_{y,2}(R_y, R_z)$$

$$\triangleq \min_V \left\{ D(V\|W|P) + \left[ I(U;Y) + \left[ I(X;Y|U) - R_y \right]_+ - R_z \right]_+ \right\} \quad (242)$$

$$= \min_V \left\{ D(V\|W|P) + \max_{\mu \in [0,1]} \left\{ \mu \cdot \left[ I(U;Y) + \max_{\rho \in [0,1]} \left\{ \rho \cdot \left[ I(X;Y|U) - R_y \right] \right\} - R_z \right] \right\} \right\} \quad (243)$$

$$= \min_V \left\{ D(V\|W|P) + \max_{\mu \in [0,1]} \max_{\rho \in [0,1]} \left\{ \mu \cdot \left[ I(U;Y) - R_z \right] + \mu\rho \cdot \left[ I(X;Y|U) - R_y \right] \right\} \right\} \quad (244)$$

$$= \min_V \left\{ D(V\|W|P) + \max_{\mu \in [0,1]} \max_{\lambda \in [0,\mu]} \left\{ \mu \cdot \left[ I(U;Y) - R_z \right] + \lambda \cdot \left[ I(X;Y|U) - R_y \right] \right\} \right\} \quad (245)$$

$$= \max_{\mu \in [0,1]} \max_{\lambda \in [0,\mu]} \left\{ -\lambda R_y - \mu R_z + \min_V \left\{ D(V\|W|P) + \mu \cdot I(U;Y) + \lambda \cdot I(X;Y|U) \right\} \right\}. \quad (246)$$

Now, for the inner-most minimization,

$$\begin{aligned} & \min_V \left\{ D(V\|W|P) + \mu \cdot I(U;Y) + \lambda \cdot I(X;Y|U) \right\} \\ &= \min_{V,Q,T} \left\{ \sum_{x,u} P(x,u) \sum_y V(y|x,u) \log \frac{V(y|x,u)}{W(y|x)} \right. \\ & \quad \left. + \mu \sum_{x,u,y} P(x,u) V(y|x,u) \log \frac{Q(y|u)}{T(y)} + \lambda \sum_{x,u,y} P(x,u) V(y|x,u) \log \frac{V(y|x,u)}{Q(y|u)} \right\} \end{aligned} \quad (247)$$

$$= \min_{V,Q,T} \left\{ \sum_{x,u,y} P(x,u) V(y|x,u) \left[ \log \frac{V(y|x,u)}{W(y|x)} + \mu \log \frac{Q(y|u)}{T(y)} + \lambda \log \frac{V(y|x,u)}{Q(y|u)} \right] \right\} \quad (248)$$

$$= \min_{V,Q,T} \left\{ \sum_{x,u,y} P(x,u) V(y|x,u) \log \left[ \frac{V^{1+\lambda}(y|x,u)}{W(y|x) T^\mu(y) Q^{\lambda-\mu}(y|u)} \right] \right\}. \quad (249)$$

First, we minimize over  $V$ . Holding  $Q$  and  $T$  fixed, and differentiating w.r.t.  $V(y|x,u)$ , we find that the minimizing  $V$  is given by

$$V^*(y|x,u) = \frac{W^{\frac{1}{1+\lambda}}(y|x) T^{\frac{\mu}{1+\lambda}}(y) Q^{\frac{\lambda-\mu}{1+\lambda}}(y|u)}{\sum_{y'} W^{\frac{1}{1+\lambda}}(y'|x) T^{\frac{\mu}{1+\lambda}}(y') Q^{\frac{\lambda-\mu}{1+\lambda}}(y'|u)}. \quad (250)$$

Substituting (250) into (249) and summing over  $y$ , we get

$$\begin{aligned} & \min_{Q,T} \left\{ \sum_{x,u,y} P(x,u) V^*(y|x,u) \log \left[ \frac{[V^*(y|x,u)]^{1+\lambda}}{W(y|x) T^\mu(y) Q^{\lambda-\mu}(y|u)} \right] \right\} \\ &= \min_{Q,T} \left\{ -(1+\lambda) \sum_{x,u} P(x,u) \log \left[ \sum_y W^{\frac{1}{1+\lambda}}(y|x) T^{\frac{\mu}{1+\lambda}}(y) Q^{\frac{\lambda-\mu}{1+\lambda}}(y|u) \right] \right\} \end{aligned} \quad (251)$$

$$= \min_{Q,T} \left\{ -(1+\lambda) \sum_u P(u) \sum_x P(x|u) \log \left[ \sum_y W^{\frac{1}{1+\lambda}}(y|x) T^{\frac{\mu}{1+\lambda}}(y) Q^{\frac{\lambda-\mu}{1+\lambda}}(y|u) \right] \right\} \quad (252)$$

$$\geq \min_{Q,T} \left\{ -(1+\lambda) \sum_u P(u) \log \left[ \sum_x P(x|u) \sum_y W^{\frac{1}{1+\lambda}}(y|x) T^{\frac{\mu}{1+\lambda}}(y) Q^{\frac{\lambda-\mu}{1+\lambda}}(y|u) \right] \right\}, \quad (253)$$



where (253) is due to Jensen's inequality. Next, we minimize the lower bound over  $Q$ , while holding  $T$  fixed. Differentiating the last expression w.r.t.  $Q(y|u)$ , we find that the minimizing  $Q$  is given by

$$Q^*(y|u) = \frac{\left[ T^{\frac{\mu}{1+\lambda}}(y) \sum_x P(x|u) W^{\frac{1}{1+\lambda}}(y|x) \right]^{\frac{1+\lambda}{1+\mu}}}{\sum_{y'} \left[ T^{\frac{\mu}{1+\lambda}}(y') \sum_{x'} P(x'|u) W^{\frac{1}{1+\lambda}}(y'|x') \right]^{\frac{1+\lambda}{1+\mu}}}. \quad (254)$$

Substituting into (253), we have

$$\begin{aligned} & \min_T \left\{ -(1+\lambda) \sum_u P(u) \log \left[ \sum_x P(x|u) \sum_y W^{\frac{1}{1+\lambda}}(y|x) T^{\frac{\mu}{1+\lambda}}(y) [Q^*(y|u)]^{\frac{\lambda-\mu}{1+\lambda}} \right] \right\} \\ &= \min_T -(1+\lambda) \sum_u P(u) \log \left[ \sum_x P(x|u) \right. \\ & \quad \left. \times \sum_y W^{\frac{1}{1+\lambda}}(y|x) T^{\frac{\mu}{1+\lambda}}(y) \frac{\left[ T^{\frac{\mu}{1+\lambda}}(y) \Phi(u, y, \lambda) \right]^{\frac{1+\lambda}{1+\mu} \frac{\lambda-\mu}{1+\lambda}}}{\left\{ \sum_{y'} \left[ T^{\frac{\mu}{1+\lambda}}(y') \Phi(u, y', \lambda) \right]^{\frac{1+\lambda}{1+\mu}} \right\}^{\frac{\lambda-\mu}{1+\lambda}}} \right] \end{aligned} \quad (255)$$

$$= \min_T -(1+\lambda) \sum_u P(u) \log \left[ \frac{\sum_y \left[ T^{\frac{\mu}{1+\lambda}}(y) \Phi(u, y, \lambda) \right] \cdot \left[ T^{\frac{\mu}{1+\lambda}}(y) \Phi(u, y, \lambda) \right]^{\frac{\lambda-\mu}{1+\mu}}}{\left\{ \sum_{y'} \left[ T^{\frac{\mu}{1+\lambda}}(y') \Phi(u, y', \lambda) \right]^{\frac{1+\lambda}{1+\mu}} \right\}^{\frac{\lambda-\mu}{1+\lambda}}} \right] \quad (256)$$

$$= \min_T -(1+\lambda) \sum_u P(u) \log \left[ \frac{\sum_y \left[ T^{\frac{\mu}{1+\lambda}}(y) \Phi(u, y, \lambda) \right]^{\frac{1+\lambda}{1+\mu}}}{\left\{ \sum_{y'} \left[ T^{\frac{\mu}{1+\lambda}}(y') \Phi(u, y', \lambda) \right]^{\frac{1+\lambda}{1+\mu}} \right\}^{\frac{\lambda-\mu}{1+\lambda}}} \right] \quad (257)$$

$$= \min_T -(1+\lambda) \sum_u P(u) \log \left[ \left\{ \sum_y \left[ T^{\frac{\mu}{1+\lambda}}(y) \Phi(u, y, \lambda) \right]^{\frac{1+\lambda}{1+\mu}} \right\}^{\frac{1+\mu}{1+\lambda}} \right] \quad (258)$$

$$= \min_T -(1+\mu) \sum_u P(u) \log \left[ \sum_y \left[ T^{\frac{\mu}{1+\lambda}}(y) \Phi(u, y, \lambda) \right]^{\frac{1+\lambda}{1+\mu}} \right] \quad (259)$$

$$= \min_T -(1+\mu) \sum_u P(u) \log \left[ \sum_y T^{\frac{\mu}{1+\mu}}(y) [\Phi(u, y, \lambda)]^{\frac{1+\lambda}{1+\mu}} \right] \quad (260)$$

$$\geq \min_T -(1+\mu) \log \left[ \sum_u P(u) \sum_y T^{\frac{\mu}{1+\mu}}(y) [\Phi(u, y, \lambda)]^{\frac{1+\lambda}{1+\mu}} \right]. \quad (261)$$

Next, we minimize over  $T$ . Differentiating w.r.t.  $T(y)$ , we get

$$T^*(y) = \frac{\left\{ \sum_u P(u) [\Phi(u, y, \lambda)]^{\frac{1+\lambda}{1+\mu}} \right\}^{1+\mu}}{\sum_{y'} \left\{ \sum_{u'} P(u') [\Phi(u', y', \lambda)]^{\frac{1+\lambda}{1+\mu}} \right\}^{1+\mu}}. \quad (262)$$

Substituting into (261), we finally get

$$\begin{aligned} & - (1 + \mu) \log \left[ \sum_u P(u) \sum_y [T^*(y)]^{\frac{\mu}{1+\mu}} [\Phi(u, y, \lambda)]^{\frac{1+\lambda}{1+\mu}} \right] \\ &= - (1 + \mu) \log \left[ \sum_u P(u) \sum_y [\Phi(u, y, \lambda)]^{\frac{1+\lambda}{1+\mu}} \right. \\ & \quad \left. \times \frac{\left\{ \sum_{\tilde{u}} P(\tilde{u}) [\Phi(\tilde{u}, y, \lambda)]^{\frac{1+\lambda}{1+\mu}} \right\}^{\mu}}{\left\{ \sum_{y'} \left\{ \sum_{u'} P(u') [\Phi(u', y', \lambda)]^{\frac{1+\lambda}{1+\mu}} \right\}^{1+\mu} \right\}^{\frac{\mu}{1+\mu}}} \right] \end{aligned} \quad (263)$$

$$= - (1 + \mu) \log \left[ \frac{\sum_y \left( \left\{ \sum_u P(u) [\Phi(u, y, \lambda)]^{\frac{1+\lambda}{1+\mu}} \right\} \cdot \left\{ \sum_{\tilde{u}} P(\tilde{u}) [\Phi(\tilde{u}, y, \lambda)]^{\frac{1+\lambda}{1+\mu}} \right\}^{\mu} \right)}{\left\{ \sum_{y'} \left\{ \sum_{u'} P(u') [\Phi(u', y', \lambda)]^{\frac{1+\lambda}{1+\mu}} \right\}^{1+\mu} \right\}^{\frac{\mu}{1+\mu}}} \right] \quad (264)$$

$$= - (1 + \mu) \log \left[ \frac{\sum_y \left\{ \sum_u P(u) [\Phi(u, y, \lambda)]^{\frac{1+\lambda}{1+\mu}} \right\}^{1+\mu}}{\left\{ \sum_{y'} \left\{ \sum_{u'} P(u') [\Phi(u', y', \lambda)]^{\frac{1+\lambda}{1+\mu}} \right\}^{1+\mu} \right\}^{\frac{\mu}{1+\mu}}} \right] \quad (265)$$

$$= - (1 + \mu) \log \left\{ \sum_y \left\{ \sum_u P(u) [\Phi(u, y, \lambda)]^{\frac{1+\lambda}{1+\mu}} \right\}^{1+\mu} \right\}^{\frac{1}{1+\mu}} \quad (266)$$

$$= - \log \sum_y \left\{ \sum_u P(u) [\Phi(u, y, \lambda)]^{\frac{1+\lambda}{1+\mu}} \right\}^{1+\mu}. \quad (267)$$

Hence, (33) is now proved, as well as the lower bound given in (34).  $\square$

## 7 Analyzing the Gallager-Style Lower Bounds

### 7.1 A Study for $E_{y,1}(R_y)$

As in the single user case, we expect to find a critical rate and a maximal rate. By maximal rate, that will be denoted by  $R_{\max}$ , we mean  $\sup\{R_y : E_{y,1}(R_y) > 0\}$ . By critical rate, to be

denoted by  $R_{\text{crit}}$ , we mean the boundary between the range where  $E_{y,1}(R_y)$  is affine and the range where it is curvy.

Let

$$E_{y,1}(R_y) = \max_{\rho \in [0,1]} \left\{ E_0(\rho) - \rho R_y \right\}, \quad (268)$$

where we have defined

$$E_0(\rho) = - \sum_u P(u) \log \sum_y \left[ \sum_x P(x|u) W^{\frac{1}{1+\rho}}(y|x) \right]^{1+\rho}. \quad (269)$$

Setting the partial derivative of the bracketed part of (268) equal to 0, we get

$$R_y = \frac{\partial E_0(\rho)}{\partial \rho}. \quad (270)$$

Following the same considerations as in [12, Section 5.6], if some  $\rho \in [0, 1]$  satisfies (270), then it must maximize (268). It turns out that a solution to (270) exists if

$$\frac{\partial E_0(\rho)}{\partial \rho} \Big|_{\rho=1} \leq R_y \leq \frac{\partial E_0(\rho)}{\partial \rho} \Big|_{\rho=0}. \quad (271)$$

In this range, it is convenient to use (270) to relate  $E_{y,1}(R_y)$  and  $R_y$  parametrically as functions of  $\rho$ . For the interval  $0 \leq \rho \leq 1$ , this gives

$$E_{y,1}(R_y) = E_0(\rho) - \rho \cdot \frac{\partial E_0(\rho)}{\partial \rho}, \quad (272)$$

$$R_y = \frac{\partial E_0(\rho)}{\partial \rho}. \quad (273)$$

For  $R_y < \partial E_0(\rho)/\partial \rho|_{\rho=1}$ , the parametric equations are not valid. In this case, the maximum occurs at  $\rho = 1$ . Thus,  $E_{y,1}(R_y)$  is affine with slope  $-1$ :

$$E_{y,1}(R_y) = E_0(1) - R_y, \quad (274)$$

where

$$E_0(1) = - \sum_u P(u) \log \sum_y \left[ \sum_x P(x|u) \sqrt{W(y|x)} \right]^2. \quad (275)$$

Now, we can find  $R_{\text{max}}$  and  $R_{\text{crit}}$ , which are given by the right-most side and the left-most side of (271), respectively. Differentiating  $E_0(\rho)$  w.r.t.  $\rho$  and substituting  $\rho = 0$  gives

$$R_{\text{max}} = \frac{\partial E_0(\rho)}{\partial \rho} \Big|_{\rho=0} = I_{P,W}(X; Y|U), \quad (276)$$

where  $I_{P,W}(X; Y|U)$  is the conditional mutual information induced by the channel  $W(y|x)$  and the code distribution  $P(u, x)$ . Next, define

$$F(u, y) = \sum_x P(x|u) \sqrt{W(y|x)}, \quad (277)$$

and

$$G(u, y) = \sum_x P(x|u) \sqrt{W(y|x)} \log W(y|x). \quad (278)$$

After some algebra, we find that

$$R_{\text{crit}} = \left. \frac{\partial E_0(\rho)}{\partial \rho} \right|_{\rho=1} = - \sum_u P(u) \frac{\sum_y \left[ F^2(u, y) \log F(u, y) - \frac{1}{2} F(u, y) G(u, y) \right]}{\sum_{y'} F^2(u, y')}. \quad (279)$$

## 7.2 A Study for $E_{y,2}(R_y, R_z)$

Let

$$E_{y,2}(R_y, R_z) = \max_{\mu \in [0,1]} \max_{\lambda \in [0,\mu]} \left\{ - \log \sum_y \left( \sum_u P(u) \left[ \sum_x P(x|u) W^{\frac{1}{1+\lambda}}(y|x) \right]^{\frac{1+\lambda}{1+\mu}} \right)^{1+\mu} - \lambda R_y - \mu R_z \right\} \quad (280)$$

$$= \max_{\mu \in [0,1]} \max_{s \in [0,1]} \left\{ - \log \sum_y \left( \sum_u P(u) \left[ \sum_x P(x|u) W^{\frac{1}{1+s\mu}}(y|x) \right]^{\frac{1+s\mu}{1+\mu}} \right)^{1+\mu} - s\mu R_y - \mu R_z \right\}, \quad (281)$$

and define

$$E_1(s, \mu) = - \log \sum_y \left\{ \sum_u P(u) \left[ \sum_x P(x|u) W^{\frac{1}{1+s\mu}}(y|x) \right]^{\frac{1+s\mu}{1+\mu}} \right\}^{1+\mu}, \quad (282)$$

such that

$$E_{y,2}(R_y, R_z) = \max_{\mu \in [0,1]} \max_{s \in [0,1]} \left\{ E_1(s, \mu) - s\mu R_y - \mu R_z \right\}. \quad (283)$$

Setting the partial derivatives of the bracketed part of (283) to zero, we get

$$\frac{\partial}{\partial s} E_1(s, \mu) = \mu R_y \quad (284)$$

$$\frac{\partial}{\partial \mu} E_1(s, \mu) = s R_y + R_z, \quad (285)$$

or, equivalently,

$$R_y = \frac{1}{\mu} \cdot \frac{\partial}{\partial s} E_1(s, \mu) \quad (286)$$

$$R_z = \frac{\partial}{\partial \mu} E_1(s, \mu) - \frac{s}{\mu} \cdot \frac{\partial}{\partial s} E_1(s, \mu). \quad (287)$$

Now, if some  $(\mu, s) \in [0, 1]^2$  satisfies (284) and (285), we may relate  $E_{y,2}(R_y, R_z)$ ,  $R_y$  and  $R_z$  parametrically as functions of  $s$  and  $\mu$ . This gives

$$E_{y,2}(R_y, R_z) = E_1(s, \mu) - \mu \cdot \frac{\partial}{\partial \mu} E_1(s, \mu) \quad (288)$$

$$R_y = \frac{1}{\mu} \cdot \frac{\partial}{\partial s} E_1(s, \mu) \quad (289)$$

$$R_z = \frac{\partial}{\partial \mu} E_1(s, \mu) - \frac{s}{\mu} \cdot \frac{\partial}{\partial s} E_1(s, \mu). \quad (290)$$

For explicit expressions for the partial derivatives of  $E_1(s, \mu)$  w.r.t.  $s$  and  $\mu$ , we first define

$$A(y, s, \mu) = \sum_u P(u) \left[ \sum_x P(x|u) W^{\frac{1}{1+s\mu}}(y|x) \right]^{\frac{1+s\mu}{1+\mu}} \quad (291)$$

$$B(u, y, s, \mu) = \sum_x P(x|u) W^{\frac{1}{1+s\mu}}(y|x) \quad (292)$$

$$E(u, y, s, \mu) = \sum_x P(x|u) W^{\frac{1}{1+s\mu}}(y|x) \log W(y|x) \quad (293)$$

$$C(y, s, \mu) = \sum_u P(u) [B(u, y, s, \mu)]^{\frac{1+s\mu}{1+\mu}} \left( \frac{s-1}{1+\mu} \cdot \log B(u, y, s, \mu) - \frac{s}{1+s\mu} \cdot \frac{E(u, y, s, \mu)}{B(u, y, s, \mu)} \right) \quad (294)$$

$$D(y, s, \mu) = \sum_u P(u) [B(u, y, s, \mu)]^{\frac{1+s\mu}{1+\mu}} \left( \log B(u, y, s, \mu) - \frac{1}{1+s\mu} \cdot \frac{E(u, y, s, \mu)}{B(u, y, s, \mu)} \right), \quad (295)$$

and get that the partial derivative w.r.t.  $s$  is given by

$$\frac{\partial}{\partial s} E_1(s, \mu) = -\mu \cdot \frac{\sum_y A^\mu(y, s, \mu) D(y, s, \mu)}{\sum_{y'} A^{1+\mu}(y', s, \mu)}, \quad (296)$$

and the partial derivatives w.r.t.  $\mu$  is given by

$$\frac{\partial}{\partial \mu} E_1(s, \mu) = -\frac{\sum_y A^{1+\mu}(y, s, \mu) \cdot \left[ \log A(y, s, \mu) + \frac{C(y, s, \mu)}{A(y, s, \mu)} \right]}{\sum_{y'} A^{1+\mu}(y', s, \mu)}. \quad (297)$$

Consider the rate pair  $(R_y, R_z)$  for which both (284) and (285) hold with  $s = \mu = 1$ :

$$R_y = \frac{\partial}{\partial s} E_1(s, 1) \Big|_{s=1} \quad (298)$$

$$R_y + R_z = \frac{\partial}{\partial \mu} E_1(1, \mu) \Big|_{\mu=1}, \quad (299)$$

which is the corner point of the affine rate region. As an immediate consequence, due to the monotonicity of  $E_1$ , we get that for low rates, i.e., for rate pairs  $(R_y, R_z)$  that satisfy

$$R_y \leq \left. \frac{\partial}{\partial s} E_1(s, 1) \right|_{s=1}, \quad (300)$$

and

$$R_y + R_z \leq \left. \frac{\partial}{\partial \mu} E_1(1, \mu) \right|_{\mu=1}, \quad (301)$$

the maximizers are  $s^* = \mu^* = 1$ , and  $E_{y,2}(R_y, R_z)$  is given by

$$E_{y,2}(R_y, R_z) = E_1(1, 1) - (R_y + R_z), \quad (302)$$

where

$$E_1(1, 1) = -\log \sum_y \left\{ \sum_x P(x) \sqrt{W(y|x)} \right\}^2. \quad (303)$$

According to (301), we can find the maximal sum-rate in the affine region. Let

$$F(y) = \sum_x P(x) \sqrt{W(y|x)}, \quad (304)$$

and

$$G(y) = \sum_x P(x) \sqrt{W(y|x)} \log W(y|x). \quad (305)$$

After some algebra, we find that the maximal sum-rate is given by

$$R_y + R_z \leq \left. \frac{\partial}{\partial \mu} E_1(1, \mu) \right|_{\mu=1} = -\frac{\sum_y \left[ F^2(y) \log F(y) - \frac{1}{2} F(y) G(y) \right]}{\sum_{y'} F^2(y')}. \quad (306)$$

The error exponent  $E_{y,2}(R_y, R_z)$  depends solely on  $R_z + R_y$  if and only if the maximizing  $s \in [0, 1]$  is  $s^* = 1$ . In this case,

$$\begin{aligned} & E_{y,2}(R_y, R_z) \\ &= \max_{\mu \in [0,1]} \left\{ -\log \sum_y \left\{ \sum_u P(u) \left[ \sum_x P(x|u) W^{\frac{1}{1+\mu}}(y|x) \right]^{\frac{1+\mu}{1+\mu}} \right\}^{1+\mu} - \mu R_y - \mu R_z \right\} \end{aligned} \quad (307)$$

$$= \max_{\mu \in [0,1]} \left\{ -\log \sum_y \left\{ \sum_u P(u) \sum_x P(x|u) W^{\frac{1}{1+\mu}}(y|x) \right\}^{1+\mu} - \mu(R_y + R_z) \right\} \quad (308)$$

$$= \max_{\mu \in [0,1]} \left\{ -\log \sum_y \left\{ \sum_x P(x) W^{\frac{1}{1+\mu}}(y|x) \right\}^{1+\mu} - \mu(R_y + R_z) \right\}, \quad (309)$$

which means that  $E_{y,2}(R_y, R_z) = E_r(R_y + R_z, P_X)$ , i.e., the ordinary random coding error exponent at rate  $R_y + R_z$  for an i.i.d. drawn code with distribution  $P_X$ . Next, we find the rate region for which  $E_{y,2}(R_y, R_z)$  depends solely on  $R_z + R_y$ , but it is not affine in the sum-rate. Consider the rate pairs  $(R_y, R_z)$  for which both (284) and (285) hold with  $s = 1$ :

$$\mu R_y = \left. \frac{\partial}{\partial s} E_1(s, \mu) \right|_{s=1} \quad (310)$$

$$R_y + R_z = \left. \frac{\partial}{\partial \mu} E_1(1, \mu) \right|. \quad (311)$$

Let  $\Gamma_{S1}$  denote the curve given by eqs. (310)-(311):

$$\Gamma_{S1} = \left\{ (R_y, R_z) \left| R_y = \frac{1}{\mu} \cdot \left. \frac{\partial}{\partial s} E_1(s, \mu) \right|_{s=1}, \quad R_y + R_z = \left. \frac{\partial}{\partial \mu} E_1(1, \mu) \right|, \quad 0 < \mu \leq 1 \right\}. \quad (312)$$

Now, in the set of  $(R_y, R_z)$ , with

$$\left. \frac{\partial}{\partial \mu} E_1(1, \mu) \right|_{\mu=1} \leq R_y + R_z \leq \left. \frac{\partial}{\partial \mu} E_1(1, \mu) \right|_{\mu=0}, \quad (313)$$

and being underneath the curve  $\Gamma_{S1}$ , the maximizer is  $s^* = 1$ , and  $E_{y,2}(R_y, R_z)$  is given by (309). Notice that the left-hand-side and the right-hand-side of (313) are expressions for the critical-rate and the maximal-rate for the channel  $W_1(y|x)$ , respectively, and hence, the latter cannot be smaller than the former. Before moving forward, let us obtain a simple information-theoretic expression for the maximal sum-rate. According to the right-hand side of (313), we only have to differentiate  $E_1(1, \mu)$  w.r.t.  $\mu$  and then substitute  $\mu = 0$ . We get

$$R_y + R_z \leq \left. \frac{\partial}{\partial \mu} E_1(1, \mu) \right|_{\mu=0} = I_{P_X, W}(X; Y), \quad (314)$$

where  $I_{P_X, W}(X; Y)$  is the mutual information induced by the channel  $W_1(y|x)$  and the code distribution  $P(x)$ .

Let us now turn to the other extreme, where  $E_{y,2}(R_y, R_z)$  depends solely on  $R_z$ . This happens if and only if the maximizing  $s \in [0, 1]$  is  $s^* = 0$ . In this case,

$$E_{y,2}(R_y, R_z) = \max_{\mu \in [0,1]} \left\{ -\log \sum_y \left\{ \sum_u P(u) \left[ \sum_x P(x|u) W(y|x) \right]^{\frac{1}{1+\mu}} \right\}^{1+\mu} - \mu R_z \right\} \quad (315)$$

$$= \max_{\mu \in [0,1]} \left\{ -\log \sum_y \left\{ \sum_u P(u) V^{\frac{1}{1+\mu}}(y|u) \right\}^{1+\mu} - \mu R_z \right\}, \quad (316)$$

which means that  $E_{y,2}(R_y, R_z) = E_r(R_z, P_U)$ , i.e., the ordinary random coding error exponent at rate  $R_z$  for an i.i.d. code with distribution  $P_U$ , where  $V$  is defined to be the equivalent channel

from  $U$  to  $Y$ . The simple explanation for the fact that  $E_{y,2}(R_y, R_z)$  becomes independent of  $R_y$ , for high  $R_y$ , is that the satellite codewords behave like pure noise.

Next, we find the region where  $E_{y,2}(R_y, R_z)$  depends solely on  $R_z$ . Consider the rate pairs  $(R_y, R_z)$  for which both (284) and (285) hold with  $s = 0$ :

$$\mu R_y = \left. \frac{\partial}{\partial s} E_1(s, \mu) \right|_{s=0} \quad (317)$$

$$R_z = \left. \frac{\partial}{\partial \mu} E_1(0, \mu) \right|_{\mu=1}, \quad (318)$$

Let  $\Gamma_{S0}$  denote the curve given by eqs. (317)-(318):

$$\Gamma_{S0} = \left\{ (R_y, R_z) \left| R_y = \frac{1}{\mu} \cdot \left. \frac{\partial}{\partial s} E_1(s, \mu) \right|_{s=0}, \quad R_z = \left. \frac{\partial}{\partial \mu} E_1(0, \mu) \right|_{\mu=1}, \quad 0 < \mu \leq 1 \right\}. \quad (319)$$

In addition, we have the following corner point for  $\mu = 1$ :

$$(\tilde{R}_y, \tilde{R}_z) = \left( \left. \frac{\partial}{\partial s} E_1(s, 1) \right|_{s=0}, \left. \frac{\partial}{\partial \mu} E_1(0, \mu) \right|_{\mu=1} \right), \quad (320)$$

and we use it to define the straight line connecting that corner point to the  $R_y$ -axis:

$$\tilde{\Gamma}_{S0} = \left\{ (R_y, R_z) \left| R_y = \tilde{R}_y, 0 \leq R_z \leq \tilde{R}_z \right. \right\}, \quad (321)$$

which is the set of all  $(R_y, R_z)$  for which the maximizers are  $s^* = 0$  and  $\mu^* = 1$ .

Let  $\hat{\Gamma}_{S0}$  be defined by  $\Gamma_{S0} \cup \tilde{\Gamma}_{S0}$ . In fact, the curve  $\hat{\Gamma}_{S0}$  is the borderline between the region where  $E_{y,2}(R_y, R_z)$  depends on  $R_y$  (affine or curvy) and the region where  $E_{y,2}(R_y, R_z)$  is independent of  $R_y$ . The set of all  $(R_y, R_z)$  that are above the curve  $\hat{\Gamma}_{S0}$  defines the region where  $E_{y,2}(R_y, R_z)$  is independent of  $R_y$ . In addition, let us obtain a simple informational expression for the maximum of  $R_z$ . According to (319), we only have to differentiate  $E_1(0, \mu)$  w.r.t.  $\mu$  and then substitute  $\mu = 0$ . We get

$$R_z \leq \left. \frac{\partial}{\partial \mu} E_1(0, \mu) \right|_{\mu=0} = I_{P_U, V}(U; Y), \quad (322)$$

where  $I_{P_U, V}(U; Y)$  is the mutual information induced by the channel  $\{V(y|u)\}$  and  $\{P(u)\}$ . In the region

$$R_z \leq \left. \frac{\partial}{\partial \mu} E_1(0, \mu) \right|_{\mu=1}, \quad (323)$$



the maximizer is  $\mu^* = 1$ , and  $E_{y,2}(R_y, R_z)$  is affine in  $R_z$  and is given by

$$E_{y,2}(R_y, R_z) = E_1(0, 1) - R_z, \quad (324)$$

where

$$E_1(0, 1) = -\log \sum_y \left\{ \sum_u P(u) \sqrt{V(y|u)} \right\}^2. \quad (325)$$

The third region is the set of all  $(R_y, R_z)$  for which the maximizing  $s$  is in  $(0, 1)$ . In this case, we use (288)-(290), which hold for every  $s \in (0, 1)$  and  $\mu \in [0, 1]$  such that both (284) and (285) are satisfied. This region can be divided into two complementary regions. In the first one, the maximizer is  $\mu^* = 1$ , and  $E_{y,2}(R_y, R_z)$  is affine in  $R_z$  and curvy in  $R_y$ , while in the second one, the maximizer  $\mu^*$  is in  $(0, 1)$ , and  $E_{y,2}(R_y, R_z)$  is curvy in both  $R_z$  and  $R_y$ . The borderline between those two regions is given by the curve

$$\Gamma_{\mu 1} = \left\{ (R_y, R_z) \left| R_y = \frac{\partial}{\partial s} E_1(s, 1), \quad sR_y + R_z = \frac{\partial}{\partial \mu} E_1(s, \mu) \right|_{\mu=1}, \quad 0 \leq s \leq 1 \right\}. \quad (326)$$

For

$$\begin{aligned} R_z &\geq I_{P_U, V}(U; Y) \\ R_y + R_z &\geq I_{P_X, W}(X; Y), \end{aligned} \quad (327)$$

the maximizers are  $s^* = \mu^* = 0$ , and then  $E_{y,2}(R_y, R_z) = 0$ .

## Appendix

### Proof of Theorem 3

Regarding decoder (95)-(96), let us select

$$f(Q_{UXZ}) = I_Q(U; Z) + I_Q(X; Z|U). \quad (A.1)$$

We show that (A.1) achieves the maximum of  $E_2(s_1(Q_{U_0 X_0 0 Z}), Q_Z)$ , given by Lemma 3, and therefore, the error exponent of this decoder is as large as that of the optimal decoder. First,

the threshold  $s_0(Q_{U_0Z})$  can be easily simplified as

$$s_0(Q_{U_0Z}) = R_y + \max_{\{Q_{X|U_0Z} \in \mathcal{S}(Q_{U_0Z}): I_Q(X;Z|U_0) \leq R_y\}} \left[ f(Q_{U_0XZ}) - I_Q(X;Z|U_0) \right] \quad (\text{A.2})$$

$$= R_y + \max_{\{Q_{X|U_0Z} \in \mathcal{S}(Q_{U_0Z}): I_Q(X;Z|U_0) \leq R_y\}} \left[ I_Q(U_0;Z) + I_Q(X;Z|U_0) - I_Q(X;Z|U_0) \right] \quad (\text{A.3})$$

$$= R_y + I_Q(U_0;Z), \quad (\text{A.4})$$

such that

$$s_1(Q_{U_0X_0Z}) = \max \left\{ R_y + I_Q(U_0;Z), I_Q(U_0;Z) + I_Q(X_{00};Z|U_0) \right\} \quad (\text{A.5})$$

$$= I_Q(U_0;Z) + \max \left\{ R_y, I_Q(X_{00};Z|U_0) \right\}. \quad (\text{A.6})$$

In general, the constraint of the inner minimization problem defining  $E_2(s_1(Q_{U_0X_0Z}), Q_Z)$  is given by

$$f(Q_{UXZ}) + [R_y - I_Q(X;Z|U)]_+ \geq s_1(Q_{U_0X_0Z}), \quad (\text{A.7})$$

which can now be written as

$$\begin{aligned} I_Q(U;Z) + I_Q(X;Z|U) + [R_y - I_Q(X;Z|U)]_+ \\ \geq I_Q(U_0;Z) + \max \left\{ R_y, I_Q(X_{00};Z|U_0) \right\}. \end{aligned} \quad (\text{A.8})$$

Subtracting  $R_y$  from both sides gives

$$\begin{aligned} I_Q(U;Z) + I_Q(X;Z|U) - R_y + [R_y - I_Q(X;Z|U)]_+ \\ \geq I_Q(U_0;Z) + \max \left\{ 0, I_Q(X_{00};Z|U_0) - R_y \right\}, \end{aligned} \quad (\text{A.9})$$

or,

$$I_Q(U;Z) + [I_Q(X;Z|U) - R_y]_+ \geq I_Q(U_0;Z) + [I_Q(X_{00};Z|U_0) - R_y]_+. \quad (\text{A.10})$$

Defining  $D(Q_{UXZ}) \triangleq I_Q(U;Z) + [I_Q(X;Z|U) - R_y]_+$ , we have

$$E_2(s_1(Q_{U_0X_0Z}), Q_Z) = \min_{Q_{UX|Z} \in \mathcal{S}(Q_Z): D(Q_{UXZ}) \geq D(Q_{U_0X_0Z})} D(Q_{UXZ}) \quad (\text{A.11})$$

$$= I_Q(U_0;Z) + [I_Q(X_{00};Z|U_0) - R_y]_+, \quad (\text{A.12})$$

which is the same as on the right hand side of (143).  $\square$

## References

- [1] T. M. Cover, “Broadcast channels,” *IEEE Trans. on Inform. Theory*, vol. 18, no. 1, pp. 2–14, January 1972.
- [2] P. P. Bergmans, “Random coding theorem for broadcast channels with degraded components,” *IEEE Trans. on Inform. Theory*, vol. IT-19, pp. 197-207, March 1973.
- [3] P. P. Bergmans, “A simple converse for broadcast channels with additive white Gaussian noise,” *IEEE Trans. on Inform. Theory*, vol. IT-20, pp. 279-280, March 1974.
- [4] R. G. Gallager, “Capacity and coding for degraded broadcast channels,” *Probl. Pered. Inform.*, vol. 10, no. 3, pp. 3-14, July-September 1974; translated in *Probl. Inform. Transm.*, pp. 185-193, July-September 1974.
- [5] J. Körner and K. Marton, “General broadcast channels with degraded message sets,” *IEEE Trans. on Inform. Theory*, vol. IT-23, pp. 60-64, January 1977.
- [6] I. Csiszár and J. Körner, “Broadcast channels with confidential messages,” *IEEE Trans. on Inform. Theory*, vol. IT-24, pp. 339-348, May 1978.
- [7] A. El Gamal, “The capacity of a class of broadcast channels,” *IEEE Trans. on Inform. Theory*, vol. IT-25, pp. 166-169, March 1979.
- [8] J. Körner and A. Sgarro, “Universally attainable error exponents for broadcast channels with degraded message sets,” *IEEE Trans. on Inform. Theory*, vol. 26, no.6, pp. 670-679, November 1980.
- [9] Y. Kaspi and N. Merhav, “Error exponents for broadcast channels with degraded message sets,” *IEEE Trans. on Inform. Theory*, vol. 57, no. 1, pp. 101–123, January 2011.
- [10] N. Merhav, “Exact random coding error exponents of optimal bin index decoding,” *IEEE Trans. on Inform. Theory*, vol. 60, no. 10, pp. 6024–6031, October 2014.
- [11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Cambridge University Press, 2011.
- [12] R. G. Gallager, *Information Theory and Reliable Communication*, New York, Wiley 1968.

- [13] N. Shulman, *Communication over an Unknown Channel via Common Broadcasting*, Ph.D. dissertation, Department of Electrical Engineering - Systems, Tel Aviv University, July 2003. [http://www.eng.tau.ac.il/~shulman/papers/Nadav\\_PhD.pdf](http://www.eng.tau.ac.il/~shulman/papers/Nadav_PhD.pdf)
- [14] N. Merhav, "Statistical physics and information theory," *Foundations and Trends in Communications and Information Theory*, vol. 6, nos. 12, pp. 1–212, 2009.