IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Error Exponents for Broadcast Channels with Degraded Message Sets

Yonatan Kaspi and Neri Merhav

CCIT Report #734
May 2009

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL

CCIT REPORT #734 May 2009

Error Exponents for Broadcast Channels with Degraded Message Sets

Yonatan Kaspi and Neri Merhav

May 27, 2009

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, ISRAEL
Email: {kaspi@tx, merhav@ee}.technion.ac.il

Abstract

We consider a broadcast channel with a degraded message set, in which a single transmitter sends a common message to two receivers and a private message to one of the receivers only. The main goal of this work is to find new lower bounds to the error exponents of the strong user, the one that should decode both messages, and of the weak user, that should decode only the common message. Unlike previous works, where suboptimal decoders where used, the exponents we derive in this work pertain to optimal decoding and depend on both rates. We take two different approaches.

The first approach is based, in part, on variations of Gallager-type bounding techniques that were presented in a much earlier work on error exponents for erasure/list decoding. The resulting lower bounds are quite simple to understand and to compute.

The second approach is based on a technique that is rooted in statistical physics, and it is exponentially tight from the initial step and onward. This technique is based on analyzing the statistics of certain enumerators. Numerical results show that the bounds obtained by this technique are tighter than those obtained by the first approach and previous results. The derivation, however, is more complex than the first approach and the retrieved exponents are harder to compute.

1 Introduction

In the broadcast channel (BC), as introduced by Cover [1], a single source is communicating to two or more receivers. In this work, we concentrate on the case of two receivers. The

encoder sends a common message, to be decoded by both receivers, and a private message for each decoder. In the case of a degraded message set, one of the private messages is absent. The capacity region of the BC with a degraded message set was found in [2]. A coding theorem for degraded broadcast channels was given by Bergmans [3] and the converse for the degraded channel case was given by Gallager [4]. Bergmans suggested the use of a hierarchical random code: First draw "cloud centers". Next, around each "cloud center", draw a cloud of codewords. The sender sends a specific codeword from one of the clouds. The strong decoder (the one with the better channel) can identify the specific codeword while the weak decoder can only identify the cloud it originated from (see Section II and [3]).

The error exponent is the rate of exponential decay of the average probability of error as a function of the block length. Unlike in the single user regime, where the error exponent is a function of the rate at which the transmitter operates, in the multiuser regime, the error exponent for each user is a function of all rates in the system. We can define an error exponent region, that is, a set of achievable error exponents for fixed rates of both users (see [5]). The tradeoff between the exponents is controlled by the choice of the random coding distributions.

Earlier work on error exponents for general degraded broadcast channels includes [4] and [6]. Both [4] and [6] used the coding scheme of [3], but did not use optimal decoding. In [4], a direct channel from the cloud center to the weak user is defined and the error exponent is calculated for this channel. By defining this channel, the decoder does not use its knowledge of the refined codebook of each cloud. The resulting exponent depends only on one of the rates - the one corresponding to the number of clouds. When the clouds are "full" (high rate of the private message), not much is lost by the use of the defined direct channel. However, for low rates of the private message, the decoding quality can be improved by knowing the codebook. In [6], universally attainable error exponents are given for a suboptimal decoder. Lower and upper bounds to the error exponents, that depend on both rates, are given.

In this work, we derive new lower bounds to the error exponents for both the weak and

the strong decoder of a degraded BC with degraded message sets. The derived exponents pertain to optimum decoding and they depend simultaneously on both rates. We present two approaches to derive the exponents, which start from the same initial step, but are substantially different otherwise.

The first approach is based, in part, on variations of Gallager-type bounding techniques along with refinements that were used in Forney's work on error exponents for erasure/list decoding [7]. Using these techniques, we derive new lower bounds which are quite simple to understand and compute. Both this approach and the approach of [4] use Jensen's inequality, as well as other inequalities, which possibly risk the tightness of the obtained bounds in the exponential scale.

Our second approach avoids the use of these inequalities. Instead, an exponentially tight evaluation of the relevant expressions is derived by assessing the moments of a certain type class enumerators. The underlying ideas behind the second approach are inspired from the statistical mechanical point of view on random code ensembles [8],[9]. The analysis tools we use in this approach are applicable to other problem settings as well, e.g., [10] and [11], where they lead to tighter bounds than those of other methods previously used. The second approach, after its initial step, is guaranteed to be exponentially tight, and is shown to obtain tighter bounds than the first approach and previous results. However, this tightness comes at the price of the complexity of both the derivation and the final results, which makes the task of obtaining numerical results quite involved.

The outline of the remaining part of this work is as follows: Section 2 gives the formal setting and notation. In Section 3 we summarize the main results of this paper, giving the resulting exponents of each of the approaches. in Sections 4 and 5, we derive the exponents using the first and second approach, respectively. At the end of each of the sections, we give numerical results for the degraded binary symmetric channel (BSC). We conclude our work in section VI.

2 Preliminaries

We begin with notation conventions. Capital letters represent scalar random variables (RVs) and specific realizations of them are denoted by the corresponding lower case letters. Random vectors of dimension n will be denoted by bold-face letters. Indicator functions of events will be denoted by $\mathcal{I}(\cdot)$. We write $[x]^+$ for the positive part of a real number x, i.e $[x]^+ \stackrel{\triangle}{=} \max(x, 0)$. The expectation operator will be denoted by $\boldsymbol{E}\{\cdot\}$. When we wish to emphasize the dependence of the expectation on a certain underlying probability distribution, say, Q, we subscript it by Q. i.e. $\boldsymbol{E}_Q\{\cdot\}$. We consider a memoryless broadcast channel with a finite input alphabet $\mathcal{X}$ and finite output alphabets $\mathcal{Y}$ and $\mathcal{Z}$, of the strong decoder and the weak decoder, respectively, given by $P(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{x}) = \prod_{t=1}^n P(y_t, z_t|x_t), \quad (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$. We are interested in sending one of $M_{yz} = e^{nR_{yz}}$ messages to both receivers and one of $M_y = e^{nR_y}$ to the strong receiver, that observes $\boldsymbol{y}$.

Consider a random selection of a hierarchical code [3] as follows: First, $M_{yz} = e^{nR_{yz}}$ "cloud centers" $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{M_{yz}} \in \mathcal{U}^n$ are drawn independently, each one using a distribution $P(\boldsymbol{u}) = \prod_{t=1}^n P(u_t)$, where $u \in \mathcal{U}$ is an auxiliary random variable. Then, for each $m = 1, 2, \ldots, M_{yz}$, $M_y = e^{nR_y}$ codewords $\boldsymbol{x}_{m,1}, \ldots, \boldsymbol{x}_{m,M_y} \in \mathcal{X}^n$ are drawn according to $P(\boldsymbol{x}|\boldsymbol{u}) = \prod_{t=1}^n P(x_t|u_t)$, with $\boldsymbol{u} = \boldsymbol{u}_m$.

The strong decoder is interested in decoding both indices (m, i) of the transmitted codeword $\boldsymbol{x}_{m,i}$, whereas the weak decoder, the one that observes $\boldsymbol{z}$, is only interested in decoding the index m. Thus, while the strong decoder best applies full maximum likelihood (ML) decoding, $(\hat{m}(\boldsymbol{y}), \hat{i}(\boldsymbol{y})) = \arg\max_{m,i} P_1(\boldsymbol{y}|\boldsymbol{x}_{m,i})$, the best decoding rule for the weak decoder is given by $\tilde{m}(\boldsymbol{z}) = \arg\max_m \frac{1}{M_y} \sum_{i=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{x}_{m,i})$, where $P_3(\boldsymbol{z}|\boldsymbol{x}) = \prod_{t=1}^n P_3(z_t|x_t) = \prod_{t=1}^n \sum_y P(y, z_t|x_t)$.

The capacity region for a BC with degraded message sets is given [2] by the closure of:

$$\{R_{yz}, R_y : \quad R_{yz} \le I(U; Z), R_y \le I(X; Y|U), R_{yz} + R_y \le I(X; Y)\}$$

for some $P(u, x, y, z) = P(u)P(x|u)P(y, z|x)$ and $|\mathcal{U}| \le |\mathcal{X}| + 2$. If the channel is degraded,

since we have $U \leftrightarrow X \leftrightarrow Y \leftrightarrow Z$, the restriction on the sum of rates is trivially satisfied and can be omitted. The capacity region for the general BC is still an open problem. The best inner bound for it is given by Marton [12] and, in a simpler manner, by El Gamal and Meulen [13]:

$$\{R_{yz}, R_y : \quad R_{yz} \leq I(U;Z), R_y \leq I(V,Y), R_{yz} + R_y \leq I(U;Z) + I(V;Y) - I(U;V)\}$$

for some $p(x,u,v)$, where u, v are auxiliary random variables with finite ranges.

Denote the average error probability of the strong decoder by $\overline{P_E^y} = Pr\left\{(\hat{m}(\boldsymbol{y}), \hat{i}(\boldsymbol{y})) \neq (m,i)\right\}$ and the average error probability of the weak decoder by $\overline{P_E^z} = Pr\left\{\tilde{m}(\boldsymbol{z}) \neq m\right\}$. The exponents of the strong and weak decoders will be denoted by E_y and E_z, respectively. A pair (E_y, E_z) is said to be an *attainable pair in the random coding sense*, for a given (R_y, R_{yz}), if there exist random coding distributions $\{P(u)\}$ and $\{P(x|u)\}$ such that the random coding exponents satisfy $E_y \leq \liminf_{n\to\infty} -\frac{1}{n}\log \overline{P_E^y}$ and $E_z \leq \liminf_{n\to\infty} -\frac{1}{n}\log \overline{P_E^z}$, where all logarithms throughout the sequel are taken to the natural base. For a given pair (R_y, R_{yz}), we say that E_z is *an attainable exponent for the weak user* if there there exists $E_y > 0$ such that the pair (E_y, E_z) is attainable in the random coding sense.

3 Main Results

In this section, we outline the main results of this paper. As described in the Introduction, we use two different approaches to derive the error exponents of a general degraded broadcast channel, pertaining to optimal decoding. We introduce the resulting exponents of each of these approaches in the following two subsections.

3.1 Gallager-type bound

Denoting $f(a,b,z) = \sum_u P(u) \left[\sum_x P(x|u) P_3(z|x)^{a/b}\right]^b$, we define:

$$E_0(\rho,\lambda,\alpha,\mu) = -\log\left[\sum_z f(1-\rho\lambda,\alpha,z)\cdot f(\lambda,\mu,z)\right],$$

$$E_y^1(R_y,\rho) = -\rho R_y - \log\sum_y\sum_u P(u)\left[\sum_x P(x|u)P_1(y|x)^{\frac{1}{1+\rho}}\right]^{1+\rho},$$

$$E_y^2(R_y,R_{yz},\rho) = -\rho(R_y+R_{yz}) - \log\left\{\sum_y\left[\sum_x P(x)P_1(y|x)^{\frac{1}{1+\rho}}\right]^{1+\rho}\right\} \tag{1}$$

Let

$$E_{z,1}(R_{yz},R_y) = \max_{0\le\rho\le1,0\le\lambda\le\mu\le1,1-\rho\lambda\le\alpha\le1}\{E_0(\rho,\lambda,\alpha,\mu) - (\alpha+\rho\mu-1)R_y - \rho R_{yz}\}$$

$$E_{y,1}(R_{yz},R_y) = \min\left(\max_{0<\rho<1} E_y^1(R_y,\rho), \max_{0<\rho<1} E_y^2(R_y,R_{yz},\rho)\right) \tag{2}$$

The first main result of this paper is the following theorem.

Theorem 1: For the degraded broadcast channel defined in Section II, the pair $(E_{z,1}(R_{yz},R_y), E_{y,1}(R_{yz},R_y))$, as defined in eq. (2), is an attainable pair in the random coding sense.

We prove this theorem in Section 4. Unlike in earlier papers [4], [6], [5], the exponents of Theorem 1 pertain to *optimal* decoding and depend on both rates. For the weak decoder exponent, the optimization on all parameters, although possible, is hard computationally. We therefore examine a few interesting choices of the parameters, in order to reduce the dimensionality of the optimization process.

1. Let $\alpha = \mu$. In this case, we show in Appendix A.1 that $\forall\lambda: \; E_0(\rho,\frac{1}{1+\rho},\alpha,\alpha) \ge E_0(\rho,\lambda,\alpha,\alpha)$, thus, the choice of $\lambda = \frac{1}{1+\rho}$ is optimal. Applying $\alpha=\mu, \lambda=\frac{1}{1+\rho}$ our bound becomes:

$$E(R_y,R_{yz}) = \max_{0\le\rho\le1,\frac{1}{1+\rho}\le\alpha\le1} E_0\left(\rho,\frac{1}{1+\rho},\alpha,\alpha\right) - [\alpha(1+\rho)-1]R_y - \rho R_{yz}. \tag{3}$$

This is a somewhat more compact expression with only two parameters. Numerical results indicate that, at least for the BSC we tested, the choice $\alpha = \mu$ is the optimal choice. However, we do not have a proof that this is true in general.

2. As a further restriction of item no. 1 above, consider the choice $\alpha = \mu = \frac{1}{1+\rho}$. In this case, the expressions in the inner–most brackets of (17) and (18) become $\sum_x Q(x|u)P_3(z|x) \stackrel{\triangle}{=} P_4(z|u)$, and $\alpha + \rho\mu - 1 = 0$. Thus, we get an exponent given by

$$E_0\left(\rho, \frac{1}{1+\rho}, \frac{1}{1+\rho}, \frac{1}{1+\rho}\right) - \rho R_{yz} = \\ -\log\left\{\sum_z \left[\sum_u P(u)P_4(z|u)^{1/(1+\rho)}\right]^{1+\rho}\right\} - \rho R_{yz} \tag{4}$$

which is exactly the ordinary Gallager function for the channel $P(z|u)$, obtained by sub–optimal decoding at the weak user [4], ignoring the knowledge of the refined codebook of each cloud center. This means that the exponents of Theorem 1 are at least as tight as the result of [4]. Numerical results show that, at least for the degraded BSC case, the exponents of Theorem 1 are tighter.

3. Another further restriction of item no. 1 is the choice $\alpha = \mu = 1$, which gives:

$$E_0\left(\rho, \frac{1}{1+\rho}, 1, 1\right) - \rho(R_y + R_{yz}) = -\rho(R_y + R_{yz}) \\ -\log\left\{\sum_z \left[\sum_x Q(x)P_3(z|x)^{1/(1+\rho)}\right]^{1+\rho}\right\}. \tag{5}$$

This corresponds to i.i.d. random coding according to $Q(x) \stackrel{\triangle}{=} \sum_u Q(u)Q(x|u)$ at rate $R_y + R_{yz}$.

3.2 A bound based on Type class enumerators

Let (X, U, Y, Z) be a quadruplet of random variables, taking values in $\mathcal{X} \times \mathcal{U} \times \mathcal{Y} \times \mathcal{Z}$, and being governed by a generic joint distribution $Q_{XUYZ} = \{Q_{XUYZ}(x, u, y, z),\ x \in \mathcal{X},\ u \in \mathcal{U},\ y \in \mathcal{Y}\ z \in \mathcal{Z}\}$, where, as introduced in Section 2, $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are, respectively, the channel

input and output alphabets and $\mathcal{U}$ is the alphabet of the auxiliary random variable which is of finite cardinality. Let us denote the various marginals and conditional distributions derived from Q_{XUYZ}, using the standard conventions, e.g., Q_X is the marginal distribution of X, $Q_{U|Z}$ is the conditional distribution of U given Z, etc. Expectation w.r.t. Q_{XUYZ}, or Q for short, will be denoted by $\boldsymbol{E}_Q$. Similarly, information measures, like entropy and conditional entropy induced by Q, will be subscripted by Q, e.g., $H_Q(X|U,Z)$ is the conditional entropy of X given U and Z under $Q = Q_{XUZY}$. In the following description, we allow various joint distributions $\{Q\}$ to govern (X,U,Y,Z).

Let Q_Y, Q_Z be given. We define $\mathcal{G}(R_y, Q_{U|Z})$ to be the set of conditional distributions $\{Q_{X|U,Z}\}$ that satisfy $R_y + \boldsymbol{E}_Q \log P(X|U) + H_Q(X|U,Z) > 0$, where, as described in Section 2, $P(x|u)$ is the random coding distribution according to which the codewords $\{\boldsymbol{x}_{m,i}\}$ are drawn given $\boldsymbol{u}_m$. Similarly, let $\mathcal{G}(R_y, Q_{U|Y})$ be the set of conditional distributions $\{Q_{X|U,Y}\}$ that satisfy $R_y + \boldsymbol{E}_Q \log P(X|U) + H_Q(X|U,Y) > 0$. Next define,

$$\begin{aligned}\alpha(Q_{U|Z}) &\stackrel{\triangle}{=} (1-\rho\lambda) \max_{Q_{X|UZ}\in\mathcal{G}(R_y,Q_{U|Z})} [\boldsymbol{E}_Q \log P(X|U) + \\ &\qquad H_Q(X|U,Z) + \boldsymbol{E}_Q \log P_3(Z|X)] \qquad (6)\\ \beta(Q_{U|Z}) &\stackrel{\triangle}{=} \rho\lambda R_y + \max_{Q_{X|UZ}\in\mathcal{G}^c(R_y,Q_{U|Z})} [\boldsymbol{E}_Q \log P(X|U) + \\ &\qquad H_Q(X|U,Z) + (1-\rho\lambda)\boldsymbol{E}_Q \log P_3(Z|X)]\,, \\ E_{\alpha\beta}(Q_{U|Z}) &= \max\{\alpha(Q_{U|Z}), \beta(Q_{U|Z})\}. \qquad (7)\end{aligned}$$

where, as described in Section 2, $P_3(\cdot|\cdot)$ is the overall channel to the weak user. Similarly, define:

$$\begin{aligned}\gamma(Q_{U|Y}) &\stackrel{\triangle}{=} \rho\left(R_y + \max_{Q_{X|U,Y}\in\mathcal{G}(R_y,Q_{U|Y})} [\boldsymbol{E}_Q \log P(X|U) + \right. \\ &\qquad \left. H_Q(X|Y,U) + \lambda \boldsymbol{E}_Q \log P_1(Y|X)]\right) \qquad (8)\\ \zeta(Q_{U|Y}) &\stackrel{\triangle}{=} R_y + \max_{Q_{X|U,Y}\in\mathcal{G}^c(R_y,Q_{U|Y})} [\boldsymbol{E}_Q \log P(X|U) + \\ &\qquad H_Q(X|U,Y) + (\rho\lambda)\boldsymbol{E}_Q \log P(Y|X)] \qquad (9)\\ E_{\gamma\zeta}(Q_{U|Z}) &= \max\{\gamma(Q_{U|Z}), \zeta(Q_{U|Z})\}. \qquad (10)\end{aligned}$$

Also, define

$$\bar{m}(Q_{U|Z}) \stackrel{\triangle}{=} R_{yz} + H_Q(U|Z) + \boldsymbol{E}_Q \log P(U)$$

where, as said, $\{P(u)\}$ is the random coding distribution of the cloud centers $\{\boldsymbol{u}_m\}$. Now,

$$N(Q_{X|Z}, Q_{U|Z}, R_y) \stackrel{\triangle}{=} R_y + \max_{Q_{X|UZ}} [\boldsymbol{E}_Q \log P(X|U) + H_Q(X|U,Z)], \tag{11}$$

where the maximization is over all $\{Q_{X|UZ}\}$ that are consistent with $Q_{X|Z}$. Next, we define

$$\begin{aligned} &\mathcal{G}_z(R_{yz}) \stackrel{\triangle}{=} \{Q_{U|Z}: \ R_{yz} + H_Q(U|Z) + \boldsymbol{E} \log P(U) \geq 0\}, \\ &B(Q_{X|Z}, Q_{U|Z}, R_y) = \rho N(Q_{X|Z}, Q_{U|Z}, R_y) \cdot \lambda^{\mathcal{I}\{N(Q_{X|Z}, Q_{U|Z}, R_y) > 0\}} \end{aligned} \tag{12}$$

and

$$C(Q_{X|Z}, Q_{U|Z}, R_y) = N(Q_{X|Z}, Q_{U|Z}, R_y) \cdot (\rho\lambda)^{\mathcal{I}\{N(Q_{X|Z}, Q_{U|Z}, R_y) > 0\}}, \tag{13}$$

We similarly define $\mathcal{G}_y(R_{yz})$, $N(Q_{X|Y}, Q_{U|Y}, R_y)$ and $\bar{m}(Q_{U|Y})$ by replacing the respective role of Z by Y. Next define

$$D(Q_{X|Y}, Q_{U|Y}, R_y) = N(Q_{X|Y}, Q_{U|Y}, R_y) \cdot \rho^{\mathcal{I}\{N(Q_{X|Y}, Q_{U|Y}, R_y) > 0\}}, \tag{14}$$

We also define

$$\begin{aligned} E(Q_{X|Z}) \stackrel{\triangle}{=} \max \Bigg\{ &\max_{Q_{U|Z} \in \mathcal{G}_z(R_{yz})} [B(Q_{X|Z}, Q_{U|Z}, R_y) + \\ &\rho\bar{m}(Q_{U|Z})], \max_{Q_{U|Z} \in \mathcal{G}_z^c(R_{yz})} [C(Q_{X|Z}, Q_{U|Z}, R_y) + \bar{m}(Q_{U|Z})] \Bigg\}, \\ E(Q_{X|Y}) \stackrel{\triangle}{=} \max \Bigg\{ &\rho \max_{Q_{U|Y} \in \mathcal{G}_y(R_{yz})} [N(Q_{X|Y}, Q_{U|Y}, R_y) + \bar{m}(Q_{U|Y})], \\ &\max_{Q_{U|Y} \in \mathcal{G}_y^c(R_{yz})} [D(Q_{X|Y}, Q_{U|Y}, R_y) + \bar{m}(Q_{U|Y})] \Bigg\}, \end{aligned}$$

$$E_1(Q_Z, R_y, R_{yz}, \rho, \lambda) \stackrel{\triangle}{=} \min_{Q_{U|Z}} \left[\boldsymbol{E}_Q \log \frac{1}{P(U)} - H_Q(U|Z) - E_{\alpha\beta}(Q_{U|Z}) \right],$$

$$E_2(Q_Z, R_y, R_{yz}, \rho, \lambda) \stackrel{\triangle}{=} \min_{Q_{X|Z}} \left[\rho\lambda \log \frac{1}{P_3(Z|X)} - E(Q_{X|Z}) + \rho\lambda R_y \right],$$

$$E_3(Q_Y, \rho, \lambda) \stackrel{\triangle}{=} \min_{Q_{X,U|Y}} \left[\boldsymbol{E}_Q \log \frac{1}{P(U,X)} - H_Q(X,U|Y) + (1-\rho\lambda) \boldsymbol{E}_Q \log \frac{1}{P(Y|X)} \right]$$

$$E_4(Q_Y, R_y, R_{yz}, \rho, \lambda) \stackrel{\triangle}{=} \min_{Q_{U|Y}} \left[\boldsymbol{E}_Q \log \frac{1}{P(U)} - E_{\gamma\zeta}(Q_{U|Y}) - H(U|Y) \right]$$

$$E_5(Q_Y, R_y, R_{yz}, \rho, \lambda) \stackrel{\triangle}{=} \min_{Q_{X|Y}} \left[\lambda\rho \hat{\mathbf{E}}_{\boldsymbol{yx}} \log \frac{1}{P_1(Y|X)} - E(Q_{X|Y}) \right]$$

Finally,

$$\begin{aligned} E_{z,2}(R_{yz}, R_y) = \max_{\rho \geq 0} \max_{0 \leq \lambda \leq 1/\rho} \min_{Q_Z} [& E_1(Q_Z, R_y, R_{yz}, \rho, \lambda) + \\ & E_2(Q_Z, R_y, R_{yz}, \rho, \lambda) - H_Q(Z)]. \end{aligned}$$

$$\begin{aligned} E_{y,2}(R_{yz}, R_y) = \max_{\rho \geq 0} \max_{\lambda \geq 0} \min_{Q_Y} [& E_3(Q_Y, \rho, \lambda) + \max\{E_4(Q_Y, R_y, R_{yz}, \rho, \lambda), \\ & E_5(Q_Y, R_y, R_{yz}, \rho, \lambda)\} - H_Q(Y)]. \end{aligned} \tag{15}$$

The second main result of this paper is given in the following theorem:

Theorem 2: For the degraded broadcast channel defined in Section II, the pair $(E_{z,2}(R_{yz}, R_y), E_{y,2}(R_{yz}, R_y))$, as defined in eq. (15), is an attainable pair in the random coding sense.

These exponents also pertain to *optimal* decoding and they depend on both rates. Unlike the exponent of Theorem 1, where the weak decoder exponent had four free parameters, here, $E_{z,2}$ has only two free parameters (λ, ρ). Moreover, $(E_{z,2}(R_{yz}, R_y), E_y(R_{yz}, R_y))$ are at least as tight as the exponents of the previous section since, as we will see in the following, their derivation is exponentially tight after the same initial step we take in the proof of Theorem 1. Numerical results show that $E_{z,2}$ is tighter, at least for the binary symmetric case.

4 Derivation of the Gallager Type Bound

In this section we prove Theorem 1.

4.1 The Weak Decoder

Applying Gallager's general upper bound [14, p. 65] to the "channel" $P(\boldsymbol{z}|m) = \frac{1}{M_y}\sum_{i=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{x}_{m,i})$, we have for $\lambda \geq 0, \rho \geq 0$:

$$P_{E_m}^z \leq \sum_{\boldsymbol{z}} \left[\frac{1}{M_y}\sum_{i=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{x}_{m,i})\right]^{1-\rho\lambda} \times \left[\sum_{m'\neq m}\left(\frac{1}{M_y}\sum_{j=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{x}_{m',j})\right)^{\lambda}\right]^{\rho}.$$

Thus, the average error probability w.r.t. the ensemble of codes is upper bounded in terms of the expectations of each of the bracketed terms above (since messages from different clouds are independent). Define:

$$A \stackrel{\triangle}{=} \boldsymbol{E}\left\{\left[\frac{1}{M_y}\sum_{i=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{X}_{m,i})\right]^{1-\rho\lambda}\right\}$$
$$B \stackrel{\triangle}{=} \boldsymbol{E}\left\{\left[\sum_{m'\neq m}\left(\frac{1}{M_y}\sum_{j=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{X}_{m',j})\right)^{\lambda}\right]^{\rho}\right\}$$

As for A, we have

$$\begin{aligned}
A &= \boldsymbol{E}\left\{\left[\frac{1}{M_y}\sum_{i=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{X}_{m,i})\right]^{1-\rho\lambda}\right\} \\
&= M_y^{\rho\lambda-1}\cdot\boldsymbol{E}\left\{\left[\sum_{i=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{X}_{m,i})\right]^{1-\rho\lambda}\right\} \\
&= M_y^{\rho\lambda-1}\cdot\sum_{\boldsymbol{u}} P(\boldsymbol{u})\cdot\boldsymbol{E}\left\{\left[\left(\sum_{j=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{X}_{m,i})\right)^{(1-\rho\lambda)/\alpha}\right]^{\alpha}|\boldsymbol{u}\right\} \\
&\leq M_y^{\rho\lambda-1}\cdot\sum_{\boldsymbol{u}} P(\boldsymbol{u})\cdot\boldsymbol{E}\left\{\left[\sum_{j=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{X}_{m,i})^{(1-\rho\lambda)/\alpha}\right]^{\alpha}|\boldsymbol{u}\right\} \qquad \alpha \geq 1-\rho\lambda \\
&\leq M_y^{\alpha+\rho\lambda-1}\cdot\sum_{\boldsymbol{u}} P(\boldsymbol{u})\cdot\left[\sum_{\boldsymbol{x}} P(\boldsymbol{x}|\boldsymbol{u})P_3(\boldsymbol{z}|\boldsymbol{x})^{(1-\rho\lambda)/\alpha}\right]^{\alpha} \qquad \alpha \leq 1 \qquad (16)
\end{aligned}$$

For a memoryless channel and $Q(\boldsymbol{u}), Q(\boldsymbol{x}|\boldsymbol{u})$ as defined in Section 2, we have

$$
\begin{aligned}
&= M_y^{\alpha+\rho\lambda-1} \cdot \sum_{\boldsymbol{u}} P(\boldsymbol{u}) \cdot \left[\sum_{\boldsymbol{x}} \prod_{t=1}^{n} P(x_t|u_t) P_3(z_t|x_t)^{(1-\rho\lambda)/\alpha} \right]^{\alpha} \\
&= M_y^{\alpha+\rho\lambda-1} \cdot \sum_{\boldsymbol{u}} P(\boldsymbol{u}) \cdot \left[\prod_{t=1}^{n} \sum_{x} P(x|u_t) P_3(z_t|x)^{(1-\rho\lambda)/\alpha} \right]^{\alpha} \\
&= M_y^{\alpha+\rho\lambda-1} \cdot \sum_{\boldsymbol{u}} P(\boldsymbol{u}) \cdot \prod_{t=1}^{n} \left[\sum_{x} P(x|u_t) P_3(z_t|x)^{(1-\rho\lambda)/\alpha} \right]^{\alpha} \\
&= M_y^{\alpha+\rho\lambda-1} \cdot \prod_{t=1}^{n} \left(\sum_{u} P(u) \left[\sum_{x} P(x|u) P_3(z_t|x)^{(1-\rho\lambda)/\alpha} \right]^{\alpha} \right). \qquad (17)
\end{aligned}
$$

Regarding B, we similarly obtain:

$$
\begin{aligned}
B &= \boldsymbol{E} \left\{ \left[\sum_{m' \neq m} \left(\frac{1}{M_y} \sum_{j=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{X}_{m',j}) \right)^{\lambda} \right]^{\rho} \right\} \\
&= M_y^{-\rho\lambda} \cdot \boldsymbol{E} \left\{ \left[\sum_{m' \neq m} \left(\sum_{j=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{X}_{m',j}) \right)^{\lambda} \right]^{\rho} \right\} \\
&\leq M_y^{-\rho\lambda} \cdot \left[\boldsymbol{E} \left\{ \sum_{m' \neq m} \left(\sum_{j=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{X}_{m',j}) \right)^{\lambda} \right\} \right]^{\rho} \qquad 0 \leq \rho \leq 1 \\
&\leq M_y^{-\rho\lambda} M_{yz}^{\rho} \cdot \left[\boldsymbol{E} \left\{ \left(\sum_{j=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{X}_{m',j}) \right)^{\lambda} \right\} \right]^{\rho} \\
&= M_y^{-\rho\lambda} M_{yz}^{\rho} \cdot \left[\boldsymbol{E} \left\{ \left(\left[\sum_{j=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{X}_{m',j}) \right]^{\lambda/\mu} \right)^{\mu} \right\} \right]^{\rho} \\
&\leq M_y^{-\rho\lambda} M_{yz}^{\rho} \cdot \left[\boldsymbol{E} \left\{ \left(\sum_{j=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{X}_{m',j})^{\lambda/\mu} \right)^{\mu} \right\} \right]^{\rho} \qquad \mu \geq \lambda \\
&\leq M_y^{(\mu-\lambda)\rho} M_{yz}^{\rho} \cdot \left[\sum_{\boldsymbol{u}'} P(\boldsymbol{u}') \left(\sum_{\boldsymbol{x}'} P(\boldsymbol{x}'|\boldsymbol{u}') P_3(\boldsymbol{z}|\boldsymbol{x}')^{\lambda/\mu} \right)^{\mu} \right]^{\rho} \qquad \mu \leq 1 \\
&= M_y^{(\mu-\lambda)\rho} M_{yz}^{\rho} \cdot \prod_{t=1}^{n} \left[\sum_{u'} P(u') \left(\sum_{x'} P(x'|u') P_3(z_t|x')^{\lambda/\mu} \right)^{\mu} \right]^{\rho}. \qquad (18)
\end{aligned}
$$

Denoting $f(a,b,z)=\sum_u Q(u)\left[\sum_x Q(x|u)P_3(z|x)^{a/b}\right]^b$, we obtain:

$$\begin{aligned}\overline{P_E^z} &\le M_y^{\alpha+\rho\mu-1}M_{yz}^{\rho}\times\left\{\sum_z f(1-\rho\lambda,\alpha,z)\cdot f^{\rho}(\lambda,\mu,z)\right\}^n \\ &= e^{-n[E_0(\rho,\lambda,\alpha,\mu)-(\alpha+\rho\mu-1)R_y-\rho R_{yz}]}\end{aligned} \tag{19}$$

where

$$\begin{aligned}&E_0(\rho,\lambda,\alpha,\mu)= \\ &\quad -\log\left[\sum_z f(1-\rho\lambda,\alpha,z)\cdot f(\lambda,\mu,z)\right].\end{aligned} \tag{20}$$

After optimizing over all free parameters, we get $\overline{P_E^z}\le\exp\{-nE(R_y,R_{yz})\}$, where

$$E(R_y,R_{yz})=\max_{0\le\rho\le1,0\le\lambda\le\mu\le1,1-\rho\lambda\le\alpha\le1}\{E_0(\rho,\lambda,\alpha,\mu)-(\alpha+\rho\mu-1)R_y-\rho R_{yz}\} \tag{21}$$

which is the weak decoder exponent of Theorem 1.

4.2 The Strong Decoder

The strong decoder (Y decoder) has to decode correctly both indices (m,i) of the transmitted $\boldsymbol{x}_{m,i}$. Applying Gallager's bound [14, p. 65], and assuming, without loss of generality, that $(m,i)=(1,1)$ was sent, we have for $\lambda\ge0,\rho\ge0$:

$$\begin{aligned}P_{E_{1,1}}^y &\le \sum_{\boldsymbol{y}} P_1(\boldsymbol{y}|\boldsymbol{x}_{1,1})\left(\sum_{(m,i)\ne(1,1)}\frac{P_1(\boldsymbol{y}|\boldsymbol{x}_{m,i})^{\lambda}}{P_1(\boldsymbol{y}|\boldsymbol{x}_{1,1})^{\lambda}}\right)^{\rho} \\ &= \sum_{\boldsymbol{y}} P_1(\boldsymbol{y}|\boldsymbol{x}_{1,1})^{1-\lambda\rho}\left(\sum_{i=2}^{M_y}P_1(\boldsymbol{y}|\boldsymbol{x}_{1,i})^{\lambda}+\sum_{m=2}^{M_{yz}}\sum_{i=1}^{M_y}P_1(\boldsymbol{y}|\boldsymbol{x}_{m,i})^{\lambda}\right)^{\rho} \\ &\overset{\rho\le1}{\le} \sum_{\boldsymbol{y}} P_1(\boldsymbol{y}|\boldsymbol{x}_{1,1})^{1-\lambda\rho}\left[\left(\sum_{i=2}^{M_y}P_1(\boldsymbol{y}|\boldsymbol{x}_{1,i})^{\lambda}\right)^{\rho}+\left(\sum_{m=2}^{M_{yz}}\sum_{i=1}^{M_y}P_1(\boldsymbol{y}|\boldsymbol{x}_{m,i})^{\lambda}\right)^{\rho}\right] \\ &\triangleq P_{E_{y1}}+P_{E_{y2}}\end{aligned} \tag{22}$$

The two resulting expressions deal, respectively, with two separate error events:

1. The Y decoder chose a different private message from the correct cloud.

2. The Y decoder chose a message from a wrong cloud.

The first expression was treated in [4]. We have: $\overline{P_{E_{y1}}} \leq 2^{-nE_{y1}(R_y,\rho)}$, where,

$$\begin{aligned} E_{y1}(R_y,\rho) &= -\rho R_y \\ &\quad - \log \sum_y \sum_u Q(u) \left[\sum_x Q(x|u) P_1(y|x)^{\frac{1}{1+\rho}} \right]^{1+\rho} \end{aligned} \tag{23}$$

We now turn to the second term in (22).

$$P_{E_{y2}} = \sum_{\boldsymbol{y}} P_1(\boldsymbol{y}|\boldsymbol{x}_{1,1})^{1-\lambda\rho} \left[\sum_{m=2}^{M_{yz}} \sum_{i=1}^{M_y} P_1(\boldsymbol{y}|\boldsymbol{x}_{i,m})^{\lambda} \right]^{\rho} \tag{24}$$

Here, when averaging over the ensemble, since the term in brackets of (24) originates from a different cloud, it is independent of the first term. Thus,

$$\begin{aligned} \overline{P_{E_{y2}}} &= \sum_{\boldsymbol{y}} \boldsymbol{E} \left[P_1(\boldsymbol{y}|\boldsymbol{X}_{1,1})^{1-\lambda\rho} \right] \boldsymbol{E} \left[\sum_{m=2}^{M_{yz}} \sum_{i=1}^{M_y} P_1(\boldsymbol{y}|\boldsymbol{X}_{m,i})^{\lambda} \right]^{\rho} \\ &\leq \sum_{\boldsymbol{y}} \boldsymbol{E} \left[P_1(\boldsymbol{y}|\boldsymbol{X}_{1,1})^{1-\lambda\rho} \right] \left[\boldsymbol{E} \sum_{m=2}^{M_{yz}} \sum_{i=1}^{M_y} P_1(\boldsymbol{y}|\boldsymbol{X}_{m,i})^{\lambda} \right]^{\rho} \qquad \rho \leq 1 \\ &\leq \sum_{\boldsymbol{y}} \left[\sum_{\boldsymbol{x}} P(\boldsymbol{x}) P_1(\boldsymbol{y}|\boldsymbol{x})^{1-\lambda\rho} \right] \left[\sum_{m=2}^{M_{yz}} \sum_{i=1}^{M_y} \sum_{\boldsymbol{x}} Q(\boldsymbol{x}) P_1(\boldsymbol{y}|\boldsymbol{x})^{\lambda} \right]^{\rho} \\ &\leq M_y^{\rho} M_{yz}^{\rho} \sum_{\boldsymbol{y}} \left[\sum_{\boldsymbol{x}} P(\boldsymbol{x}) P_1(\boldsymbol{y}|\boldsymbol{x})^{1-\lambda\rho} \right] \left[\sum_{\boldsymbol{x}} Q(\boldsymbol{x}) P_1(\boldsymbol{y}|\boldsymbol{x})^{\lambda} \right]^{\rho} \end{aligned} \tag{25}$$

Selecting [1] $\lambda = \frac{1}{1+\rho}$ yields

$$\overline{P_{E_{y2}}} \leq M_y^{\rho} M_{yz}^{\rho} \sum_{\boldsymbol{y}} \left[\sum_{\boldsymbol{x}} P(\boldsymbol{x}) P_1(\boldsymbol{y}|\boldsymbol{x})^{\frac{1}{1+\rho}} \right]^{1+\rho}$$

For a memoryless channel, we get:

$$\begin{aligned} \overline{P_{E_{y2}}} &\leq M_y^{\rho} M_{yz}^{\rho} \left\{ \sum_y \left[\sum_x P(x) P_1(y|x)^{\frac{1}{1+\rho}} \right]^{1+\rho} \right\}^n \\ &= 2^{-nE_{y2}(R_y,R_{yz},\rho)} \end{aligned} \tag{26}$$

[1] This choice is optimal for the same reason it is optimal in the single user regime. see [15] Prob. 5.6

where

$$E_{y2}(R_y, R_{yz}, \rho) = -\rho(R_y + R_{yz}) - \log\left\{\sum_y \left[\sum_x P(x) P_1(y|x)^{\frac{1}{1+\rho}}\right]^{1+\rho}\right\}$$

Note that this corresponds to the random coding exponent for the channel $X \to Y$ at rate $R_y + R_{yz}$.

To summarize, we have:

$$\overline{P_E^y}(R_y, R_{yz}) \leq 2^{-n \max_{0<\rho<1} E_{Y1}(R_y,\rho)} + 2^{-n \max_{0<\rho<1} E_{Y2}(R_y,R_{yz},\rho)}$$

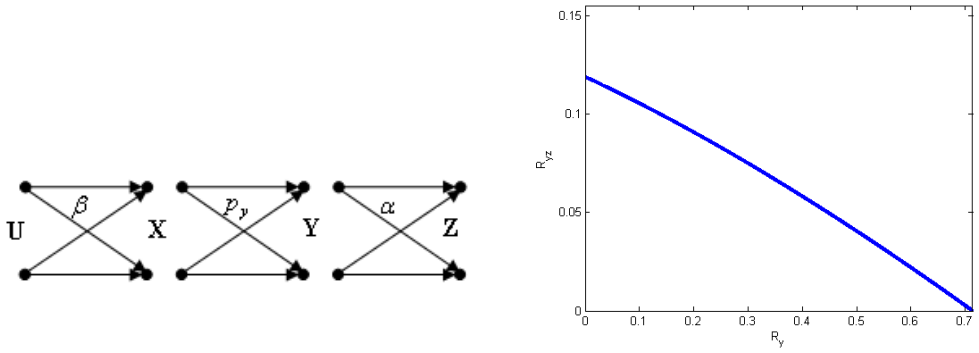
Taking the dominant exponent of the above sum yields the strong decoder exponent of Theorem 1.

4.3 Numerical Results for the Degraded BSC

In this section, we show some numerical results of our error exponents and compare them to the exponents that were derived in [4]. Our setup is that of a binary broadcast channel with a binary input X and separate binary symmetric channels to Y and Z with parameters p_y, p_z ($p_y < p_z < \frac{1}{2}$) respectively. This channel can be recast into a cascade of (degraded) binary symmetric channels with parameters p_y, α, where $\alpha = p(z \neq y) = \frac{p_z - p_y}{1-2p_y}$. In this case, the auxiliary random variable U is also binary. By symmetry, U is distributed uniformly on $\{0, 1\}$ and connected to X by another BSC with parameter β (see Fig. 1a). The capacity region is given by [16]:

$$R_z \leq 1 - h(\beta * p_z)$$
$$R_y \leq h(\beta * p_y) - h(p_y)$$

where $\beta * p = \beta(1-p) + (1-\beta)p$ and $h(x)$ is the binary entropy function given by $-x \log x - (1-x)\log(1-x)$ for $0 \leq x \leq 1$.



(a) (b)

Figure 1: (a)The recast channel with the auxiliary variable. (b)The capacity region $R_{yz}(R_y)$ with $p_y = 0.05, p_z = 0.3$

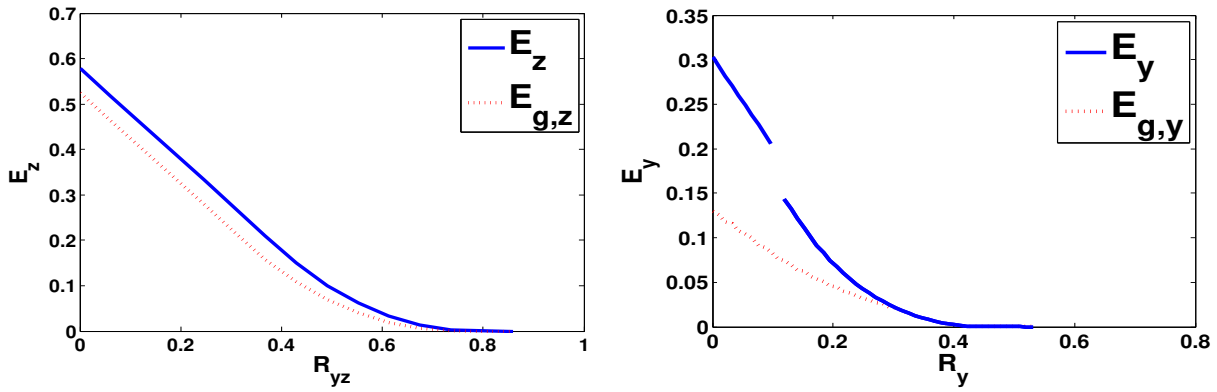
Denote the exponents of [4], calculated for this model, by $E_{g,y}, E_{g,z}$ for the strong and weak decoder, respectively. For a general channel, $E_{g,z}$ is given by (4). $E_{g,y}$ is the minimum between (23) and

$$\max_{\rho} \left\{ -\log \left[\sum_y \sum_u Q(u) \left(\sum_x Q(x|u) P_1^{\frac{1}{1+\rho}}(y|x) \right)^{1+\rho} \right] - \rho R_{yz} \right\}. \tag{27}$$

For given R_y and R_{yz}, β controls the tradeoff between the exponents (E_y, E_z). For example, if we are interested in finding the attainable pair (E_y, E_z) with maximal E_z for a given pair (R_y, R_{yz}), the maximizing β will be the smallest β s.t. E_y is positive, i.e., the value of β that maximizes $1 - H(\beta * p_z)$ while keeping $E_y > 0$. In Fig. 5, we show the best attainable (maximized over β) $E_y(R_y)$ for a given R_{yz} and the best attainable $E_z(R_{yz})$ for a given R_y compared to $E_{g,y}(R_y)$ and $E_{g,z}(R_{yz})$. In both cases the new exponents are better.

Note that the exponent value vanishes when the operating point is outside the capacity region (see Fig. 1b). The reason for this is that in Fig. 5a and Fig. 2b, we allowed the error exponents of the strong and weak decoders respectively, to be arbitrarily small. This allowed us to get arbitrarily close to the capacity region curve.

Although the values of E_z and $E_{g,z}$ in Fig. 5a are close, in the numerical calculation, it turned out that $\alpha = \mu \neq \frac{1}{1+\rho}$. We said above that in this case, the maximizing λ equals $\frac{1}{1+\rho}$. Therefore, since different parameters maximized E_z then the parameters in (4), the



(a) (b)

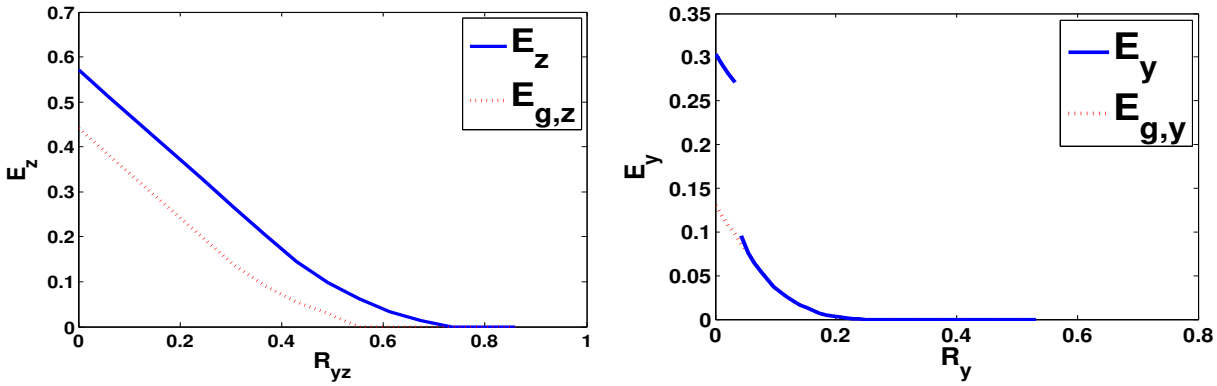
Figure 2: Comparing E_y, E_z (solid curves) to $E_{g,y}, E_{g,z}$(dotted curves) maximized over β. (a) $E_z(R_{yz})$ vs $E_{g,2}(R_{yz})$ for a fixed $R_y = 10^{-4}$. (b)$E_y(R_y)$ vs $E_{g,1}(R_y)$ for fixed $R_{yz} = 0.005$

new exponent is strictly larger than the exponent in [4] for all R_{yz} and the given R_y as long as $R_{yz} < 1 - h(p_z)$.

Denote the maximal value[2] of E_y, E_z by $E_{y_{max}}, E_{z_{max}}$ respectively. In Fig. 3 we repeat the calculation of Fig. 5. However, here we restrict $E_y \geq E_y^{min} = E_{y_{max}}/4, E_z \geq E_z^{min} = E_{z_{max}}/4$ in Fig. 3a and Fig. 3b, respectively. This time the exponents vanish deep inside the capacity region.

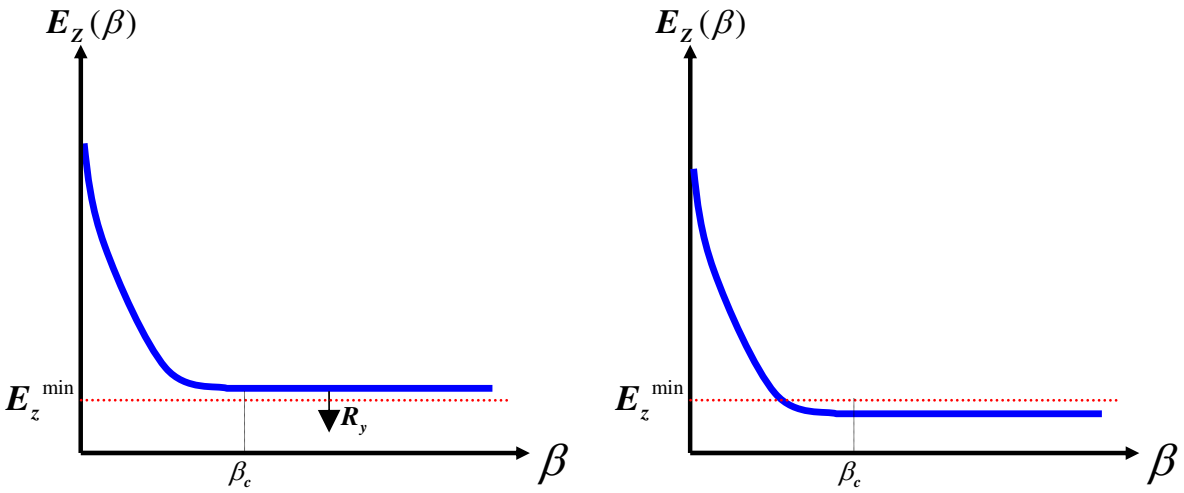
The reason for the singular points of E_y in Fig. 2b and Fig. 3b is the behavior of E_z as a function of β (illustrated in Fig. 4). Note that as β increases, the channel $U \to Z$ becomes noisier. Therefore $E_z(R_{yz}, R_y)$ is non increasing in β. For a given (R_{yz}, R_y) there is a critical value, β_c, such that for every $\beta \geq \beta_c$, $E_z(R_y, R_{yz}, \beta \geq \beta_c) \stackrel{\triangle}{=} E_{z_0}(R_y, R_{yz})$ is constant and has the form of (5), which is the single user error exponent ([14] p. 65) for the channel $X \to Z$ at rate $R_y + R_{yz}$. If $E_{z_0}(R_y, R_{yz})$ is greater than the threshold (for example $E_{z_0} \geq E_{z_{max}}/4$ in Fig. 3b) then the maximization over $E_y(R_y, R_{yz})$ is unconstrained and is attained by $\beta = 0.5$. However, as R_y increases, $E_{z_0}(R_y, R_{yz})$ decreases and at some critical R_{y_c}, $E_{z_0}(R_{y_c}, R_{yz})$ becomes smaller than the threshold (Illustrated in Fig 4.b).

[2]The maximal value is the single user error exponent ([14] p. 65) for the channel from X to Y and from X to Z for the strong and weak decoders respectively. i.e for a given R_{yz}, the maximal value for E_z is obtained with $R_y = 0$. For a given R_y the maximal E_y is obtained with $R_z = 0, \beta = 0.5$



(a) (b)

Figure 3: Comparing E_y, E_z (solid curves) to $E_{g,y}, E_{g,z}$(dotted curves) maximized over β. (a) $E_z(R_{yz})$ vs $E_{g,2}(R_{yz})$ for a fixed $R_y = 10^{-4}$ with $E_y \geq E_{y_{max}}/4$. (b)$E_y(R_y)$ vs $E_{g,1}(R_y)$ for fixed $R_{yz} = 0.005$ with $E_z \geq E_{z_{max}}/4$



(a) (b)

Figure 4: Illustration of E_z as a function of β. (a) for some $R_y < R_{y_c}$. E_z is above the threshold. (b) for $R_y > R_{y_c}$.

Thus, for $R_y \geq R_{y_c}$, the maximization of E_y becomes constrained and the largest valid β is much smaller than 0.5. Hence the sudden drop in the value of E_y. This phenomenon is not seen in $E_{g,y}$ since $E_{g,z}$ does not depend on R_y and the maximizing β is the same for all R_y.

5 Derivation for the Type Class Enumerators Approach

In this section, we prove Theorem 2. Throughout, we rely on the method of types [17]. We start with the notation we use in this section.

The empirical distribution pertaining to a vector $\boldsymbol{x} \in \mathcal{X}^n$ will be denoted by $\hat{Q}_{\boldsymbol{x}}$ and its type class by $T_{\boldsymbol{x}}$. In other words, $\hat{Q}_{\boldsymbol{x}} = \{\hat{q}_{\boldsymbol{x}}(a),\ a \in \mathcal{X}\}$, where $q_{\boldsymbol{x}}(a) = n_{\boldsymbol{x}}(a)/n$, $n_{\boldsymbol{x}}(a)$ being the number of occurrences of the letter a in $\boldsymbol{x}$. Similar conventions apply to empirical joint distributions of pairs of letters, $(a,b) \in \mathcal{X} \times \mathcal{Y}$, extracted from the corresponding pairs of vectors $(\boldsymbol{x}, \boldsymbol{y})$. Similarly, $\hat{q}_{\boldsymbol{x}|\boldsymbol{y}}(a|b) = \hat{q}_{\boldsymbol{x}\boldsymbol{y}}(a,b)/\hat{q}_{\boldsymbol{y}}(b)$ will denote the empirical conditional probability of $X = a$ given $Y = b$ (with convention that $0/0 = 0$), and $\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}$ will denote $\{\hat{q}_{\boldsymbol{x}|\boldsymbol{y}}(a|b),\ a \in \mathcal{X},\ b \in \mathcal{Y}\}$. $T_{\boldsymbol{x}|\boldsymbol{y}}$ will denote the conditional type class of $\boldsymbol{x}$ given $\boldsymbol{y}$. The expectation w.r.t. the empirical distribution of $(\boldsymbol{x}, \boldsymbol{y})$ will be denoted by $\hat{\boldsymbol{E}}_{\boldsymbol{x}\boldsymbol{y}}\{\cdot\}$, i.e., for a given function $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, we define $\hat{\boldsymbol{E}}_{\boldsymbol{x}\boldsymbol{y}}\{f(X,Y)\}$ as $\sum_{(a,b)\in\mathcal{X}\times\mathcal{Y}} \hat{q}_{\boldsymbol{x}\boldsymbol{y}}(a,b) f(a,b)$, where in this notation, X and Y are understood to be random variables jointly distributed according to $\hat{Q}_{\boldsymbol{x}\boldsymbol{y}}$. The entropy with respect to the empirical distribution of a vector $\boldsymbol{x}$ will be denoted by $\hat{H}(\boldsymbol{x})$. Finally, the notation $a_n \doteq b_n$ means that $\frac{1}{n}\log\frac{a_n}{b_n} \to 0$ as $n \to \infty$.

We start this section with the same initial step we used in the previous section. Namely, Gallager's general upper bound [14, p. 65] to the "channel" $P(\boldsymbol{z}|m) = \frac{1}{M_y}\sum_{i=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{x}_{m,i})$. The average error probability w.r.t. the ensemble of codes for $\lambda \geq 0, \rho \geq 0$ is given by:

$$\overline{P_{E_m}} \leq \sum_z \boldsymbol{E}\left[\frac{1}{M_y}\sum_{i=1}^{M_y} P(z|x_{m,i})\right]^{1-\rho\lambda} \times \boldsymbol{E}\left[\sum_{m'\neq m}\left(\frac{1}{M_y}\sum_{j=1}^{M_y} P(z|x_{m',j})\right)^{\lambda}\right]^{\rho} \quad \lambda \geq 0, \rho \geq 0. \tag{28}$$

We will see that both expectations depend on the $\boldsymbol{z}$ only through its empirical distribution. All the analysis is done for a given $\boldsymbol{z}$. The summation over all possible empirical distributions of $\boldsymbol{z}$ is done in the last step. $E_1(Q_z, R_y, R_{yz}, \rho, \lambda)$ and $E_2(Q_z, R_y, R_{yz}, \rho, \lambda)$ of Theorem 2 are the exponential rates of the first and second expectations in (28), respectively. After this initial step, our analysis is exponentially tight, whereas in the previous section, this is not necessarily the case. The price for this tightness is that the derivation and the resulting expression are much more involved, as we will see in the following subsections that derive $E_1(Q_z, R_y, R_{yz}, \rho, \lambda)$ and $E_2(Q_z, R_y, R_{yz}, \rho, \lambda)$.

5.1 Deriving $E_1(Q_z, R_y, R_{yz}, \rho, \lambda)$

Let $N_{z,m}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}})$ be a type class enumerator, that is, the number of codewords within cloud m having the same empirical conditional probability $\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}}$.

$$
\begin{aligned}
&\boldsymbol{E}\left[\frac{1}{M_y}\sum_{i=1}^{M_y} P(z|x_{m,i})\right]^{1-\rho\lambda} \\
&= M_y^{\rho\lambda-1}\boldsymbol{E}_u\boldsymbol{E}_{x|u}\left[\sum_{i=1}^{M_y} P(z|x_{mi})\right]^{1-\rho\lambda} \\
&= M_y^{\rho\lambda-1}\boldsymbol{E}_u\boldsymbol{E}_{x|u}\left[\sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}}} N_{z,m}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}})e^{n\hat{\mathbf{E}}\boldsymbol{z}\boldsymbol{x}\log P(Z|X)}\right]^{1-\rho\lambda} \\
&\doteq M_y^{\rho\lambda-1}\boldsymbol{E}_u\left[\sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}}} \boldsymbol{E}_{x|u}N_{z,m}^{1-\rho\lambda}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}})e^{n(1-\rho\lambda)\hat{\mathbf{E}}\boldsymbol{z}\boldsymbol{x}\log P(Z|X)}\right] \qquad (29)
\end{aligned}
$$

The last exponential equality is the first main point in our approach: It holds, even before taking the expectations because the summation over $\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}}$ consists of a sub-exponential number of terms. Thus, the key issue here is how to assess the moments of the type class enumerator.

Note that the probability, under $P(x^n|u^n) = \prod_{i=1}^n P(x_i|u_i)$, to fall into $T_{\boldsymbol{x}|\boldsymbol{u},\boldsymbol{z}}$ is

$$
|T_{\boldsymbol{x}|\boldsymbol{u},\boldsymbol{z}}| \cdot \prod_{a\in\mathcal{U},b\in\mathcal{X},c\in\mathcal{Z}} P(b|a)^{n\hat{P}(a,b,c)} \doteq e^{n(\hat{\mathbf{E}}\boldsymbol{x}\boldsymbol{u}\log P(X|U)+\hat{H}(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}))}
$$

Given $\boldsymbol{u}$, we independently generate e^{nR_y} codewords under $P(x^n|u^n) = \prod_{i=1}^n P(x_i|u_i)$. Therefore:

$$\boldsymbol{E}_{x|u} N_{z,m}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}}) \doteq e^{n(R_y + \hat{\mathbf{E}}_{\boldsymbol{x}\boldsymbol{u}} \log P(X|U) + \hat{H}(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}))} \tag{30}$$

The second main point of our approach is that the moments of the type class enumerator behave differently when the last exponent is positive or not (equivalently, $\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}} \in \mathcal{G}(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ or not). By the same arguments as in [10, Appendix]

$$\boldsymbol{E}_{x|u} N_{z,m}^{1-\rho\lambda}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}}) \tag{31}$$

$$\doteq \begin{cases} e^{n(1-\rho\lambda)(R_y + \hat{\mathbf{E}}_{\boldsymbol{x}\boldsymbol{u}} \log P(X|U) + \hat{H}(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}))} & \hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}} \in \mathcal{G}(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \\ e^{n(R_y + \hat{\mathbf{E}}_{\boldsymbol{x}\boldsymbol{u}} \log P(X|U) + \hat{H}(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}))} & \hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}} \in \mathcal{G}^c(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \end{cases} \tag{32}$$

We require $\rho\lambda \leq 1$ since the probability of $\{N_{z,m}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}}) = 0\}$ is positive, and so, negative moments of $N_{z,m}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{u},\boldsymbol{z}})$ diverge. The intuition behind this different behavior is that when $\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}} \in \mathcal{G}(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$, the enumerator concentrates extremely rapidly (double exponentially fast) around its expectation. However, when $\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}} \in \mathcal{G}^c(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ the enumerator is typically zero, and thus the dominant term when calculating the moment is $1^{1-\rho\lambda} \cdot \Pr(N_{z,m}^{1-\rho\lambda}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}}) = 1)$.

We continue from (29) by splitting the sum over all conditional types to those that belong

to $\mathcal{G}(R_y, Q_{u|z})$ and those that do not.

$$
\begin{aligned}
& M_y^{\rho\lambda-1} \boldsymbol{E}_u \left[\sum_{T_{x|z,u}} \boldsymbol{E}_{x|u} N_{z,m}^{1-\rho\lambda}(\hat{Q}_{x|z,u}) e^{n(1-\rho\lambda)\hat{\mathbf{E}}_{\boldsymbol{zx}} \log P(Z|X)} \right] \\
\dot{=} & \boldsymbol{E}_u \left\{ \sum_{\mathcal{G}(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})} e^{n(1-\rho\lambda)(\hat{\mathbf{E}}_{\boldsymbol{xu}} \log P(X|U) + \hat{H}(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}) + \hat{\mathbf{E}}_{\boldsymbol{zx}} \log P(Z|X))} + \right. \\
& \left. \sum_{\mathcal{G}^c(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})} e^{n((\rho\lambda)R_y + \hat{\mathbf{E}}_{\boldsymbol{xu}} \log P(X|U) + \hat{H}(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}) + (1-\rho\lambda)\hat{\mathbf{E}}_{\boldsymbol{zx}} \log P(Z|X))} \right\} \\
\dot{=} & \boldsymbol{E}_u (e^{n\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})} + e^{n\beta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}) \\
\dot{=} & \max_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \Pr(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}|\boldsymbol{z}) (e^{n\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})} + e^{n\beta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})})
\end{aligned} \tag{33}
$$

the last line is true since $\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ and $\beta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ (cf. (6), (7)) depend on $\boldsymbol{u}$ through $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}$. $\Pr(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}|\boldsymbol{z})$ is the probability, under $P(u^n) = \prod_{i=1}^n P(u_i)$, to belong to $T_{\boldsymbol{u}|\boldsymbol{z}}$ which equals (exponentially) to $e^{n(\hat{\mathbf{E}}_{\boldsymbol{u}} \log P(U) + \hat{H}(\boldsymbol{u}|\boldsymbol{z}))}$. If we have used Jensen's inequality, instead of the above tight steps, the last sum would contain only $e^{n\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}$ and the expression of $\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ would contain a global maximization rather than the constrained optimization of (6). Therefore, Jensen inequality is tight whenever the unconstrained achiever of $\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ is in $\mathcal{G}(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ and $\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \geq \beta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ (See [18, Appendix E] for more detains)

We start by evaluating $\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$: The unconstrained achiever of the optimization in (6) is $P(x|z,u)$ and it belongs to $\mathcal{G}(R_y, \hat{Q}\boldsymbol{u}|\boldsymbol{z})$ for large enough R_y if $R_y - \hat{I}(\boldsymbol{x};\boldsymbol{z}|\boldsymbol{u}) \geq 0$ (Here, unlike the single user case [10], such R_y can be in the capacity region). If $P(x|z,u) \in \mathcal{G}(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ The maximum in (6) will be obtained with the empirical distribution $\hat{Q}(x|u,z) = P(x|u,z)$ (as $n \to \infty$).

We now consider the case in which $P(x|z,u) \in \mathcal{G}^c(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$. Following the exact arguments of [10, Section 4.3], any internal point of $\mathcal{G}(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ can be improved by a point on the boundary of $\mathcal{G}(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ when $P(x|z,u) \in \mathcal{G}^c(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$. The achieving pmf will thus be

$$
Q^*(x|z,u) = \frac{P(x|u) P_3^{\delta_R(u)}(z|x)}{\sum_x P(x|u) P_3^{\delta_R(u)}(z|x)} \tag{34}
$$

where $\delta_R(u)$ is such that $-R_y = \hat{\mathbf{E}}_{Q^*} \log P(x|u) + \hat{H}_{Q^*}(x|z,u)$. The existence of $\delta_R(u)$ is discussed in Section A.2. Using the above arguments, since the constrained maximizer will be on the boundary of $\mathcal{G}(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$, we can use the fact that on the boundary $-R_y = \hat{\mathbf{E}}_Q \log P(x|u) + \hat{H}_Q(x|z,u)$ to get:

$$\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \doteq (1-\rho\lambda)(-R_y + \max_{\mathcal{G}(R_y, \hat{Q}\boldsymbol{u}|\boldsymbol{z})} \hat{\mathbf{E}}_{\boldsymbol{z}\boldsymbol{x}} \log P(Z|X)) \tag{35}$$

$$= (1-\rho\lambda)(-R_y + \hat{\mathbf{E}}_{Q^*} \log P(Z|X)) \tag{36}$$

To summarize, when $P(x|z,u) \in \mathcal{G}(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ we have

$$\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \doteq (1-\rho\lambda)\boldsymbol{E}_{P_{x|u,z}} \log P(X|U) + H_{P_{x|u,z}}(X|U,Z) + \boldsymbol{E}_{P_{x|u,z}} \log P_3(Z|X) \tag{37}$$

and when $P(x|z,u) \in \mathcal{G}^c(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ we have

$$\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \doteq (1-\rho\lambda)(-R_y + \hat{\mathbf{E}}_{Q^*} \log P_3(Z|X)). \tag{38}$$

We now proceed by evaluating $\beta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$.

The unconstrained achiever of (7) is

$$Q_{1-\rho\lambda}(x|u,z) = \frac{P(x|u)P^{1-\rho\lambda}(z|x)}{\sum_{x'} P(x'|u)P_3^{1-\rho\lambda}(z|x')}.$$

$R_y, (1-\rho\lambda)$ will determine if $Q_{1-\rho\lambda}(x|u,z) \in \mathcal{G}^c(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$. From the proof of the existence of $\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ (Section A.2) it is easily seen that the unconstrained achiever is outside $\mathcal{G}^c(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ when $P(x|u,z) \in \mathcal{G}(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ or when $1-\rho\lambda \leq \delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$. In this case, by the same arguments as before, the constrained achiever will be on the boundary and therefore:

$$\beta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \doteq (1-\rho\lambda)\left[-R_y + \hat{\mathbf{E}}_{Q^*} \log P(z|x)\right] \tag{39}$$

where $Q^*(x|u,z)$ is defined in (34).

In the case where $Q_{1-\rho\lambda}(x|u,z) \in \mathcal{G}^c(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ $(1-\rho\lambda \leq \delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}))$, for simplicity, set

$c(1-\rho\lambda, U, Z) = \sum_X P(X|U)P_3^{1-\rho\lambda}(Z|X)$. We have

$$\begin{aligned}
\beta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) &= \rho\lambda R_y + \hat{\mathbf{E}}_{Q_{1-\rho\lambda}} \log[P(X|U)P^{1-\rho\lambda}(Z|X)] + \hat{H}_{Q_{1-\rho\lambda}}(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}) \\
&= \rho\lambda R_y + \hat{\mathbf{E}}_{Q_{1-\rho\lambda}} \left\{\log[P(X|U)P^{1-\rho\lambda}(Z|X)] - \log Q_{1-\rho\lambda}(X|U,Z)\right\} \\
&= \rho\lambda R_y + \hat{\mathbf{E}}_{Q_{1-\rho\lambda}} \left\{\log[P(X|U)P^{1-\rho\lambda}(Z|X)] - \log \frac{P(X|U)P^{1-\rho\lambda}(Z|X)}{c(1-\rho\lambda,U,Z)}\right\} \\
&= \rho\lambda R_y + \hat{\mathbf{E}}_{uz} \log c(1-\rho\lambda,U,Z) \qquad (40)
\end{aligned}$$

To summarize:

$$\beta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \doteq \begin{cases} (1-\rho\lambda)(-R_y + \hat{\mathbf{E}}_{Q^*} \log P(Z|X)) & P(x|z,u) \in \mathcal{G}(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \text{ or } \rho\lambda \geq 1-\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \\ \rho\lambda R_y + \hat{\mathbf{E}}_{uz} c(\rho\lambda, u, z) & \rho\lambda < 1-\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \end{cases} \qquad (41)$$

And finally, letting $E_{\alpha\beta} = \max\{\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}), \beta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})\}$, substituting it into (33) and letting $n \to \infty$ yields $E_1(Q_z, R_y, R_{yz}, \rho, \lambda)$.

5.2 Deriving $E_2(Q_z, R_y, R_{yz}, \rho, \lambda)$

We now proceed to the second expectation of the original bound.

$$\begin{aligned}
&\boldsymbol{E}\left[\sum_{m'\neq m}\left(\frac{1}{M_y}\sum_{j=1}^{M_y} P(\boldsymbol{z}|\boldsymbol{x}_{j,m'})\right)^{\lambda}\right]^{\rho} \\
=&M_y^{-\rho\lambda}\boldsymbol{E}\left[\sum_{m'\neq m}\left(\sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}} N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}) e^{n\hat{\mathbf{E}}\boldsymbol{z}\boldsymbol{x}\log P_3(Z|X)}\right)^{\lambda}\right]^{\rho} \\
\doteq&M_y^{-\rho\lambda}\boldsymbol{E}\left[\sum_{m'\neq m}\sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}} N_{z,m'}^{\lambda}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}) e^{n\lambda\hat{\mathbf{E}}\boldsymbol{z}\boldsymbol{x}\log P_3(Z|X)}\right]^{\rho} \\
\doteq&M_y^{-\rho\lambda}\boldsymbol{E}\left[\sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}}\sum_{m'\neq m} N_{z,m'}^{\lambda}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}) e^{n\lambda\hat{\mathbf{E}}\boldsymbol{z}\boldsymbol{x}\log P_3(Z|X)}\right]^{\rho} \\
\doteq&M_y^{-\rho\lambda}\sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}} e^{n\lambda\rho\hat{\mathbf{E}}\boldsymbol{z}\boldsymbol{x}\log P_3(Z|X)}\boldsymbol{E}\left[\sum_{m'\neq m} N_{z,m'}^{\lambda}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}})\right]^{\rho} \qquad (42)
\end{aligned}$$

Here, unlike the previous subsection, there are two main obstacles. The first is the inner sum over $m' \neq m$ which has an exponential number of terms. In the previous subsection, when we used the enumerators, the resulting sums had only a polynomial number of terms, which allowed us to distribute the expectation operator and moments over the summands without loosing exponential tightness. Here we have to use a different approach. The second obstacle is that the enumerators, $N^{\lambda}_{z,m'}(T_{x|z})$, are distributed differently for every m' (since the codewords are drawn given $\boldsymbol{u}'_m$). Note however, that for all $\boldsymbol{u}_m$ that belong to the same conditional type $T_{\boldsymbol{u}|\boldsymbol{z}}$ the corresponding enumerators are identically distributed. We use this fact in the following.

We continue by dividing $[0, R_{yz}]$ into a grid with a sub-exponential number of intervals in n (for example, $d = \frac{R_{yz}}{n}$). Evaluating the last expectation in (42), we have:

$$\begin{aligned}
&\boldsymbol{E}\left[\sum_{m'\neq m} N^{\lambda}_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}})\right]^{\rho} \\
&= \boldsymbol{E}\left[\sum_{A=0}^{R_{yz}} (\text{number of times } N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}) \doteq e^{nA}) e^{n\lambda A}\right]^{\rho} \\
&\doteq \sum_{A=0}^{R_{yz}} e^{n\lambda\rho A} \boldsymbol{E}\left[(\text{number of times } N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}) \doteq e^{nA})\right]^{\rho} \\
&\doteq \sum_{A=0}^{R_{yz}} e^{n\lambda\rho A} \boldsymbol{E}\left[\sum_{m'\neq m} I_{m'}(A)\right]^{\rho} \qquad (43)
\end{aligned}$$

where $I_{m'}(A) \stackrel{\triangle}{=} \mathcal{I}\left(N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}) \doteq e^{nA}\right)$, omitting the dependence on $\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}$ to simplify notation). Next, we partition the summation over m' into subsets in which the enumerators are identically distributed as described above.

$$\boldsymbol{E}\left[\sum_{m'\neq m} I_{m'}(A)\right]^{\rho} = \boldsymbol{E}\left[\sum_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(A)\right]^{\rho}$$
$$\doteq \sum_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(A)\right]^{\rho} \tag{44}$$

Note that the number of terms in the inner summation of (44) is a random variable. Define $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \stackrel{\triangle}{=} |m' : \boldsymbol{u}_{m'} \in T_{\boldsymbol{u}|\boldsymbol{z}}|$ - the number of cloud centers that belong to the same conditional type. Since we draw $e^{nR_{yz}}$ cloud centers independently with $P(u^n) = \prod_{i=1}^{n} P(u_i)$ we have:

$$\boldsymbol{E}\left[M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}\right] \doteq e^{n(R_{yz}+\hat{H}(\boldsymbol{u}|\boldsymbol{z})+\hat{\mathbf{E}}_{\boldsymbol{u}}\log P(U))} \stackrel{\triangle}{=} e^{n\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}$$

The sign of the last exponent determines if we are likely to find an exponential number of cloud centers of this type. We show in Section A.3 that when $\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) > 0$ (i.e $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z(R_{yz})$), $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$ converges to its expectation double exponentially fast. When $\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \leq 0$, $\Pr\left(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} > e^{n\epsilon}\right)$ vanishes double exponentially fast.
Let $P_A(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \stackrel{\triangle}{=} \Pr\{I_{m'}(A) = 1\}$ denote the probability that we have e^{nA} codewords around cloud m' that belong to $T_{\boldsymbol{x}|\boldsymbol{z}}$. Define

$$A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) = \left[N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}, R_y)\right]^{+}$$

We show in Section A.4 that when $A = A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) > 0$, $P_{A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ converges to 1 and vanishes for every other A double exponentially fast. When $A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) = 0$, we show that $P_{A=0}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) = e^{nN(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}, R_y)}$. Thus, the outer summation in (43) consists only of those $A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ and the number of elements in the summation is upper bounded by $|\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}| \times |\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}|$ which is sub-exponential in n.

Continuing (44), there are four cases: the combinations of $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z(R_{yz})$ or not and $A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) > 0$ or $A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) = 0$. We start with the case $A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) > 0$.

5.2.1 The case $A = A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) > 0$

We need to evaluate:

$$\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(A)\right]^{\rho} \tag{45}$$

We use the fact that for $A = A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$, $P_A(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) > 1-\epsilon$, for some $\epsilon > 0$ that vanishes double exponentially fast (see Section A.4), to show that the probability that all the indicators, $I_{m'}(A)$, equal one is very likely. Denote this event by A:

$$\Pr(\mathcal{A}) \geq (1-\epsilon)^{M_{\hat{Q}}\boldsymbol{u}|\boldsymbol{z}} = e^{M_{\hat{Q}}\boldsymbol{u}|\boldsymbol{z}\log(1-\epsilon)} \geq e^{M_{\hat{Q}}\boldsymbol{u}|\boldsymbol{z}\frac{-\epsilon}{1-\epsilon}} \tag{46}$$

$M_{\hat{Q}\boldsymbol{u}|\boldsymbol{z}}$ is a random variable in $[0, e^{nR_{yz}}]$. Since ϵ vanishes double exponentially fast we have $\Pr(\mathcal{A}) \to 1$ double exponentially fast.

$$\begin{aligned}
&\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(A)\right]^{\rho} \\
&\Pr(\mathcal{A})\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(A)|\mathcal{A}\right]^{\rho} + \Pr(\mathcal{A}^c)\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(A)|\mathcal{A}^c\right]^{\rho} \\
&= \boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(A)|\mathcal{A}\right]^{\rho} \\
&= \boldsymbol{E}\left[M_{\hat{Q}\boldsymbol{u}|\boldsymbol{z}}|\mathcal{A}\right]^{\rho}
\end{aligned} \tag{47}$$

In the second to the last line we used the fact that $\Pr(\mathcal{A}^c) \to 0$ fast enough to make the second term in the summation negligible (note that the expectation value can grow, at most, at an exponential rate while $\Pr(\mathcal{A}^c)$ vanishes double exponentially fast). In the last step we used the fact that given $\mathcal{A}$, all the indicators are equal to one. Note that the conditioning on the event $\mathcal{A}$ introduces dependencies between the drawings of the codewords x and clouds u. (given $\mathcal{A}$ for instance, there might be some $u \in \mathcal{U}$ which cannot be drawn. therefore the clouds are no longer drawn according to $\prod_{i=1}^{n} P(u_i)$). We claim that since the conditioning in (47) is on an event which is very likely (its probability is very close to 1), we can remove

the conditioning without changing much the resulting value. To see this, Let $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$ be distributed with some distribution measure Q.

$$Q(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}) = \Pr(\mathcal{A})Q(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}|\mathcal{A}) + \Pr(\mathcal{A}^c)Q(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}|\mathcal{A}^c) \geq (1-\epsilon)Q(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}|\mathcal{A}) \tag{48}$$

on the other hand,

$$Q(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}) = \Pr(\mathcal{A})Q(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}|\mathcal{A}) + \Pr(\mathcal{A}^c)Q(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}|\mathcal{A}^c) \leq Q(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}|\mathcal{A}) + \epsilon \cdot 1. \tag{49}$$

therefore,

$$Q(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}) - \epsilon \leq Q(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}|\mathcal{A}) \leq \frac{Q(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}})}{1-\epsilon}. \tag{50}$$

Since $\epsilon \to 0$ double exponentially fast, we can replace $Q(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}|\mathcal{A})$ by $Q(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}})$ in the calculation of the expectation in (47) and preserve exponential tightness. Using Section A.3 for $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z(R_{yz})$ we have:

$$\begin{aligned}\boldsymbol{E}\left[M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}\right]^{\rho} &\leq e^{n\rho\left[\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+\epsilon\right]} Pr\left\{M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \leq e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+\epsilon)}\right\} + e^{nR_{yz}} Pr\left\{M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \geq e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+\epsilon)}\right\} \\ &\leq e^{n\rho\left[\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+\epsilon\right]} + e^{nR_{yz}} e^{-n\epsilon e^{n\left[\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+\epsilon\right]}}\end{aligned} \tag{51}$$

On the other hand:

$$\begin{aligned}\boldsymbol{E}\left[M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}\right]^{\rho} &\geq e^{n\rho\left[\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})-\epsilon\right]} Pr\left\{M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \geq e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})-\epsilon)}\right\} \\ &= e^{n\rho\left[\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})-\epsilon\right]} \left\{1 - Pr\left\{M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} < e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})-\epsilon)}\right\}\right\} \\ &\geq e^{n\rho\left[\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})-\epsilon\right]} \left\{1 - e^{-n\epsilon e^{n\left[\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})-\epsilon\right]}}\right\}\end{aligned} \tag{52}$$

Finally we have for $\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \geq 0$

$$\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'} \in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(A)\right]^{\rho} \doteq e^{n\rho\left[\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})\right]} \tag{53}$$

When $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z^c(R_{yz})$ we have:

$$\boldsymbol{E}\left[M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}\right]^{\rho} \leq \qquad e^{n\rho\epsilon} \Pr\left\{1 \leq M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \leq e^{n\epsilon}\right\} + e^{nR_{yz}} \Pr\left\{M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \geq e^{n\epsilon}\right\} \tag{54}$$

The second term vanishes since the probability that $M_{\hat{Q}\boldsymbol{u}|\boldsymbol{z}} > e^{n\epsilon}$ vanishes double exponentially fast for $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z^c(R_{yz})$. Neglecting the second term and using the properties of $M_{\hat{Q}\boldsymbol{u}|\boldsymbol{z}}$, proved in Section A.3, we continue:

$$\begin{aligned}\boldsymbol{E}\left[M_{\hat{Q}\boldsymbol{u}|\boldsymbol{z}}\right]^{\rho} &\le e^{n\rho\epsilon}\Pr\left\{M_{\hat{Q}\boldsymbol{u}|\boldsymbol{z}} \ge 1\right\} \\ &\le e^{n\rho\epsilon}\boldsymbol{E}\left\{M_{\hat{Q}\boldsymbol{u}|\boldsymbol{z}}\right\} \\ &= e^{n\rho\epsilon}e^{n\bar{m}(\hat{Q}\boldsymbol{u}|\boldsymbol{z})}\end{aligned} \tag{55}$$

On the other hand:

$$\boldsymbol{E}\left[M_{\hat{Q}\boldsymbol{u}|\boldsymbol{z}}\right]^{\rho} \ge 1\cdot\Pr\left\{M_{\hat{Q}\boldsymbol{u}|\boldsymbol{z}} = 1\right\} = e^{n\bar{m}(\hat{Q}\boldsymbol{u}|\boldsymbol{z})} \tag{56}$$

Therefore, since we can let ϵ vanish sufficiently slowly with n, e.g. $\epsilon = 1/\sqrt{n}$, we have for $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z^c(R_{yz})$:

$$\boldsymbol{E}\left[M_{\hat{Q}\boldsymbol{u}|\boldsymbol{z}}\right]^{\rho} \doteq e^{n\bar{m}(\hat{Q}\boldsymbol{u}|\boldsymbol{z})} \tag{57}$$

To conclude this subsection, when $A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) > 0$:

$$\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(A)\right]^{\rho} \doteq \begin{cases} e^{n\rho\bar{m}(\hat{Q}\boldsymbol{u}|\boldsymbol{z})} & \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z(R_{yz}) \\ e^{n\bar{m}(\hat{Q}\boldsymbol{u}|\boldsymbol{z})} & \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z^c(R_{yz}) \end{cases} \tag{58}$$

5.2.2 The case $A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) = 0$

Here, as before, we divide into two cases: $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z(R_{yz})$ or $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z^c(R_{yz})$. Unlike the previous case, where we knew that $P_{A,\hat{Q}\boldsymbol{u}|\boldsymbol{z}}$ converges to 1 double exponentially fast, here, we know that $P_0(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \doteq e^{nN(\hat{Q}\boldsymbol{x}|\boldsymbol{z},\hat{Q}\boldsymbol{u}|\boldsymbol{z},R_y)}$ ($N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}, R_y) \le 0$, see Section A.4). Therefore, we have to use a somewhat different approach. We start with the case of

$\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z(R_{yz})$

$$\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0)\right]^{\rho} \leq$$

$$e^{n\rho\left[\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)+\epsilon\right]} Pr\left\{\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0) \leq e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y))+\epsilon}\right\} +$$

$$e^{nR_{yz}} Pr\left\{\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0) \geq e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)+\epsilon)}\right\} \tag{59}$$

Focusing on the probability in second term:

$$\begin{aligned}
&Pr\left\{\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0) \geq e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y))+\epsilon}\right\} \\
&= \sum_{m=0}^{e^{nR_{yz}}} Pr\left\{M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \doteq e^{nm}\right\} \times \\
&\qquad Pr\left\{\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0) \geq e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)+\epsilon)} | M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \doteq e^{nm}\right\} \\
&= Pr\left\{M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \doteq e^{n\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}\right\} \times \\
&\qquad Pr\left\{\sum_{m':\boldsymbol{u}_{m'}\in \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0) \geq e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)+\epsilon)} | M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \doteq e^{\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}\right\}
\end{aligned} \tag{60}$$

The last step is true because of the concentration of $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$ around its expectation when $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z(R_{yz})$. Therefore $Pr\left\{M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \doteq e^{n\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}\right\} \to 1$ double exponentially fast (see Section A.3). Here, as in the previous subsection, we condition on an event which is extremely likely. By the same arguments we used in the previous subsection, we remove the conditioning. Continuing (60) we have:

$$= Pr\left\{\sum_{m=1}^{e^{n\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}} I_{m'}(0) \geq e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)+\epsilon)}\right\} \tag{61}$$

We are left with analyzing the probability that we have more than $e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)+\epsilon)}$ successes in $e^{n\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}$ independent Bernoulli trials with prob-

ability $e^{nN(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)}$ each. By using the Chernoff bound, it is easily seen that the probability that this will happen, vanishes double exponentially fast, since we have an exponential number of trials. We therefore have:

$$\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0)\right]^{\rho} \leq e^{\rho\left[n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)+\epsilon)\right]} \tag{62}$$

The upper bound for $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z(R_{yz})$ is given by

$$\begin{aligned}
&\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0)\right]^{\rho} \geq \\
&e^{\rho\left[n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)-\epsilon)\right]} Pr\left\{\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0) \geq e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)-\epsilon)}\right\} \\
&= e^{\rho\left[n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)-\epsilon)\right]}\times \\
&\quad\left\{1-Pr\left\{\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0) < e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)-\epsilon)}\right\}\right\}
\end{aligned} \tag{63}$$

By the same arguments we used in the upper bound, the last probability vanishes double exponentially fast. So we have for $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z(R_{yz})$:

$$\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0)\right]^{\rho} \doteq e^{n\rho\left[\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y))\right]} \tag{64}$$

We now continue to the case $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z^c(R_{yz})$. Here, we know that $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$ is sub-exponential (the probability that $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$ in sub exponential converges to 1 double exponentially fast). Therefore, we will not be able to apply the Chernoff bound as we did before in (61). Again, we use a different approach.

$$
\begin{aligned}
&\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0)\right]^{\rho} \\
&= Pr\left\{M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} < e^{n\epsilon}\right\}\boldsymbol{E}\left\{\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0)\right]^{\rho}\middle| M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} < e^{n\epsilon}\right\} \\
&\quad + Pr\left\{M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \geq e^{n\epsilon}\right\}\boldsymbol{E}\left\{\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0)\right]^{\rho}\middle| M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \geq e^{n\epsilon}\right\}
\end{aligned} \tag{65}
$$

The second term can be neglected since the $Pr\left\{M_{T_{\boldsymbol{u}|\boldsymbol{z}}} \geq e^{n\epsilon}\right\}$ vanishes double exponentially fast for $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z^c(R_{yz})$ and the expectation grows at most at an exponential rate. Since we know that the number of elements in the sum over m' is of sub exponential order, we can distribute ρ over the summands and still preserve exponential tightness.

$$
\doteq \boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}^{\rho}(0)\middle| M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} < e^{n\epsilon}\right] \tag{66}
$$

We now condition on $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$. Doing this alone would introduce dependencies between the $\boldsymbol{u}$'s and $\boldsymbol{x}$ and change the probability law of the indicator function. To avoid this, we condition also on $\boldsymbol{u}_{m'}$. Given a specific $\boldsymbol{u}_{m'}$ all drawing of $\boldsymbol{x}_{m',i}$ are independent and $P_{A=0}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ remains intact.

$$
= \boldsymbol{E}_{M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}}\boldsymbol{E}_{\boldsymbol{u}}\left\{\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} \boldsymbol{E}\left[I_{m'}(0)|M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}, \boldsymbol{u}\right]\right\} \tag{67}
$$

Given $\boldsymbol{u}$ the inner expectation is independent of the number of such $\boldsymbol{u}$'s ($M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$) and becomes $P_{A=0}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$. Now, since $P_{A=0}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ is constant for all $\boldsymbol{u}$'s in the conditional type $T_{\boldsymbol{u}|\boldsymbol{z}}$ the expectation over $\boldsymbol{u}$ doesn't change the value and we are left with:

$$
\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0)\right]^{\rho} \doteq e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y))} \tag{68}
$$

To summarize this subsection: When $A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) = 0$ we have

$$\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(A)\right]^{\rho} \doteq \begin{cases} e^{n\rho\left[\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)\right]} & \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z(R_{yz}) \\ e^{n\left[\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)\right]} & \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z^c(R_{yz}) \end{cases} \tag{69}$$

5.2.3 Wrapping up

Using the results we obtained in the previous two subsections, we are now ready to continue (43).

$$\begin{aligned} \boldsymbol{E}\left[\sum_{m'\neq m} N^{\lambda}_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}})\right]^{\rho} &\doteq \sum_{A\geq 0}^{R_{yz}} e^{n\lambda\rho A} \sum_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(A)\right]^{\rho} \\ &= \sum_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \sum_{A\geq 0}^{R_{yz}} e^{n\lambda\rho A} \boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(A)\right]^{\rho} \end{aligned} \tag{70}$$

We saw that for all $A \neq A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ the inner sum vanishes. Using definitions (12) and (13) we continue:

$$\begin{aligned} &= \sum_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} e^{n\lambda\rho A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})} \boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(A^*)\right]^{\rho} \\ &\doteq \sum_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}\in\mathcal{G}_z(R_{yz})} e^{n(B(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)+\rho\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}))} + \sum_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}\in\mathcal{G}_z^c(R_{yz})} e^{n(C(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)+\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}))} \\ &\doteq e^{n\cdot\max\left\{max_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}\in\mathcal{G}_z(R_{yz})}\left[B(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)+\rho\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})\right],\max_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}\in\mathcal{G}_z^c(R_{yz})}\left[C(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)+\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})\right]\right\}} \\ &\triangleq e^{nE(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}})}. \end{aligned} \tag{71}$$

Substituting this into (42), we have:

$$\begin{aligned} &\boldsymbol{E}\left[\sum_{m'\neq m}\left(\frac{1}{M_y}\sum_{j=1}^{M_y} P(z|x)\right)^{\lambda}\right]^{\rho} \\ &\doteq e^{-n\left\{\max_{\hat{Q}}\boldsymbol{x}|\boldsymbol{z}\ \lambda\rho\hat{\mathbf{E}}\boldsymbol{z}\boldsymbol{x}\log\frac{1}{P(Z|X)}-E(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}})+\rho\lambda R_y\right\}} \end{aligned} \tag{72}$$

When $n \to \infty$, this is the expression of $E_2(Q_Z, R_y, R_{yz}, \rho, \lambda)$ of Theorem 2.

5.3 The Strong Decoder

We now proceed to the derivation of the strong decoder exponent. We start with the same steps as in the Gallager-type approach (22):

$$\begin{aligned}
\overline{P^{y}_{E_{m,i}}} &\leq \boldsymbol{E} \sum_{\boldsymbol{y}} P_1(\boldsymbol{y}|\boldsymbol{x}_{m,i}) \left(\sum_{(m',i')\neq(m,i)} \frac{P_1(\boldsymbol{y}|\boldsymbol{x}_{m',i'})^{\lambda}}{P_1(\boldsymbol{y}|\boldsymbol{x}_{m,i})^{\lambda}} \right)^{\rho} \\
&= \boldsymbol{E} \sum_{\boldsymbol{y}} P_1(\boldsymbol{y}|\boldsymbol{x}_{m,i})^{1-\lambda\rho} \left(\sum_{i'\neq i} P_1(\boldsymbol{y}|\boldsymbol{x}_{m,i'})^{\lambda} + \sum_{m'\neq m} \sum_{i'=1}^{M_y} P_1(\boldsymbol{y}|\boldsymbol{x}_{m',i'})^{\lambda} \right)^{\rho} \\
&\doteq \boldsymbol{E} \sum_{\boldsymbol{y}} P_1(\boldsymbol{y}|\boldsymbol{x}_{m,i})^{1-\lambda\rho} \left[\left(\sum_{i'\neq i} P_1(\boldsymbol{y}|\boldsymbol{x}_{m,i'})^{\lambda} \right)^{\rho} + \left(\sum_{m'\neq m} \sum_{i'=1}^{M_y} P_1(\boldsymbol{y}|\boldsymbol{x}_{m',i'})^{\lambda} \right)^{\rho} \right] \\
&\triangleq \boldsymbol{E} P_{E_{y1}} + \boldsymbol{E} P_{E_{y2}}
\end{aligned} \tag{73}$$

As before, we evaluate the expressions for a given $\boldsymbol{y}$ and sum over all $\boldsymbol{y}$ in the last step. We start with $P_{E_{y1}}$

$$\begin{aligned}
P_{E_{y1}} &= \boldsymbol{E} \sum_{\boldsymbol{y}} P_1(\boldsymbol{y}|\boldsymbol{X}_{m,i})^{1-\lambda\rho} \left(\sum_{i'\neq i} P_1(\boldsymbol{y}|\boldsymbol{X}_{m,i'})^{\lambda} \right)^{\rho} \\
&= \sum_{\boldsymbol{y}} \boldsymbol{E} P_1(\boldsymbol{y}|\boldsymbol{X}_{m,i})^{1-\lambda\rho} \boldsymbol{E} \left(\sum_{i'\neq i} P_1(\boldsymbol{y}|\boldsymbol{X}_{m,i'})^{\lambda} \right)^{\rho}
\end{aligned} \tag{74}$$

The first expectation becomes:

$$
\begin{aligned}
\boldsymbol{E}P_1(\boldsymbol{y}|\boldsymbol{X}_{m,i})^{1-\lambda\rho} &= \boldsymbol{E}_u\boldsymbol{E}_{x|u}P_1(\boldsymbol{y}|\boldsymbol{X}_{m,i})^{1-\lambda\rho} \\
&= \boldsymbol{E}_u \sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{u}\boldsymbol{y}}} \Pr\left(\hat{Q}_{\boldsymbol{x}|\boldsymbol{u}\boldsymbol{y}}\right) e^{n(1-\rho\lambda)\hat{\mathbf{E}}\boldsymbol{y}\boldsymbol{x}\log P(Y|X)} \\
&\doteq \boldsymbol{E}_u \max_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{u}\boldsymbol{y}}} \Pr\left(\hat{Q}_{\boldsymbol{x}|\boldsymbol{u}\boldsymbol{y}}\right) e^{n(1-\rho\lambda)\hat{\mathbf{E}}\boldsymbol{y}\boldsymbol{x}\log P(Y|X)} \\
&\doteq \max_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}}} \Pr\left(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}}\right) e^{n\max_{\hat{Q}}\boldsymbol{x}|\boldsymbol{u}\boldsymbol{y}\,(\hat{\mathbf{E}}\boldsymbol{x}\boldsymbol{u}\log P(X|U)+\hat{H}(\boldsymbol{x}|\boldsymbol{u},\boldsymbol{y})+(1-\rho\lambda)\hat{\mathbf{E}}\boldsymbol{y}\boldsymbol{x}\log P(Y|X))} \\
&\doteq \max_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}}} e^{n(\hat{\mathbf{E}}\boldsymbol{u}\log P(U)+\hat{H}(\boldsymbol{u}|\boldsymbol{y}))} e^{n\max_{\hat{Q}}\boldsymbol{x}|\boldsymbol{u}\boldsymbol{y}\,(\hat{\mathbf{E}}\boldsymbol{x}\boldsymbol{u}\log P(X|U)+\hat{H}(\boldsymbol{x}|\boldsymbol{u},\boldsymbol{y})+(1-\rho\lambda)\hat{\mathbf{E}}\boldsymbol{y}\boldsymbol{x}\log P(Y|X))} \\
&\doteq \max_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}}} \max_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{u}\boldsymbol{y}}} e^{n(\hat{\mathbf{E}}\boldsymbol{u}\boldsymbol{x}\log P(U,X)+\hat{H}(\boldsymbol{x},\boldsymbol{u}|\boldsymbol{y})+(1-\rho\lambda)\hat{\mathbf{E}}\boldsymbol{y}\boldsymbol{x}\log P(Y|X))} \\
&\doteq \max_{\hat{Q}_{\boldsymbol{x},\boldsymbol{u}|\boldsymbol{y}}} e^{n(\hat{\mathbf{E}}\boldsymbol{u}\boldsymbol{x}\log P(U,X)+\hat{H}(\boldsymbol{x},\boldsymbol{u}|\boldsymbol{y})+(1-\rho\lambda)\hat{\mathbf{E}}\boldsymbol{y}\boldsymbol{x}\log P(Y|X))} \qquad (75)
\end{aligned}
$$

The last exponent is $E_3(Q_Y, R_y, R_{yz}, \rho, \lambda)$ of Theorem 2 as $n \to \infty$. The derivation of the exponent of the second expectation is quite similar to the steps of following (29) in the weak decoder exponent. We therefore only outline the derivation here. For the second expectation we have:

$$
\begin{aligned}
\boldsymbol{E}\left(\sum_{i'\neq i} P_1(\boldsymbol{y}|\boldsymbol{X}_{m,i'})^{\lambda}\right)^{\rho} &= \boldsymbol{E}_u\boldsymbol{E}_{x|u}\left(\sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{u}\boldsymbol{y}}} N_{y,m}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{u}\boldsymbol{y}})e^{n\lambda\hat{\mathbf{E}}\boldsymbol{y}\boldsymbol{x}\log P(Y|X)}\right)^{\rho} \\
&\doteq \boldsymbol{E}_u\left(\sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{u}\boldsymbol{y}}} \boldsymbol{E}_{x|u}N^{\rho}_{y,m}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{u}\boldsymbol{y}})e^{n\rho\lambda\hat{\mathbf{E}}\boldsymbol{y}\boldsymbol{x}\log P(Y|X)}\right) \qquad (76)
\end{aligned}
$$

As in the case of the weak decoder we define:

$$
\mathcal{G}(R_y, Q_{U|Y}) = \left\{Q_{X|U,Y} : R_y + \boldsymbol{E}_Q \log P(X|U) + H_Q(X|U,Y) > 0\right\} \qquad (77)
$$

and we have

$$
\begin{aligned}
&\boldsymbol{E}_{x|u}N^{\rho}_{y,m}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}}) \\
&\quad\doteq \begin{cases} e^{n\rho(R_y+\hat{\mathbf{E}}\log P(X|U)+\hat{H}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{u}))} & \hat{Q}_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{u}} \in \mathcal{G}(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{y}}) \\ e^{n(R_y+\hat{\mathbf{E}}\log P(X|U)+\hat{H}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{u}))} & \hat{Q}_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{u}} \in \mathcal{G}^c(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{y}}) \end{cases} \qquad (78)
\end{aligned}
$$

Now define:

$$\gamma(Q_{U|Y}) \stackrel{\triangle}{=} \rho\left(R_y + \max_{Q_{X|U,Y}\in\mathcal{G}(R_y,Q_{U|Y})} (\boldsymbol{E}_Q \log P(X|U) + H_Q(X|Y,U) + \lambda \boldsymbol{E}_Q \log P(Y|X))\right) \tag{79}$$

where, as described in Section 2, $P_1(\cdot|\cdot)$ is the channel to the strong user. Similarly, define:

$$\zeta(Q_{U|Y}) \stackrel{\triangle}{=} R_y + \max_{Q_{X|U,Y}\in\mathcal{G}^c(R_y,Q_{U|Y})} [\boldsymbol{E}_Q \log P(X|U) + H_Q(X|U,Y) + (\rho\lambda)\boldsymbol{E}_Q \log P(Y|X)] \tag{80}$$

We now continue (76) by splitting the sum over all $\hat{Q}_{\boldsymbol{x}|\boldsymbol{uy}}$ into $\hat{Q}_{\boldsymbol{x}|\boldsymbol{uy}} \in \mathcal{G}(R_y, Q_{U|Y})$ and $\hat{Q}_{\boldsymbol{x}|\boldsymbol{uy}} \in \mathcal{G}^c(R_y, Q_{U|Y})$.

$$\begin{aligned}\boldsymbol{E}\left(\sum_{i'\neq i} P_1(\boldsymbol{y}|\boldsymbol{X}_{m,i'})^\lambda\right)^\rho &\doteq E_u\left[e^{n\gamma(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}})} + e^{n\zeta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}})}\right] \\ &\doteq \max_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}}} \Pr\left(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}}\right)\left[e^{n\gamma(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}})} + e^{n\zeta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}})}\right]\end{aligned} \tag{81}$$

We begin with the evaluation of $\gamma(Q_{\boldsymbol{u}|\boldsymbol{y}})$. The unconstrained achiever in (79) is:

$$Q_\lambda(x|u,y) = \frac{P(x|u)P^\lambda(y|x)}{\sum_{x'} P(x'|u)P_3^\lambda(y|x')}.$$

If $Q_\lambda(x|u,y) \in \mathcal{G}(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{y}})$ than we can calculate $\gamma(Q_{U|Y})$ with it. If $Q_\lambda(x|u,y) \in \mathcal{G}^c(R_y, Q_{U|Y})$ Since $Q_{\lambda=0}(x|u,y) \in \mathcal{G}(R_y, Q_{U|Y})$, we know that $\mathcal{G}(R_y, Q_{U|Y})$ is not empty, and there is a $\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}}) \in (0,\lambda)$ for which $Q_{\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}})}$ is on the boundary of $\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}}$. As before, our constrained optimizer is on the boundary. So we have for $\gamma(Q_{\boldsymbol{u}|\boldsymbol{y}})$:

$$\gamma(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}}) = \begin{cases} \rho\left(R_y + \boldsymbol{E}_{Q_\lambda} \log P(X|U) + H_{Q_\lambda}(X|Y,U) + \lambda \boldsymbol{E}_{Q_\lambda} \log P(Y|X)\right) & Q_\lambda(x|u,y) \in \mathcal{G}(R_y, Q_{\boldsymbol{u}|\boldsymbol{y}}) \\ \rho\lambda \boldsymbol{E}_{Q_{\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}})}} \log P(Y|X) & Q_\lambda(x|u,y) \in \mathcal{G}^c(R_y, Q_{\boldsymbol{u}|\boldsymbol{y}}) \end{cases} \tag{82}$$

By the same arguments:

$$\zeta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}}) = \begin{cases} \rho\lambda \boldsymbol{E}_{Q_{\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}})}} \log P(Y|X) & Q_{\rho\lambda}(x|u,y) \in \mathcal{G}(R_y, Q_{\boldsymbol{u}|\boldsymbol{y}}) \\ R_y + \boldsymbol{E}_{Q_{\rho\lambda}} \log P(X|U) + H_{Q_{\rho\lambda}}(X|Y,U) + \rho\lambda \boldsymbol{E}_{Q_{\rho\lambda}} \log P(Y|X) & Q_{\rho\lambda}(x|u,y) \in \mathcal{G}^c(R_y, Q_{\boldsymbol{u}|\boldsymbol{y}}) \end{cases} \tag{83}$$

Letting $E_{\gamma\zeta}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}})$ be the dominant term between $\gamma(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}})$ and $\zeta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}})$,the second expectation of $P_{E_{y1}}$ is:

$$\boldsymbol{E}\left(\sum_{i'\neq i} P_1(\boldsymbol{y}|\boldsymbol{X}_{m,i'})^{\lambda}\right)^{\rho} \doteq e^{n\max_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}}}(E_{\gamma\zeta}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{y}})+\hat{\mathbf{E}}_{\boldsymbol{u}}\log P(U)+\hat{H}(\boldsymbol{u}|\boldsymbol{y}))} \tag{84}$$

the last exponent is $E_4(Q_Y, R_y, R_{yz}, \rho, \lambda)$ of Theorem 2 as $n \to \infty$.

We now proceed to the evaluation of:

$$\boldsymbol{E}P_{E_{y2}} = \sum_{\boldsymbol{y}} \boldsymbol{E}P_1(\boldsymbol{y}|\boldsymbol{x}_{m,i})^{1-\lambda\rho}\boldsymbol{E}\left[\sum_{m'\neq m}\sum_{i'=1}^{M_y} P_1(\boldsymbol{y}|\boldsymbol{x}_{m',i'})^{\lambda}\right]^{\rho} \tag{85}$$

The fist expectation is the same as before. For the second expectation, following the same steps as is (42) we have

$$\boldsymbol{E}\left[\sum_{m'\neq m}\sum_{i'=1}^{M_y} P_1(\boldsymbol{y}|\boldsymbol{x}_{m',i'})^{\lambda}\right]^{\rho} \doteq \sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}} e^{n\lambda\rho\hat{\mathbf{E}}_{\boldsymbol{y}\boldsymbol{x}}\log P_1(Y|X)}\boldsymbol{E}\left[\sum_{m'\neq m} N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}})\right]^{\rho} \tag{86}$$

and by the arguments that led to (43) we have:

$$\boldsymbol{E}\left[\sum_{m'\neq m} N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}})\right]^{\rho} \doteq \sum_{A\geq 0}^{R_{yz}} e^{n\rho A}\boldsymbol{E}\left[\sum_{m'\neq m} I_{m'}(A)\right]^{\rho} \tag{87}$$

where, here, $I_{m'}(A) \triangleq \mathcal{I}\left(N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}) \doteq e^{nA}\right)$ (as before, we omit the dependence on $\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}$ to simplify notation). The only difference between (87) and (43) is that here only ρ multiplies A in the exponent whereas in (43) we had $\rho\lambda$ multiplying A. This fact will change the final result, however, the evaluation of $\boldsymbol{E}\left[\sum_{m'\neq m} I_{m'}(A)\right]^{\rho}$ is identical to the weak decoder case by replacing the role of $\boldsymbol{z}$ with $\boldsymbol{y}$ and $P_3(Z|X)$ with $P_1(Y|X)$. We therefore have:

$$\boldsymbol{E}\left[\sum_{m'\neq m} N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}})\right]^{\rho} \doteq e^{nE(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}})} \tag{88}$$

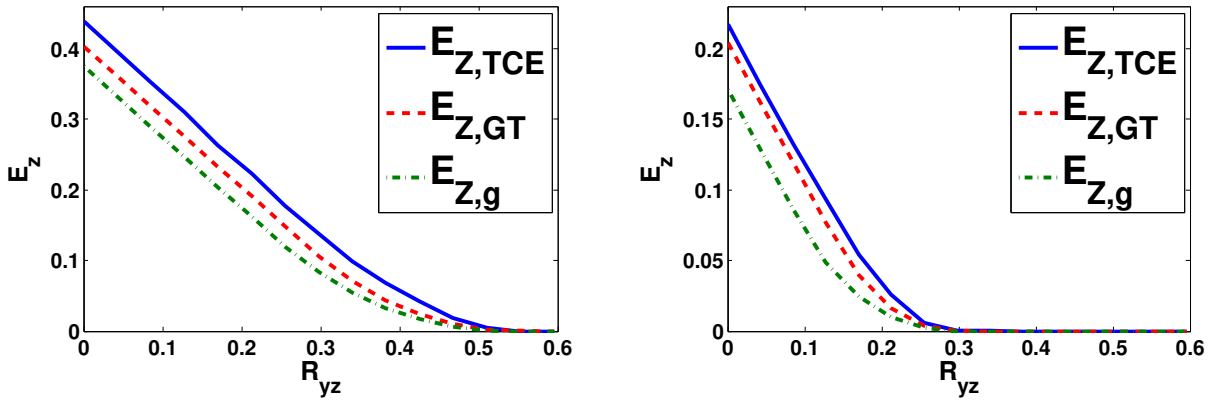
and for the second expectation we have:

$$\boldsymbol{E}\left[\sum_{m'\neq m}\sum_{i'=1}^{M_y} P_1(\boldsymbol{y}|\boldsymbol{x}_{m',i'})^{\lambda}\right]^{\rho} \doteq e^{n\max_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}}}\left[\lambda\rho\hat{\mathbf{E}}_{\boldsymbol{y}\boldsymbol{x}}\log P_1(Y|X)+E(\hat{Q}_{\boldsymbol{x}|\boldsymbol{y}})\right]} \tag{89}$$

the last exponent is $E_5(Q_Y, R_y, R_{yz}, \rho, \lambda)$ of Theorem 2 as $n \to infty$ Taking the maximum of and and using we arrive at $E_{y,2}(R_{yz}, R_y)$ after optimizing over the free parameters.

5.4 Numerical Results

In this subsection, we revisit the same setup as in Section 4.3. We show some numerical results of the error exponents obtained by the type class enumerators approach and compare them to the exponents of our Gallager type approach and to Gallager's results [4]. Unlike the calculation of the numerical results of Section 4, which, after setting $\alpha = \mu$ had a straightforward implementation and reasonable computation time, here the calculation is much more complex. For every ρ, λ searched, we need to optimize over $Q(u|z), Q(x|z)$ in the intermediate steps 71,72 and finally over $Q(z)$. In Fig. 5, we show the best attainable $E_z(R_y, R_{yz})$ (maximized over β) for two values of R_y, compared to results in [4] and of Section 4. In both cases, although we confined ρ to $[0, 1]$ in order to limit the computation time, the new exponents are better. We used E_y that was derived in Section 4 and allowed it to be arbitrarily small (yet positive), thus complying with the definition of an attainable exponent for the weak user.



(a) (b)

Figure 5: E_z for (a) $R_y = 0.05$[nats] and (b) $R_y = 0.3$[nats]. $E_{Z,g}$ is Gallager's 74 result, $E_{Z,GT}$ is Gallager-type approach exponent and $E_{Z,TCE}$ is the type class enumerators approach result.

In both plots of Fig. 5, the exponent becomes zero when the pair (R_y, R_{yz}) is outside the capacity region. The improvement gained by the type class enumerators approach is more substantial when R_y is small. As discussed in [18, Appendix E], when the number of elements in the sum of likelihoods (28) is large enough, Jensen's inequality becomes tighter

and the results of the Gallager-type approach will be closer to the tight approach results.

A Appendix

A.1 proof of $\lambda = \frac{1}{1+\rho}$ when $\alpha = \mu$

It will be shown bellow that

$$\forall \lambda: \quad E_0(\rho, \frac{1}{1+\rho}, \alpha, \alpha) \geq E_0(\rho, \lambda, \alpha, \alpha)$$

where $E_0(\rho, \lambda, \alpha, \alpha)$ was defined in (1). We use the following variant of Hölder's inequality [15, p. 523]: Let a_i, b_i, P_i be non negative numbers defined over a finite set of i with $\sum_i P_i = 1$ and $0 < \gamma < 1$

$$\sum_i P_i a_i b_i \leq \left(\sum_i P_i a_i^{\frac{1}{\gamma}} \right)^{\gamma} \left[\sum_i P_i b_i^{\frac{1}{1-\gamma}} \right]^{1-\gamma} \tag{90}$$

We have for the weak decoder:

$$E(R_1, R_2) = \max_{0 \leq \rho \leq 1} \max_{0 \leq \lambda \leq \mu \leq 1} \max_{1-\rho\lambda \leq \alpha \leq 1} \{E_0(\rho, \lambda, \alpha, \mu) - (\alpha + \rho\mu - 1)R_1 - \rho R_2\}$$

where

$$E_0(\rho, \lambda, \alpha, \mu) = -\log \left\{ \sum_z \left[\sum_u Q_1(u) \left(\sum_x Q_2(x|u) P_3(z|x)^{(1-\rho\lambda)/\alpha} \right)^{\alpha} \right] \times \left[\sum_{u'} Q_1(u') \left(\sum_{x'} Q_2(x'|u') P_3(z|x')^{\lambda/\mu} \right)^{\mu} \right]^{\rho} \right\}.$$

Substituting $\alpha = \mu$, $(max(\lambda, 1-\lambda\rho) \leq \alpha \leq 1)$ we have for E_0:

$$E_0(\rho, \lambda, \alpha, \alpha) = -\log \left\{ \sum_z \left[\sum_u Q_1(u) \left(\sum_x Q_2(x|u) P_3(z|x)^{(1-\rho\lambda)/\alpha} \right)^{\alpha} \right] \times \left[\sum_{u'} Q_1(u') \left(\sum_{x'} Q_2(x'|u') P_3(z|x')^{\lambda/\alpha} \right)^{\alpha} \right]^{\rho} \right\}. \tag{91}$$

Finally,

$$E_0(\rho, \frac{1}{1+\rho}, \alpha, \alpha) = -\log \sum_z \left\{ \sum_u Q_1(u) \left(\sum_x Q_2(x|u) P_3(z|x)^{1/\alpha(1+\rho)} \right)^{\alpha} \right\}^{1+\rho}$$

The proof holds for $1 \geq \rho > 0$. Since when $\rho = 0$ (note that in this case $\alpha = 1$) we have for all λ: $E_0(\rho = 0, \lambda, 1, 1) = 0$, this is sufficient for our case.

Proof. Let us observe the inner term of $E_0(\rho, \frac{1}{1+\rho}, \alpha, \alpha)$:

$$\left\{ \sum_u Q_1(u) \left(\sum_x Q_2(x|u) P_3(z|x)^{1/\alpha(1+\rho)} \right)^{\alpha} \right\}^{1+\rho} \tag{92}$$

It is sufficient to show, that for every z, this term lower bounds the same term with λ instead of $\frac{1}{1+\rho}$ (as in (91)).
To Start, we use (90) with the following assignments: $P_i = Q_2(x|u), a_i = P_3(z|x)^{\frac{1-\lambda\rho}{\alpha(1+\rho)}}, b_i = P_3(z|x)^{\frac{\lambda\rho}{\alpha(1+\rho)}}$. Applying this we have for $0 \leq \delta \leq 1$:

$$\begin{aligned}
&\left\{ \sum_u Q_1(u) \left(\sum_x Q_2(x|u) P_3(z|x)^{1/\alpha(1+\rho)} \right)^{\alpha} \right\}^{1+\rho} \leq \\
&\leq \left\{ \sum_u Q_1(u) \left[\left(\sum_x Q_2(x|u) P_3(z|x)^{\frac{1-\lambda\rho}{\delta\alpha(1+\rho)}} \right)^{\delta} \left(\sum_x Q_2(x|u) P_3(z|x)^{\frac{\lambda\rho}{(1-\delta)\alpha(1+\rho)}} \right)^{1-\delta} \right]^{\alpha} \right\}^{1+\rho}.
\end{aligned} \tag{93}$$

At this point we use (90) again over the whole term with the following assignments:

$$\begin{aligned}
P_i &= Q(u) \\
a_i &= \left(\sum_x Q_2(x|u) P_3(z|x)^{\frac{1-\lambda\rho}{\delta\alpha(1+\rho)}} \right)^{\delta\alpha} \\
b_i &= \left(\sum_x Q_2(x|u) P_3(z|x)^{\frac{\lambda\rho}{(1-\delta)\alpha(1+\rho)}} \right)^{\alpha(1-\delta)}
\end{aligned}$$

Continuing from(93):

$$\leq \left\{ \begin{array}{l} \left[\sum_u Q_1(u) \left(\sum_x Q_2(x|u) P_3(z|x)^{\frac{1-\lambda\rho}{\delta\alpha(1+\rho)}} \right)^{\delta\alpha/\gamma} \right]^{\gamma} \times \\ \left[\sum_u Q_1(u) \left(\sum_x Q_2(x|u) P_3(z|x)^{\frac{\lambda\rho}{(1-\delta)\alpha(1+\rho)}} \right)^{\frac{\alpha(1-\delta)}{(1-\gamma)}} \right]^{1-\gamma} \end{array} \right\}^{1+\rho} \qquad 0 \leq \gamma \leq 1.$$

Assigning $\gamma = \delta = \frac{1}{1+\rho}$ we have:

$$\left\{\sum_u Q_1(u)\left(\sum_x Q_2(x|u)P_3(z|x)^{1/\alpha(1+\rho)}\right)^{\alpha}\right\}^{1+\rho} \leq$$
$$\left[\sum_u Q_1(u)\left(\sum_x Q_2(x|u)P_3(z|x)^{(1-\rho\lambda)/\alpha}\right)^{\alpha}\right]\left[\sum_{u'} Q_1(u')\left(\sum_{x'} Q_2(x'|u')P_3(z|x')^{\lambda/\alpha}\right)^{\alpha}\right]^{\rho}$$

Note that the last term is equivalent to (92) when $\lambda = \frac{1}{1+\rho}$ and greater or equal for every other value of λ. Since this is true for every z the proof is completed. □

A.2 The Existence of $\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$

We need to show that for $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}$, there exist a $\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ such that, when $P(x|u,z) \in \mathcal{G}^c(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$, the partition function of $\mathcal{G}^c(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ is zero. Namely:

$$R_y + \boldsymbol{E}_Q \log P(X|U) + H_Q(X|Z,U) = 0 \tag{94}$$

where the above entropy and expectation are calculated with respect to

$$Q(x,u,z) = Q^*(x|u,z)\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}(u,z)\hat{Q}_{\boldsymbol{z}}(z)$$

($Q^*(x|u,z)$ is defined in (34)).
Denote $C(\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}), u, z) = \sum_x P(x|u)P_3^{\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}(z|x)$ and define

$$\begin{aligned} g(\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) &\triangleq R_y + \boldsymbol{E}_Q \log P(X|U) + H_Q(X|Z,U) \\ &= R_y + \boldsymbol{E}_Q \log \frac{P(X|U)C(\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}), u, z)}{P(X|U)P_3^{\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}(Z|X)} \\ &= R_y + \delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})\boldsymbol{E}_Q \log \frac{1}{P(Z|X)} + \boldsymbol{E}_{\boldsymbol{uz}} \log C(\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}, u, z) \end{aligned} \tag{95}$$

For $P(x|u,z) \in \mathcal{G}^c(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$, $g(1) \leq 0$ and since $R_y \geq 0$, $g(0) \geq 0$. Therefore, because of the continuity of $g(\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$, we conclude that there exist $\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \in [0,1)$ such that $g(\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})) = 0$. It can be shown that $g(\delta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}))$ is non increasing for $\delta > 0$.

A.3 The Behavior of $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$

$$M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} = \sum_{i=1}^{e^{nR_{yz}}} \mathcal{I}(\boldsymbol{u}_i \in T_{\boldsymbol{u}|\boldsymbol{z}}) \tag{96}$$

The probability that a cloud center $\boldsymbol{u}_m$, drawn with $P(u^n) = \prod_{i=1}^n P(u_i)$ will belong to $T_{\boldsymbol{u}|\boldsymbol{z}}$ is (exponentially) $e^{n(\hat{\mathbf{E}}_{\boldsymbol{u}} \log P(U) + \hat{H}(\boldsymbol{u}|\boldsymbol{z})}$. Using $D(a||b) > \left(\ln \frac{a}{b} - 1\right)$ ([10, Appendix]) and the Chernoff bound we have:

$$\begin{aligned}
&\Pr(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \geq e^{n\cdot a}) \leq \exp\left\{-ne^{n\cdot a}\left[a - R_{yz} - \hat{H}(\boldsymbol{u}|\boldsymbol{z}) - \hat{\mathbf{E}}_{\boldsymbol{u}} \log P(U)\right]\right\} \;\; a \geq R_{yz} + \hat{H}(\boldsymbol{u}|\boldsymbol{z}) + \hat{\mathbf{E}}_{\boldsymbol{u}} \log P(U) \\
&\Pr(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \leq e^{n\cdot a}) \leq \exp\left\{ne^{n\cdot a}\left[a - R_{yz} - \hat{H}(\boldsymbol{u}|\boldsymbol{z}) - \hat{\mathbf{E}}_{\boldsymbol{u}} \log P(U)\right]\right\} \;\; a \leq R_{yz} + \hat{H}(\boldsymbol{u}|\boldsymbol{z}) + \hat{\mathbf{E}}_{\boldsymbol{u}} \log P(U)
\end{aligned} \tag{97}$$

Therefore, for $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z(R_{yz})$, $\epsilon > 0$:

$$\begin{aligned}
\Pr(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \doteq e^{n(R_{yz} + \hat{H}(\boldsymbol{u}|\boldsymbol{z}) + \hat{\mathbf{E}}_{\boldsymbol{u}} \log P(U))}) &= 1 - Pr(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \geq e^{n(R_{yz} + \hat{H}(u|z) + \hat{\mathbf{E}} \log P(u) + \epsilon)}) \\
&\quad - Pr(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \leq e^{n(R_{yz} + \hat{H}(u|z) + \hat{\mathbf{E}} \log P(u) - \epsilon)}) \\
&\geq 1 - 2e^{-n\epsilon\epsilon e^{n(R_{yz} + \hat{H}(u|z) + \hat{\mathbf{E}} \log P(u) - \epsilon)}}
\end{aligned} \tag{98}$$

And thus, for $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z(R_{yz})$, $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$ converges to its expectation double exponentially fast. It is obvious from (97) that when $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z^c(R_{yz})$, we wont find an exponential number of cloud centers of this type. Furthermore, the dominant term in $\boldsymbol{E}M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$ will be $1 \cdot \Pr(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} = 1)$. We now show the exponential behavior of $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$ when $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_z^c(R_{yz})$

$$\begin{aligned}
\Pr(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} = 1) &= e^{nR_{yz}} e^{n(\hat{H}(\boldsymbol{u}|\boldsymbol{z}) + \hat{\mathbf{E}}_{\boldsymbol{u}} \log P(U))} (1 - e^{n(\hat{H}(\boldsymbol{u}|\boldsymbol{z}) + \hat{\mathbf{E}}_{\boldsymbol{u}} \log P(U))})^{e^{nR_{yz}} - 1} \\
&\leq e^{nR_{yz}} e^{n(\hat{H}(\boldsymbol{u}|\boldsymbol{z}) + \hat{\mathbf{E}}_{\boldsymbol{u}} \log P(u))} \\
&= e^{n\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}
\end{aligned} \tag{99}$$

$$
\begin{aligned}
\Pr(M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}=1) &= e^{nR_{yz}} e^{n(\hat{H}(\boldsymbol{u}|\boldsymbol{z})+\hat{\mathbf{E}}_{\boldsymbol{u}}\log P(U))}\big(1-e^{n(\hat{H}(\boldsymbol{u}|\boldsymbol{z})+\hat{\mathbf{E}}_{\boldsymbol{u}}\log P(U))}\big)^{e^{nR_{yz}}-1} \\
&\doteq e^{n\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}\big(1-e^{n(\hat{H}(\boldsymbol{u}|\boldsymbol{z})+\hat{\mathbf{E}}_{\boldsymbol{u}}\log P(U))}\big)^{e^{nR_{yz}}} \\
&= e^{n\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}\exp\left[\log\big(1-e^{n(\hat{H}(\boldsymbol{u}|\boldsymbol{z})+\hat{\mathbf{E}}_{\boldsymbol{u}}\log P(U))}\big)e^{nR_{yz}}\right] \\
&\geq e^{n\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}\exp\left[e^{nR_{yz}}\frac{-e^{n(\hat{H}(\boldsymbol{u}|\boldsymbol{z})+\hat{\mathbf{E}}_{\boldsymbol{u}}\log P(U))}}{1-e^{n(\hat{H}(\boldsymbol{u}|\boldsymbol{z})+\hat{\mathbf{E}}_{\boldsymbol{u}}\log P(U))}}\right] \qquad (100) \\
&= e^{n\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}\exp\left[\frac{e^{n\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}}{1-e^{n(\hat{H}(\boldsymbol{u}|\boldsymbol{z})+\hat{\mathbf{E}}_{\boldsymbol{u}}\log P(U))}}\right] \\
&\to e^{n\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})} \qquad (101)
\end{aligned}
$$

where in (100), we used $\log(1+x)\geq\frac{x}{1+x}$ and the last line is true since $e^{n\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}\to 0$ when $n\to\infty$ for $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}\in\mathcal{G}_z^c(R_{yz})$. To conclude, we have:

$$
\Pr(M_{T_{\boldsymbol{u}|\boldsymbol{z}}}=1)\doteq e^{n\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})} \qquad (102)
$$

A.4 Deriving $P_A(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$

For a given $\boldsymbol{u}^*$, the probability of drawing $\boldsymbol{x}$ with $P(\boldsymbol{x}|\boldsymbol{u})$ which will belong to $T_{\boldsymbol{x}|\boldsymbol{z}}$ is

$$
\begin{aligned}
\sum_{x\in T_{x|z}}P(\boldsymbol{x}|\boldsymbol{u}^*) &= \sum_{x\in T_{x|z}}\prod_{i=1}^{n}P(x_i|u_i^*) \\
&= \sum_{T_{x|z,u^*}}|T_{x|z,u^*}|\prod_{a\in\mathcal{U},b\in\mathcal{X},c\in\mathcal{Z}}P(b|a)^{n\hat{P}(a,b,c)} \qquad (103)
\end{aligned}
$$

where $\hat{P}(a,b,c)$ is the joint empirical distribution of the triplet $a\in\mathcal{U},b\in\mathcal{X},c\in\mathcal{Z}$. Note that for different $\boldsymbol{x}\in T_{x|z}$, $\hat{P}(a,b,c)$ have different values. Exponentially, the behavior will be according to the maximal element. Namely:

$$
\doteq e^{n\cdot\max_{T_{x|z,u}|T_{x|z}}\left\{\hat{\mathbf{E}}\log P(x|u)+\hat{H}(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u})\right\}} \qquad (104)
$$

The last expression remains true for all permutations of $\boldsymbol{u}^*$ which belong to $T_{\boldsymbol{u}^*|\boldsymbol{z}}$. This is because we can apply the same permutation to the $\boldsymbol{x}$ vector and get the same value in the exponent. This value will be the maximizer since the range of the maximization remains

constant while $\boldsymbol{u}$ belongs to the same $T_{u^*|z}$. for a given $\boldsymbol{u} \in T_{\boldsymbol{u}^*|\boldsymbol{z}}$ (if there is such a $\boldsymbol{u}$ in our random codebook) we draw e^{nR_y} $\boldsymbol{x}$ series independently according to $\prod_{i=1}^{n} P(x_i|u_i)$. Therefore, the average number of $\boldsymbol{x}$ that will belong to $T_{\boldsymbol{x}|\boldsymbol{z}}$ when $\boldsymbol{u}$ belongs to $T_{\boldsymbol{u}|\boldsymbol{z}}$ is

$$e^{n\left(R_y+\max_{\hat{Q}_{x|z,u}|\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}\left\{\hat{\mathbf{E}}_{\boldsymbol{x}\boldsymbol{u}}\log P(X|U)+\hat{H}(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u})\right\}\right)} \triangleq e^{nN(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)} \tag{105}$$

Since we are evaluating the probability of drawing an exponential number of $\boldsymbol{x}$ which will belong to $T_{\boldsymbol{x}|\boldsymbol{z}}$ we are only interested in the case where the last exponent is positive. By the same arguments in Section A.3, when $N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y) > 0$ the number of $\{\boldsymbol{x}_m\}$ which will belong to $T_{\boldsymbol{x}|\boldsymbol{z}}$ concentrates double exponentially fast around the expectation (105). Therefore, for $N_(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y) > 0$, $\epsilon > 0$:

$$\begin{aligned} &Pr\left\{\mathbf{1}\left(N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}) \doteq e^{nN(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)}\right)=1\right\} \\ &\quad \geq 1-2e^{-n\epsilon e^{n(N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)-\epsilon)}} \end{aligned} \tag{106}$$

To conclude, $P_{A,T_{\boldsymbol{u}|\boldsymbol{z}}}$ either vanishes double exponentially fast if $A \neq N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)$ or converges double exponentially fast to 1 if $A = N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)$.

When the exponent in (105) is negative, for every $A > 0$ $P_A(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ vanishes double exponentially fast. However, for $A = 0$, by the same arguments as in section A.3 we show that

$$Pr\left\{N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}) \doteq e^{n0}\right\} = Pr\left\{1 \leq N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}) < e^{n\epsilon}\right\} \doteq \Pr\left\{N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}) = 1\right\} \tag{107}$$

and

$$\Pr\left\{N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}) = 1\right\} \doteq e^{nN(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)}. \tag{108}$$

References

[1] T. M. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 2–14, January 1972.

[2] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Transactions on Information Theory*, vol. 23, no. 1, pp. 60–64, January 1977.

[3] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Transactions on Information Theory*, vol. 19, no. 2, pp. 197–207, March 1973.

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

[5] L. Weng, S. S. Pradhan, and A. Anastasopoulos, "Error exponent regions for gaussian broadcast and multiple-access channels," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 2919–2942, July 2008.

[6] J. Körner and A. Sgarro, "Universally attainable error exponents for broadcast channels with degraded message sets," *IEEE Transactions on Information Theory*, vol. 26, no. 6, pp. 670–679, November 1980.

[7] G. D. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Transactions on Information Theory*, vol. 14, no. 2, pp. 206–220, March 1968.

[8] M. Mezard and A. Montanari, *Constraint Satisfaction Networks in Physics and Computation*. Oxford University Press, 2009.

[9] N. Merhav, "Relations between random coding exponents and the statistical physics of random codes," *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 83–92, January 2009.

[10] ——, "Error exponents of erasure/list decoding revisited via moments of distance enumerators," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4439–4447, October 2008.

[11] R. Etkin, N. Merhav, and E. Ordentlich, "Error exponents of optimum decoding for the interference channel," in *Proceeding of the International Symposium on Information Theory*, 2008, pp. 1523–1527.

[12] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Transactions on Information Theory*, vol. 25, no. 3, pp. 306– 311, May 1979.

[13] A. El Gammal and E. C. van der Meulen, "A proof of marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 120–122, January 1981.

[14] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. McGraw-Hill, 1979.

[15] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.

[17] I. Csiszar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.

[18] Y. Kaspi, "Error exponents for broadcast channels with degraded message sets," Master's thesis, Technion - Isreal Institute of Technology, Haifa, Israel, April 2009.