Error Exponents of Optimum Decoding for the Degraded Broadcast Channel Using Moments of Type Class Enumerators

Yonatan Kaspi and Neri Merhav
Department of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel
Email: {kaspi@tx, merhav@ee}.technion.ac.il

***Abstract*—THIS PAPER IS ELIGIBLE FOR THE STUDENT PAPER AWARD. The analysis of random coding error exponents pertaining to optimal decoding in a degraded broadcast with degraded message sets is revisited. Instead of using Jensens inequality as well as some other inequalities in the derivation, we demonstrate that, after an initial step, an exponentially tight analysis can be carried out by assessing the relevant moments of a certain type class enumerator.**

I. Introduction

In a 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 for a BC with a degraded message set was found in [2]. The coding theorem for degraded broadcast channels was given by Bergmans [3] and the converse was given by Gallager [4]. Bergmans suggested the use of a random hierarchical 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. The tradeoff between the exponents is controlled by the random coding distributions.

Works on error exponents for general degraded broadcast channels include [4], [5] and [8]. Both [4] and [5] used the coding scheme of [3] but did not use optimal decoding. In [5], universally attainable error exponents were given for a suboptimal decoder. In [4] a direct channel from the cloud center to the weak user was defined and the error exponent was calculated for this channel. In [8] we derived the error exponents for both the weak and the strong decoders while using optimal maximum likelihood decoders. Numerical results for the binary symmetric broadcast channel demonstrated that by using optimal decoders, we can achieve better exponents compared to the results in [4]. Another effect that was shown in [8] is that when we required that one of the exponents will be greater than a given threshold, discontinuities in the other exponent may arise.

In this work, we revisit the setting of [8] using a substantially different analysis technique to derive the error exponent for the weak decoder of a degraded BC with degraded message sets that pertains to optimum decoding. Unlike [4], [8] , where Jensen’s inequality, as well as other inequalities were used, which possibly risked the tightness of the obtained bounds, our technique in this paper is guaranteed to be exponentially tight in all steps from the very beginning. The underlying ideas behind this technique are inspired from the statistical mechanical point of view on random code ensembles [10]. These analysis tools are applicable to other problem settings as well, e.g. [7] and [11] where they lead to tighter bounds than the bounds obtained by other methods that were previously used. The goal of this work is not only to improve the error exponent of the case at hand, but also to demonstrate the use of the analysis technique we offer here. With this in mind, since the derivation of the weak decoder error exponent, using our technique, is much more involved than the derivation of the strong decoder exponent, in this work we present only the derivation of the weak decoder exponent. The derivation of the other exponent will be presented in the full paper [9].

The rest of this work is structured as follows: Section II gives the formal setting and notation. In section III, we will give the main result of the paper. Section IV will outline the proof of the main result. Finally. Numerical results for the degraded BSC are given in section V.

II. Preliminaries

We begin with notation. 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 Q, we subscript it by Q. i.e $\boldsymbol{E}_Q\{\cdot\}$.

We consider a memoryless degraded broadcast channel (MDBC) 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 $P(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{x}) = \prod_{t=1}^{n} P_1(y_t|x_t)P_2(z_t|y_t), \quad (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$. 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 next a random selection of an 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 $P(\boldsymbol{u}) = \prod_{t=1}^{n} P(u_t)$, and 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 $P_3(\boldsymbol{z}|\boldsymbol{x}) = \prod_{t=1} P_3(z_t|x_t) = \prod_{t=1} \sum_y P_2(z_t|y)P_1(y|x_t)$.

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 $\overline{P_E^z} = Pr\{\tilde{m}(\boldsymbol{z}) \neq m\}$. 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 $P(u), P(x|u)$ such that the random coding exponents satisfy $E_y \leq \lim_{n\to\infty} -\frac{1}{n}\log \overline{P_E^y}$ and $E_z \leq \lim_{n\to\infty} -\frac{1}{n}\log \overline{P_E^z}$, 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.

III. Main Result

Let (X, U, Z) be a triplet of random variables, taking values in $\mathcal{X} \times \mathcal{U} \times \mathcal{Z}$, and being governed by a generic joint distribution $Q_{XUZ} = \{Q_{XUZ}(x, u, z),\ x \in \mathcal{X},\ u \in \mathcal{U},\ z \in \mathcal{Z}\}$. Let us denote the various marginals and conditional distributions derived from Q_{XUZ}, 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_{XUZ}, 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 $Q = Q_{XUZ}$. In the following description, we allow various joint distributions $\{Q\}$ to govern (X, U, Z).

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

$$\alpha(Q_{U|Z}) \stackrel{\triangle}{=} (1 - \rho\lambda) \max_{Q_{X|UZ} \in \mathcal{G}(R_y, Q_{U|Z})} [\boldsymbol{E}_Q \log P(X|U) + H_Q(X|U, Z) + \boldsymbol{E}_Q \log P_3(Z|X)] \tag{1}$$

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

$$\beta(Q_{U|Z}) \stackrel{\triangle}{=} \rho\lambda R_y + \max_{Q_{X|UZ} \in \mathcal{G}(R_y, Q_{U|Z})} [\boldsymbol{E}_Q \log P(X|U) + H_Q(X|U, Z) + (1 - \rho\lambda)\boldsymbol{E}_Q \log P_3(Z|X)] \tag{2}$$

and

$$E_{\alpha\beta}(Q_{U|Z}) = \max\{\alpha(Q_{U|Z}), \beta(Q_{U|Z})\}.$$

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,

$$N(Q_{X|Z}, Q_{U|Z}, R_y) \stackrel{\triangle}{=} R_y + \max_{Q_{X|UZ}} [\boldsymbol{E}_Q \log P(X|U) + H_Q(X|U, Z)], \tag{3}$$

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

$$\mathcal{G}(R_{yz}) \stackrel{\triangle}{=} \{Q_{U|Z} :\ R_{yz} + H_Q(U|Z) + \boldsymbol{E} \log P(U) \geq 0\},$$

$$B(Q_{X|Z}, Q_{U|Z}, R_y) = \rho N(Q_{X|Z}, Q_{U|Z}, R_y) \cdot \lambda^{\mathcal{I}\{N(Q_{X|Z}, Q_{U|Z}, R_y) > 0\}} \tag{4}$$

and

$$C(Q_{X|Z}, Q_{U|Z}, R_y) = N(Q_{X|Z}, Q_{U|Z}, R_y) \cdot (\rho\lambda)^{\mathcal{I}\{N(Q_{X|Z}, Q_{U|Z}, R_y) > 0\}}, \tag{5}$$

We also define

$$E(Q_{X|Z}) \stackrel{\triangle}{=} \max\left\{ \max_{Q_{U|Z} \in \mathcal{G}(R_y)} [B(Q_{X|Z}, Q_{U|Z}, R_y) + \rho\bar{m}(Q_{U|Z})],\ \max_{Q_{U|Z} \in \mathcal{G}^c(R_y)} [C(Q_{X|Z}, Q_{U|Z}, R_y) + \bar{m}(Q_{U|Z})] \right\},$$

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

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

Finally,

$$E_z(R_{yz}, R_y) = \max_{\rho \geq 0} \max_{0 \leq \lambda \leq 1/\rho} \min_{Q_Z} [E_A(Q_Z, R_y, R_{yz}, \rho, \lambda) + E_B(Q_Z, R_y, R_{yz}, \rho, \lambda) - H_Q(Z)]. \qquad (6)$$

Theorem 1: For the degraded broadcast channel defined in Section II, $E_z(R_{yz}, R_y)$, as defined in eq. (6), is an attainable exponent for the weak user.

Unlike the result of [8], where the exponent had four free parameters, the new bound has only two free parameters (λ, ρ). Also, it is at least as tight as the exponent of [8] since, as we will see in the next section, its derivation is exponentially tight after the same initial step of [8]. In Section V we show that the new exponent is tighter, at least for the binary symmetric case.

IV. Sketch of Proof

In this section, we outline the main ideas of the proof of Theorem 1. The full proof will appear in [9]. Throughout, we rely on the method of types [12]. We start with notation. 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 will 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 empiric 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 $n \to \infty$.

Applying Gallager's general upper bound [6, 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:

$$\overline{P_E^z} \leq \sum_{\boldsymbol{z}} \boldsymbol{E} \left[\frac{1}{M_y} \sum_{i=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{x}_{m,i}) \right]^{1-\rho\lambda} \times \boldsymbol{E} \left[\sum_{m' \neq m} \left(\frac{1}{M_y} \sum_{j=1}^{M_y} P_3(\boldsymbol{z}|\boldsymbol{x}_{m',j}) \right)^{\lambda} \right]^{\rho} \qquad (7)$$

since messages from different clouds are independent. 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. $E_A(Q_z, R_y, R_{yz}, \rho, \lambda)$ and $E_B(Q_z, R_y, R_{yz}, \rho, \lambda)$ are the exponential rates of the first and second expectations in (7), respectively. Note that (7) is the same initial step as in [8]. After this step, our analysis is exponentially tight, whereas in [8] 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 outline the derivation of $E_A(Q_z, R_y, R_{yz}, \rho, \lambda)$ and $E_B(Q_z, R_y, R_{yz}, \rho, \lambda)$.

A. Deriving $E_A(Q_z, R_y, R_{yz}, \rho, \lambda)$

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 $\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}}$.

$$\begin{aligned}
&\boldsymbol{E} \left[\frac{1}{M_y} \sum_{i=1}^{M_y} P(\boldsymbol{z}|\boldsymbol{x}_{m,i}) \right]^{1-\rho\lambda} \\
&= M_y^{\rho\lambda-1} \boldsymbol{E}_{P_u} \boldsymbol{E}_{P_{x|u}} \left[\sum_{i=1}^{M_y} P(\boldsymbol{z}|\boldsymbol{x}_{mi}) \right]^{1-\rho\lambda} \\
&= M_y^{\rho\lambda-1} \boldsymbol{E}_{P_u} \boldsymbol{E}_{P_{x|u}} \left[\sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}}} N_{z,m}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}}) e^{n\hat{\mathbf{E}}_{\boldsymbol{z}\boldsymbol{x}} \log P_3(Z|X)} \right]^{1-\rho\lambda} \\
&\doteq M_y^{\rho\lambda-1} \boldsymbol{E}_{P_u} \left[\sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}}} \boldsymbol{E}_{P_{x|u}} N_{z,m}^{1-\rho\lambda}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}}) \times \right. \\
&\qquad \left. e^{n(1-\rho\lambda)\hat{\mathbf{E}}_{\boldsymbol{x}\boldsymbol{z}} \log P_3(Z|X)} \right] \qquad (8)
\end{aligned}$$

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 subexponential 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 (exponentially) $e^{n(\hat{\mathbf{E}}_{\boldsymbol{x}\boldsymbol{u}} \log P(X|U) + \hat{H}(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}))}$. Therefore:

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

By the same arguments as in [7, Section IV]:

$$\begin{aligned}
&\boldsymbol{E}_{x|u} N_{z,m}^{1-\rho\lambda}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}}) \doteq \\
&\begin{cases} e^{n(1-\rho\lambda)(R_y + \hat{\mathbf{E}}_{\boldsymbol{x}\boldsymbol{u}} \log P(X|U) + \hat{H}(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}))} & \hat{Q}_{\boldsymbol{x}|\boldsymbol{u},\boldsymbol{z}} \in \mathcal{G}(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \\ e^{n(R_y + \hat{\mathbf{E}}_{\boldsymbol{x}\boldsymbol{u}} \log P(X|U) + \hat{H}(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u}))} & \hat{Q}_{\boldsymbol{x}|\boldsymbol{u},\boldsymbol{z}} \in \mathcal{G}^c(R_y, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \end{cases} \qquad (9)
\end{aligned}$$

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 $N_{z,m}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{u},\boldsymbol{z}})$ will diverge. We continue (8) 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. Using (9) and taking the dominant element of each sum we have that (8) is of the exponential order of

$$\max_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} \Pr(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}|\boldsymbol{z}) (e^{n\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})} + e^{n\beta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})})$$

the last line is true since $\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ and $\beta(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ (cf. (1), (2)) 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 $T_{\boldsymbol{u}|\boldsymbol{z}}$.

We sketch here only the evaluation of $\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$: The unconstrained achiever of (1) is $P(x|z,u)$ which might belong to $\mathcal{G}(R_y,\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ for large enough R_y. When $P(x|z,u)\in\mathcal{G}^c(R_y,\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$, by following the arguments of [7, Section IV], every internal point of $\mathcal{G}(R_y,\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ can be improved by a point on the boundary of $\mathcal{G}(R_y,\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$. Since the constrained maximizer is on the boundary, we use the fact that on that boundary $-R_y = \hat{\mathbf{E}}_{\boldsymbol{x}\boldsymbol{u}}\log P(X|U)+\hat{H}(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{u})$ to get $\alpha(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) = (1-\rho\lambda)(-R_y+\max_{\mathcal{G}(R_y,\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})}\hat{\mathbf{E}}_{\boldsymbol{z}\boldsymbol{x}}\log P_3(Z|X))$. The achieving p.m.f is

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

where $\delta_R(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ is such that $-R_y = \boldsymbol{E}_{Q^*}\log P(X|U)+H_{Q^*}(X|Z,U)$. It can be shown that $\delta_R(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ either exists or $\mathcal{G}^c(R_y,\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ is empty.

B. Deriving $E_B(Q_z,R_y,R_{yz},\rho,\lambda)$

Here, we evaluate the second expectation of (7).

$$\begin{aligned}
&\boldsymbol{E}\left[\sum_{m'\neq m}\left(\frac{1}{M_y}\sum_{j=1}^{M_y}P(\boldsymbol{z}|\boldsymbol{x})\right)^{\lambda}\right]^{\rho}\\
=&M_y^{-\rho\lambda}\boldsymbol{E}\left[\sum_{m'\neq m}\left(\sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}}N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}})e^{n\hat{\mathbf{E}}_{\boldsymbol{z}\boldsymbol{x}}\log P_3(Z|X)}\right)^{\lambda}\right]^{\rho}\\
\doteq&M_y^{-\rho\lambda}\sum_{\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}}e^{n\lambda\rho\hat{\mathbf{E}}_{\boldsymbol{z}\boldsymbol{x}}\log P_3(Z|X)}\boldsymbol{E}\left[\sum_{m'\neq m}N^{\lambda}_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}})\right]^{\rho}
\end{aligned}\tag{10}$$

Here, the enumerators $\{N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}})\}$ are distributed differently for every m' and there is an exponential number of such m'. We divide $[0,R_{yz}]$ into a grid with a sub-exponential number of intervals in n (for example, $d=\frac{R_{yz}}{n}$). Evaluating the last expectation in (10), we have:

$$\begin{aligned}
&\boldsymbol{E}\left[\sum_{m'\neq m}N^{\lambda}_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}})\right]^{\rho}\\
&=\boldsymbol{E}\left[\sum_{A\geq 0}^{R_{yz}}(\text{number of times } N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}})\doteq e^{nA})e^{n\lambda A}\right]^{\rho}\\
&\doteq\sum_{A\geq 0}^{R_{yz}}e^{n\lambda\rho A}\boldsymbol{E}\left[\sum_{m'\neq m}I_{m'}(A)\right]^{\rho}
\end{aligned}\tag{11}$$

where $I_{m'}(A)\triangleq\mathcal{I}\left(N_{z,m'}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}})\doteq e^{nA}\right)$ (we omit 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.

$$\boldsymbol{E}\left[\sum_{m'\neq m}I_{m'}(A)\right]^{\rho}\doteq\sum_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}}I_{m'}(A)\right]^{\rho}\tag{12}$$

Note that the number of terms in the inner summation of (12) is a random variable. Define $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}\triangleq|m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}|$ - 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:

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

The sign of the last exponent determines if we are likely to find an exponential number of codewords of this type. It can be shown [7, Appendix] that when $\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})>0$ (i.e $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}\in\mathcal{G}_{R_{yz}}$), $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$ 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 $P_A(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})\triangleq$ Pr $\{I_{m'}(A)=1\}$ and define:

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

By using the Chernoff bound, we show that if $A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})>0$, $P_{A^*}(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ converges to unity double exponentially fast and vanishes for any other A. When $A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})=0$, $P_0(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})=e^{nN(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)}$. Hence the terms of the summation of (11) are non-zero for $A=A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$.

Continuing (12), there are four cases: $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}\in\mathcal{G}_{R_{yz}}$ 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 $A^*(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})>0$.

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

We use the fact that for this case, $P_A(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})>1-e^{-n\epsilon e^{n\epsilon}}$ for some $\epsilon>0$. For $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}\in\mathcal{G}_{R_{yz}}$ we have for the expectation in (12):

$$\begin{aligned}
&\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}}I_{m'}(A)\right]^{\rho}\leq\\
&e^{n\rho\left(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+\epsilon\right)}\text{Pr}\left\{M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}\leq e^{n\left(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+\epsilon\right)}\right\}+\\
&e^{nR_{yz}}\text{Pr}\left\{M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}\geq e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+\epsilon)}\right\}\\
&\leq e^{n\rho\left(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+\epsilon\right)}+e^{nR_{yz}}e^{-n\epsilon e^{n\left(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+\epsilon\right)}}
\end{aligned}$$

On the other hand:

$$\begin{aligned}
&\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}}I_{m'}(A)\right]^{\rho}\\
&\geq e^{n\rho\left(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})-\epsilon\right)}\text{Pr}\left\{M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}\geq e^{n\left(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})-\epsilon\right)}\right\}
\end{aligned}$$

$$\geq e^{n\rho\left(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})-\epsilon\right)}\left\{1-e^{-n\epsilon e^{n\left(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})-\epsilon\right)}}\right\}$$

Handling $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}^c_{R_{yz}}$ by the same methods as in [7] we have for $A = A^*_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} > 0$, $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}^c_{R_{yz}}$:

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

The case $A^*_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} = 0$

We know that $P_0(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}) \doteq e^{nN(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y)}$ as in this case $N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}, R_y) < 0$. Here, $P_0(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ vanishes exponentially. For $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}_{R_{yz}}$, we have similarly:

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

When $\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}} \in \mathcal{G}^c_{R_{yz}}$, we use the fact that the probability that $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$ is subexponential converges to 1 double exponentially fast.

$$\begin{aligned}\boldsymbol{E}\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0)\right]^\rho &= \Pr\left\{M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} < e^{n\epsilon}\right\}\times\\ \boldsymbol{E}\left\{\left[\sum_{m':\boldsymbol{u}_{m'}\in T_{\boldsymbol{u}|\boldsymbol{z}}} I_{m'}(0)\right]^\rho |M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}} < e^{n\epsilon}\right\} &+ O(e^{-n\epsilon e^{n\epsilon}})\end{aligned} \tag{13}$$

As the sum in the last expectation is of sub exponential order, we can distribute ρ over the sum and still preserve exponential tightness. We now condition on $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$. This will introduce dependencies in the drawings of $\{\boldsymbol{u}_m\}$ and of $\{\boldsymbol{x}_{m,i}\}$ and change $P_0(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ (since given $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$, the drawings $\{\boldsymbol{u}_m\}$ are no longer independent). To avoid this, we condition also on $\boldsymbol{u}_{m'}$. Given $\boldsymbol{u}_{m'}$, $\{\boldsymbol{x}_{m',i}\}$ are independent and $P_0(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ remain intact. Thus , (13) is shown to be given by

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

Given $\boldsymbol{u}$, the inner expectation is independent of the number of $M_{\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}}}$ and becomes $P_0(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$. Now, since $P_0(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}}, \hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})$ is constant for all $\boldsymbol{u}$'s in $T_{\boldsymbol{u}|\boldsymbol{z}}$, we are left with $e^{n(\bar{m}(\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}})+N(\hat{Q}_{\boldsymbol{x}|\boldsymbol{z}},\hat{Q}_{\boldsymbol{u}|\boldsymbol{z}},R_y))}$. Using this in (11), then in (10) and letting $n \to \infty$ yields $E_B(Q_z, R_y, R_{yz}, \rho, \lambda)$.

V. Numerical Results

In this section, we show some numerical results of our error exponents and compare them to the exponents of [4] and [8]. Our setup is that of a binary BC with a binary input X and a separate BSC 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 $\alpha = p(z \neq y) = \frac{p_z - p_y}{1-2p_y}$. Here, 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. 1). In Fig. 2, 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 [8]. 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 [8] and allowed it to be arbitrarily small, thus complying with the definition of an attainable exponent for the weak user.

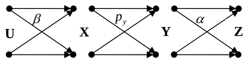
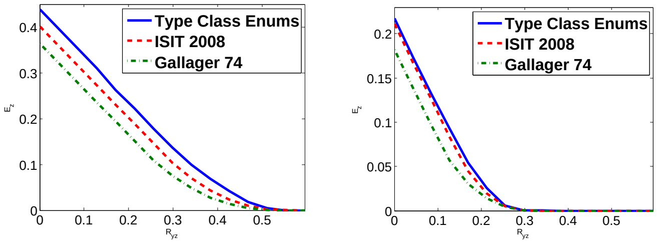


Fig. 1: The recast channel with the auxiliary variable.



(a)

(b)

Fig. 2: E_z for (a) $R_y = 0.05$[nats] and (b) $R_y = 0.3$[nats].

References

[1] T. M. Cover, "Broadcast Channels," *IEEE Transactions on Information Theory*, vol. IT–18, pp. 2–14, January 1972.

[2] J. Körner and K. Marton, "General Broadcast Channels with Degraded Message Sets," *IEEE Transactions on Information Theory*, vol. IT–23, no. 1, pp. 60–64, November 1977.

[3] P. P. Bergmans, "Random Coding Theorem for Broadcast Channels With Degraded Components," *IEEE Transactions on Information Theory*, vol. IT–19, pp. 197–207, March 1973.

[4] R. G. Gallager, "Capacity and Coding for Degraded Broadcast Channels," *Problemy Peredachi Informatsii*, vol. 10(3), pp. 3–14, 1974.

[5] J. Körner and A. Sgarro, "Universally Attainable Error Exponents for Broadcast Channels with Degraded Message Sets," *IEEE Transactions on Information Theory*, vol. IT–26, no. 6, pp. 670–679, November 1980.

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

[7] N. Merhav, "Error Exponents of Erasure/List Decoding Revisited via Moments of Distance Enumerators," *IEEE Trans. Inform. Theory*, vol. 54, no. 10 pp. 4439-4447, October 2008.

[8] Y. Kaspi, N. Merhav, "Error Exponents for Degraded Broadcast Channels with Degraded Message Sets," *Proc. ISIT 2008*, pp. 1518–1522, Toronto, Canada, July 2008.

[9] Y. Kaspi, N. Merhav, "Error Exponents of Optimum Decoding for the Degraded Broadcast Channel Using Moments of Type Class Enumerators," in preperation.

[10] N. Merhav, "Relations Between Random Coding Exponents and the Statistical Physics of Random Codes," *IEEE Trans. Inform. Theory*, vol. 55, no. 1, pp. 83–92, Jan. 2009.

[11] R. Etkin, N. Merhav, E. Ordentlich, "Error Exponents of Optimum Decoding for the Interference Channel," *Proc. ISIT 2008*, pp. 1523–1527, Toronto, Canada, July 2008.

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