

Error Exponents for Degraded Broadcast Channels with Degraded Message Sets

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—We consider a degraded broadcast channel with maximum likelihood decoders and derive lower bounds on the error exponent of each user. Unlike earlier results, our exponents pertain to optimal decoding and include both rates.

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. 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 random coding distributions.

Works on error exponents for general degraded broadcast channels include [4] and [6]. Both use 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 center. The resulting exponent

depends only on one of the rates - corresponding to the number of cloud centers. 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 error exponents for both the weak and strong decoders of a degraded BC with degraded message sets. The derived exponents pertain to optimum decoding and simultaneously depend on both rates. Numerical results show that for low rates of the private message, the new exponents improve on the previous results of [4].

The rest of this work is structured as follows: Section II gives the formal setting and notation. In Section III, we derive our new lower bounds on the error exponents for both the weak and the strong decoder. Section IV concentrates on a specific channel, the broadcast binary symmetric channel (BSC), and compare our exponents to the exponents in [4]. We conclude our work in Section V.

II. PRELIMINARIES

We begin with notation. Throughout this work, 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. 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 $P(\mathbf{y}, \mathbf{z}|\mathbf{x}) = \prod_{t=1}^n P_1(y_t|x_t)P_2(z_t|y_t)$, $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$. We are interested in sending one of $M_{yz} = 2^{nR_{yz}}$ messages to both receivers and one of $M_y = 2^{nR_y}$ to the strong receiver. Consider next a random selection of a hierarchical code [3], as follows: First, $M_{yz} = 2^{nR_{yz}}$ “cloud centers” $\mathbf{u}_1, \dots, \mathbf{u}_{M_{yz}} \in \mathcal{U}^n$ are drawn independently, each one using a distribution $Q(\mathbf{u}) = \prod_{t=1}^n Q(u_t)$, and then, for each $m = 1, 2, \dots, M_{yz}$,

$M_y = 2^{nR_y}$ codewords $\mathbf{x}_{m,1}, \dots, \mathbf{x}_{m,M_y} \in \mathcal{X}^n$ are drawn according to $Q(\mathbf{x}|\mathbf{u}) = \prod_{t=1}^n Q(x_t|u_t)$, with $\mathbf{u} = \mathbf{u}_m$.

The strong decoder, the one that observes \mathbf{y} , is interested in decoding both indices (m, i) of the transmitted codeword $\mathbf{x}_{m,i}$, whereas the weak decoder, the one that observes \mathbf{z} , is only interested in decoding the index m . Thus, while the strong decoder best applies full maximum likelihood (ML) decoding, $(\hat{m}(\mathbf{y}), \hat{i}(\mathbf{y})) = \arg \max_{m,i} P_1(\mathbf{y}|\mathbf{x}_{m,i})$, the best decoding rule for the weak decoder is given by $\tilde{m}(\mathbf{z}) = \arg \max_m \frac{1}{M_y} \sum_{i=1}^{M_y} P_3(\mathbf{z}|\mathbf{x}_{m,i})$, where $P_3(\mathbf{z}|\mathbf{x}) = \prod_{t=1}^n P_3(z_t|x_t) = \prod_{t=1}^n \sum_y P_2(z_t|y)P_1(y|x_t)$. Denote the average error probability of the strong decoder by $\overline{P}_E^y = Pr\{(\hat{m}(\mathbf{y}), \hat{i}(\mathbf{y})) \neq (m, i)\}$ and the average error probability of the weak decoder by $\overline{P}_E^z = Pr\{\tilde{m}(\mathbf{z}) \neq m\}$. We are interested in the random coding error exponents associated with both decoders. The exponents of the strong and weak decoders will be denoted by E_y, E_z respectively. A pair (E_y, E_z) is said to be attainable for a given (R_y, R_{yz}) if there exist random coding distributions $Q(u), Q(x|u)$ such that the random coding exponents satisfy

$$E_y \leq \lim_{n \rightarrow \infty} -\frac{1}{n} \log \overline{P}_E^y$$

$$E_z \leq \lim_{n \rightarrow \infty} -\frac{1}{n} \log \overline{P}_E^z$$

where all logarithms throughout the sequel are taken to the base of 2. We say that E_y or E_z is attainable, if there exist a pair (E_y, E_z) which is attainable.

III. THE EXPONENTS

A. The Weak Decoder

Applying Gallager's general upper bound ([8] p. 65) to the "channel" $P(\mathbf{z}|m) = \frac{1}{M_y} \sum_{i=1}^{M_y} P_3(\mathbf{z}|\mathbf{x}_{m,i})$, we have for $\lambda \geq 0, \rho \geq 0$:

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

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:

$$A \triangleq \mathbf{E} \left\{ \left[\frac{1}{M_y} \sum_{i=1}^{M_y} P_3(\mathbf{z}|\mathbf{X}_{m,i}) \right]^{1-\rho\lambda} \right\}$$

$$B \triangleq \mathbf{E} \left\{ \left[\sum_{m' \neq m} \left(\frac{1}{M_y} \sum_{j=1}^{M_y} P_3(\mathbf{z}|\mathbf{X}_{m',j}) \right)^\lambda \right]^\rho \right\}$$

As for A , we have

$$A = \mathbf{E} \left\{ \left[\frac{1}{M_y} \sum_{i=1}^{M_y} P_3(\mathbf{z}|\mathbf{X}_{m,i}) \right]^{1-\rho\lambda} \right\}$$

$$= M_y^{\rho\lambda-1} \cdot \mathbf{E} \left\{ \left[\sum_{i=1}^{M_y} P_3(\mathbf{z}|\mathbf{X}_{m,i}) \right]^{1-\rho\lambda} \right\}$$

$$= M_y^{\rho\lambda-1} \cdot \sum_{\mathbf{u}} Q(\mathbf{u}) \times \mathbf{E} \left\{ \left[\left(\sum_{j=1}^{M_y} P_3(\mathbf{z}|\mathbf{X}_{m,i}) \right)^{(1-\rho\lambda)/\alpha} \right]^\alpha \middle| \mathbf{u} \right\}$$

$$\stackrel{\alpha \geq 1-\rho\lambda}{\leq} M_y^{\rho\lambda-1} \cdot \sum_{\mathbf{u}} Q(\mathbf{u}) \times \mathbf{E} \left\{ \left[\sum_{j=1}^{M_y} P_3(\mathbf{z}|\mathbf{X}_{m,i})^{(1-\rho\lambda)/\alpha} \right]^\alpha \middle| \mathbf{u} \right\}$$

$$\stackrel{\alpha \leq 1}{\leq} M_y^{\alpha+\rho\lambda-1} \cdot \sum_{\mathbf{u}} Q(\mathbf{u}) \times \left[\sum_{\mathbf{x}} Q(\mathbf{x}|\mathbf{u}) P_3(\mathbf{z}|\mathbf{x})^{(1-\rho\lambda)/\alpha} \right]^\alpha \quad (1)$$

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

$$A \leq M_y^{\alpha+\rho\lambda-1} \prod_{t=1}^n \left[\sum_u Q(u) \times \left(\sum_x Q(x|u) P_3(z_t|x)^{(1-\rho\lambda)/\alpha} \right)^\alpha \right]. \quad (2)$$

Regarding B , we similarly obtain:

$$B \leq M_y^{(\mu-\lambda)\rho} M_{yz}^\rho \prod_{t=1}^n \left[\sum_{u'} Q(u') \times \left(\sum_{x'} Q(x'|u') P_3(z_t|x')^{\lambda/\mu} \right)^\mu \right]^\rho. \quad (3)$$

In the last expression, μ plays the same role as α did in the derivation of A . Denoting $f(a, b, z) = \sum_u Q(u) [\sum_x Q(x|u) P_3(z|x)^{a/b}]^b$, we obtain:

$$\overline{P}_E^z \leq M_y^{\alpha+\rho\mu-1} M_{yz}^\rho \times \left\{ \sum_z f(1-\rho\lambda, \alpha, z) \cdot f^\rho(\lambda, \mu, z) \right\}^n$$

$$= 2^{-n[E_0(\rho, \lambda, \alpha, \mu) - (\alpha+\rho\mu-1)R_y - \rho R_{yz}]} \quad (4)$$

where

$$E_0(\rho, \lambda, \alpha, \mu) = -\log \left[\sum_z f(1-\rho\lambda, \alpha, z) \cdot f(\lambda, \mu, z) \right]. \quad (5)$$

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

$$E(R_y, R_{yz}) = \max_{0 \leq \rho \leq 1, 0 \leq \lambda \leq \mu \leq 1, 1 - \rho \lambda \leq \alpha \leq 1} \{E_0(\rho, \lambda, \alpha, \mu) - (\alpha + \rho\mu - 1)R_y - \rho R_{yz}\} \quad (6)$$

Unlike in earlier papers [4], [5], [6], this exponent pertains to *optimal* decoding and it depends on both rates.

A few interesting choices of the parameters are the following:

1. Let $\alpha = \mu$. In this case, it can be shown that $\forall \lambda: E_0(\rho, \frac{1}{1+\rho}, \alpha, \alpha) \geq E_0(\rho, \lambda, \alpha, \alpha)$. Applying $\alpha = \mu, \lambda = \frac{1}{1+\rho}$ our bound becomes:

$$E(R_y, R_{yz}) = \max_{0 \leq \rho \leq 1, \frac{1}{1+\rho} \leq \alpha \leq 1} E_0\left(\rho, \frac{1}{1+\rho}, \alpha, \alpha\right) - [\alpha(1+\rho) - 1]R_y - \rho R_{yz}. \quad (7)$$

This is a somewhat more compact expression with only two parameters.

2. As a further restriction of item no. 1 above, consider the case where, in addition to the choice $\alpha = \mu$, we take $\alpha = \mu = \frac{1}{1+\rho}$. In this case, the expressions in the inner-most brackets of (2) and (3) become $\sum_x Q(x|u)P_3(z|x) \triangleq P_4(z|u)$, and $\alpha + \rho\mu - 1 = 0$. Thus, we get an exponent given by

$$E_0\left(\rho, \frac{1}{1+\rho}, \frac{1}{1+\rho}, \frac{1}{1+\rho}\right) - \rho R_{yz} - \log \left\{ \sum_z \left[\sum_u Q(u)P_4(z|u)^{1/(1+\rho)} \right]^{1+\rho} \right\} - \rho R_{yz} \quad (8)$$

which is exactly the ordinary Gallager function for the channel $P_4(z|u)$, obtained by sub-optimal decoding at the weak user [4], ignoring the knowledge of the refined codebook of each cloud center.

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

$$E_0\left(\rho, \frac{1}{1+\rho}, 1, 1\right) - \rho(R_y + R_{yz}) = -\rho(R_y + R_{yz}) - \log \left\{ \sum_z \left[\sum_x Q(x)P_3(z|x)^{1/(1+\rho)} \right]^{1+\rho} \right\}. \quad (9)$$

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

B. The Strong Decoder

The strong decoder (Y decoder) has to decode correctly both indices (m, i) of the transmitted $\mathbf{x}_{m,i}$. Applying Gallager's bound ([8] p. 65), and assuming,

without loss of generality, that $m = 1, i = 1$ was sent, we have for $\lambda \geq 0, \rho \geq 0$:

$$\begin{aligned} \overline{P_{E_{1,1}}^y} &\leq \sum_{\mathbf{y}} P_1(\mathbf{y}|\mathbf{x}_{1,1}) \left(\sum_{(m,i) \neq (1,1)} \frac{P_1(\mathbf{y}|\mathbf{x}_{m,i})^\lambda}{P_1(\mathbf{y}|\mathbf{x}_{1,1})^\lambda} \right)^\rho \\ &= \sum_{\mathbf{y}} P_1(\mathbf{y}|\mathbf{x}_{1,1})^{1-\lambda\rho} \left(\sum_{i=2}^{M_y} P_1(\mathbf{y}|\mathbf{x}_{1,i})^\lambda + \sum_{m=2}^{M_{yz}} \sum_{i=1}^{M_y} P_1(\mathbf{y}|\mathbf{x}_{m,i})^\lambda \right)^\rho \\ &\stackrel{\rho \leq 1}{\leq} \sum_{\mathbf{y}} P_1(\mathbf{y}|\mathbf{x}_{1,1})^{1-\lambda\rho} \left[\left(\sum_{i=2}^{M_y} P_1(\mathbf{y}|\mathbf{x}_{1,i})^\lambda \right)^\rho + \left(\sum_{m=2}^{M_{yz}} \sum_{i=1}^{M_y} P_1(\mathbf{y}|\mathbf{x}_{m,i})^\lambda \right)^\rho \right] \\ &\triangleq P_{E_{y1}} + P_{E_{y2}} \end{aligned} \quad (10)$$

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,

$$E_{y1}(R_y, \rho) = -\rho R_y - \log \sum_{\mathbf{y}} \sum_u Q(u) \left[\sum_x Q(x|u)P_1(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} \right]^{1+\rho} \quad (11)$$

We now turn to the second term in (10).

$$P_{E_{y2}} = \sum_{\mathbf{y}} P_1(\mathbf{y}|\mathbf{x}_{1,1})^{1-\lambda\rho} \left[\sum_{m=2}^{M_{yz}} \sum_{i=1}^{M_y} P_1(\mathbf{y}|\mathbf{x}_{i,m})^\lambda \right]^\rho \quad (12)$$

Here, when averaging over the ensemble, since the term in brackets of (12) originates from a different cloud, it is independent of the first term. Thus,

$$\begin{aligned} \overline{P_{E_{y2}}} &= \sum_{\mathbf{y}} \mathbf{E} [P_1(\mathbf{y}|\mathbf{X}_{1,1})^{1-\lambda\rho}] \times \\ &\quad \mathbf{E} \left[\sum_{m=2}^{M_{yz}} \sum_{i=1}^{M_y} P_1(\mathbf{y}|\mathbf{X}_{m,i})^\lambda \right]^\rho \\ &\stackrel{\rho \leq 1}{\leq} \sum_{\mathbf{y}} \mathbf{E} [P_1(\mathbf{y}|\mathbf{X}_{1,1})^{1-\lambda\rho}] \times \\ &\quad \left[\mathbf{E} \sum_{m=2}^{M_{yz}} \sum_{i=1}^{M_y} P_1(\mathbf{y}|\mathbf{X}_{m,i})^\lambda \right]^\rho \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} Q(\mathbf{x}) P_1(\mathbf{y}|\mathbf{x})^{1-\lambda\rho} \right] \times \\
&\quad \left[\sum_{m=2}^{M_{yz}} \sum_{i=1}^{M_y} \sum_{\mathbf{x}} Q(\mathbf{x}) P_1(\mathbf{y}|\mathbf{x})^\lambda \right]^\rho \\
&\leq M_y^\rho M_{yz}^\rho \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} Q(\mathbf{x}) P_1(\mathbf{y}|\mathbf{x})^{1-\lambda\rho} \right] \times \\
&\quad \left[\sum_{\mathbf{x}} Q(\mathbf{x}) P_1(\mathbf{y}|\mathbf{x})^\lambda \right]^\rho \quad (13)
\end{aligned}$$

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

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

For a memoryless channel, we will get:

$$\begin{aligned}
\overline{P_{E_{y2}}} &\leq M_y^\rho M_{yz}^\rho \left\{ \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} Q(\mathbf{x}) P_1(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} \right]^{1+\rho} \right\}^n \\
&= 2^{-n E_{y2}(R_y, R_{yz}, \rho)} \quad (14)
\end{aligned}$$

where

$$\begin{aligned}
E_{y2}(R_y, R_{yz}, \rho) &= -\rho(R_y + R_{yz}) \\
&\quad - \log \left\{ \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} Q(\mathbf{x}) P_1(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} \right]^{1+\rho} \right\}
\end{aligned}$$

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

To summarize, we have:

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

The dominant random coding error exponent for the Y-decoder will therefore be

$$\begin{aligned}
E_y(R_y, R_{yz}) &= \\
&\min \left(\max_{0 < \rho < 1} E_{y1}(R_y, \rho), \max_{0 < \rho < 1} E_{y2}(R_y, R_{yz}, \rho) \right) \quad (15)
\end{aligned}$$

IV. 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 $\alpha =$

¹This choice is optimal for the same reason it is optimal in the single user regime. see [7] Prob. 5.6

$p(z \neq y) = \frac{p_z - p_y}{1 - 2p_y}$. 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 [9]:

$$\begin{aligned}
R_z &\leq 1 - h(\beta * p_z) \\
R_y &\leq h(\beta * p_y) - h(p_y)
\end{aligned}$$

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 $0 \leq x \leq 1$.

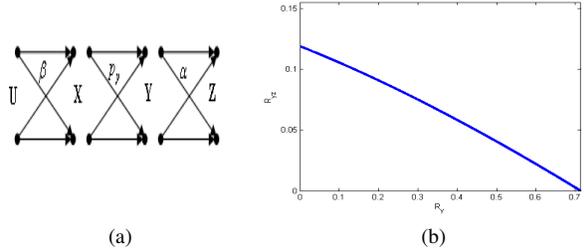


Fig. 1: (a)The recast channel with the auxiliary variable. (b)The capacity region $R_{yz}(R_y)$ with $p_y = 0.05, p_z = 0.3$

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 (8). $E_{g,y}$ is the minimum between (11) and

$$\max_{\rho} \left\{ -\log \left[\sum_{\mathbf{y}} \sum_{\mathbf{u}} Q(\mathbf{u}) \left(\sum_{\mathbf{x}} Q(\mathbf{x}|\mathbf{u}) \times P_1^{\frac{1}{1+\rho}}(\mathbf{y}|\mathbf{x}) \right)^{1+\rho} \right] - \rho R_{yz} \right\}. \quad (16)$$

For a given R_y, R_{yz} , β controls the relation 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 β that maximizes $1 - H(\beta * p_z)$ while keeping $E_y > 0$. In Fig. 2, 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. 2a 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.

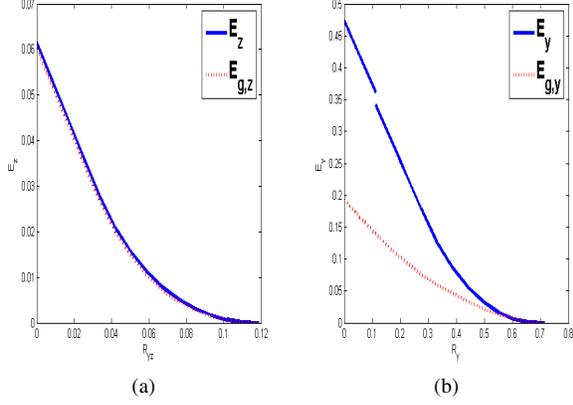


Fig. 2: Comparing E_y, E_z (solid curves) to $E_{g,y}, E_{g,z}$ (dotted curves) maximized over β . (a) $E_z(R_{yz})$ vs $E_{g,2}(R_{yz})$ for a fixed $R_y = 10^{-4}$. (b) $E_y(R_y)$ vs $E_{g,1}(R_y)$ for fixed $R_{yz} = 0.005$

Although the values of E_z and $E_{g,z}$ in Fig. 2a 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 (8), the 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 value² of E_y, E_z by $E_{y_{max}}, E_{z_{max}}$ respectively. In Fig. 3 we repeat the calculation of Fig. 2. However, here we restrict $E_y \geq E_{y_{max}}/4, E_z \geq E_{z_{max}}/4$ 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 . Note that as β increases, the channel $U \rightarrow Z$ 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 $\beta \geq \beta_c$, $E_z(R_y, R_{yz}, \beta \geq \beta_c) \triangleq E_{z_0}(R_y, R_{yz})$ is constant and has the form of (9), which is the single user error exponent ([8] p. 65) for the channel $X \rightarrow Z$ 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. Thus, for $R_y \geq R_{y_c}$, the maximization of E_y becomes constrained and the

²The maximal value is the single user error exponent ([8] 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 $R_z = 0, \beta = 0.5$

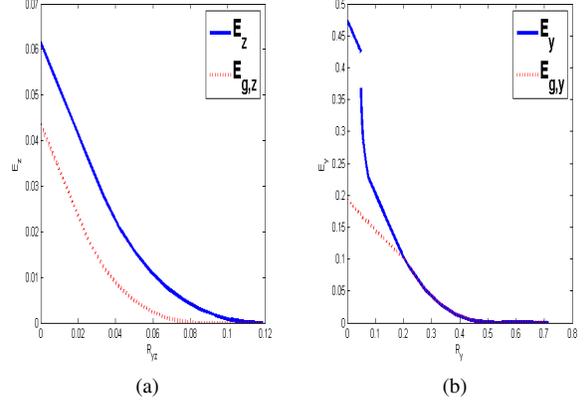


Fig. 3: Comparing E_y, E_z (solid curves) to $E_{g,y}, E_{g,z}$ (dotted curves) maximized over β . (a) $E_z(R_{yz})$ vs $E_{g,2}(R_{yz})$ for a fixed $R_y = 10^{-4}$ with $E_y \geq E_{y_{max}}/4$. (b) $E_y(R_y)$ vs $E_{g,1}(R_y)$ for fixed $R_{yz} = 0.005$ with $E_z \geq E_{z_{max}}/4$

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 .

V. CONCLUSIONS

In this work we derived new lower bounds on the error exponents of broadcast channels with degraded message sets. Our exponents pertain to optimal decoding and depend on both rates. We showed that they include previous results obtained in [4] as a special case. Numerical results for the degraded BSC showed that our exponents improve previous results.

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] L. Weng, S. S. Pradhan, and A. Anastasopoulos, "Error exponent region for Gaussian broadcast channels," *Proc. Conf. on Inform. Sciences and Systems (CISS)*, March 2004.
- [6] 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.
- [7] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, 1968.
- [8] A. J. Viterbi, J. K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill, 1979.
- [9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd Edition, Wiley, 2006.