Extremes of Random Coding Error Exponents

Albert Guillén i Fàbregas
Department of Engineering
University of Cambridge
Cambridge, CB2 1PZ, UK
guillen@ieee.org

Ingmar Land
Institute for Telecommunications Research
University of South Australia
Adelaide SA 5001, Australia
ingmar.land@ieee.org

Alfonso Martinez
Department of Engineering
University of Cambridge
Cambridge, CB2 1PZ, UK
alfonso.martinez@ieee.org

***Abstract*— We show that Gallager's random coding error exponent of an arbitrary binary-input memoryless symmetric channel is upper-bounded by that of the binary erasure channel and lower-bounded by that of the binary-symmetric channel of the same capacity. We apply the result to find the extremes of the channel dispersion for the aforementioned class of channels.**

I. Preliminaries

We consider a classical communication scenario, where equiprobable messages $m \in \{1, \ldots, M\}$ are to be transmitted over a binary-input memoryless symmetric channel described by the channel transition probabilities

$$P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} P_{Y|X}(y_i|x_i) \tag{1}$$

where n is the sequence length, $\boldsymbol{X} \in \mathcal{X}^n, \boldsymbol{Y} \in \mathcal{Y}^n$ are the random variables corresponding to the channel input and output sequences, respectively, X is the random variable corresponding to the channel inputs taking values x on the binary alphabet $\mathcal{X} = \{x_0, x_1\}$, Y is the random variable corresponding to the channel outputs taking values y on alphabet $\mathcal{Y}$. Therefore, the channel is fully characterized by the probabilities $P_{Y|X}(y|x_0), P_{Y|X}(y|x_1)$ for every $y \in \mathcal{Y}$.

We consider symmetric channels [1], i.e., channels for which the channel transition probability matrix (rows corresponding to input values) is such that it can be partitioned in submatrices for which each row is a permutation of any other row and each column is a permutation of any other column. Strongly symmetric channels are those for which the channel transition probability matrix fulfils this permutation property (without partitioning into subsets). Examples of the above channels include the binary erasure channel (BEC) and the binary symmetric channel (BSC). This also extends to channels with continuous output alphabets, such as the binary-input additive white Gaussian noise channel (BIAWGN).

Each of the M messages is mapped into a codeword $\boldsymbol{x}(m) = \big(x_1(m), \ldots, x_n(m)\big)$ with an encoder. The code $\mathcal{C}$ is the collection of all codewords. The code rate is $R = \frac{1}{n} \log M$. We consider maximum-likelihood (ML) decoding for which the estimated transmitted message $\hat{m}$ is obtained as follows

$$\hat{m} = \underset{m \in \{1,\ldots,M\}}{\arg\max}\ P_{\boldsymbol{Y}|\boldsymbol{X}}\big(\boldsymbol{y}|\boldsymbol{x}(m)\big). \tag{2}$$

[1]This work has been supported in part by the International Joint Project 2008/R2 of the Royal Society, by the Australian Research Council under ARC Discovery Grant DP0986089 and by the European Research Council under ERC grant agreement 259663.

The average probability of a message error is defined as

$$P_e = \frac{1}{M} \sum_{m=1}^{M} P_e(m) \tag{3}$$

where $P_e(m)$ is the probability of decoding message $\hat{m} \neq m$ when message m was transmitted. A code rate R is said to be achievable if for every $\epsilon > 0$ there exists a code of length n of rate not smaller than R such that $P_e < \epsilon$, for a suitably large n. The channel capacity C is the supremum of all achievable rates. For memoryless channels, the channel capacity is [2]

$$C = \max_{P_X(X)} I(X;Y) \tag{4}$$

where

$$I(X;Y) = \mathbb{E}\left[\log \frac{P_{Y|X}(Y|X)}{P_Y(Y)}\right] \tag{5}$$

is the mutual information.

Gallager proved the achievability of the channel capacity by showing that the code ensemble average error probability under ML decoding can be upperbounded by [1]

$$\bar{P}_e \leq e^{-nE_r(R)} \tag{6}$$

where

$$E_r(R) = \max_{0 \leq \rho \leq 1} E_0(\rho) - \rho R \tag{7}$$

is the random coding error exponent, and $E_0(\rho) \triangleq -\log F_0(\rho)$ is the Gallager function with

$$F_0(\rho) \triangleq \mathbb{E}\left[\left(\sum_{x' \in \mathcal{X}} P_X(x') \frac{P_{Y|X}(Y|x')^{\frac{1}{1+\rho}}}{P_{Y|X}(Y|X)^{\frac{1}{1+\rho}}}\right)^{\rho}\right]. \tag{8}$$

The above result provides the achievability of rates $R < I(X;Y)$, since the random coding exponent is positive for $R < I(X;Y)$. For symmetric channels, equiprobable inputs not only maximize the mutual information but also the error exponent [3]. For example, the BEC and BSC have that

$$F_0^{\mathrm{bec}}(\rho) = 2^{-\rho}(1-\varepsilon) + \varepsilon \tag{9}$$

$$F_0^{\mathrm{bsc}}(\rho) = 2^{-\rho}\left(\varepsilon^{\frac{1}{1+\rho}} + (1-\varepsilon)^{\frac{1}{1+\rho}}\right)^{1+\rho} \tag{10}$$

where ε denotes the erasure/crossover probability, respectively.

Figure 1 plots the error exponents for the BEC, BSC, BIAWGN and Rayleigh fading BIAWGN with perfect channel state information. The channel parameters are chosen such that the channel capacity is the same for all channels.

For future reference, we parametrize the above functions for a fixed channel capacity, i.e., $E_0(\rho; C)$ is the Gallager function

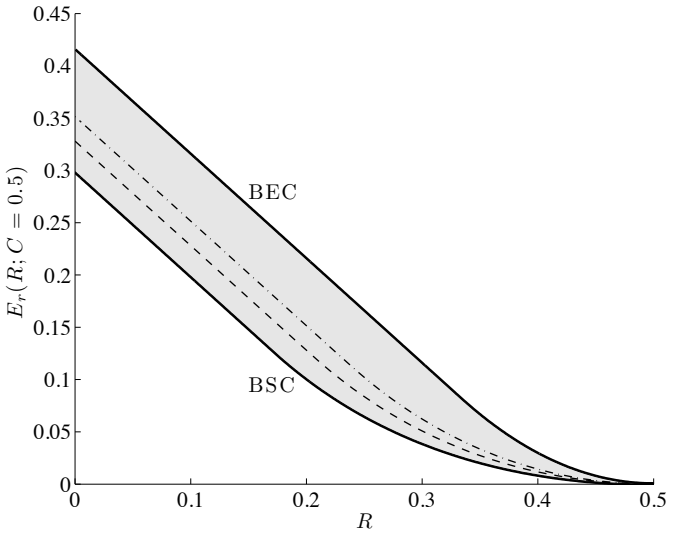

Fig. 1. Random coding error exponents of the BEC, BSC, BIAWGN (dashed), and Rayleigh fading BIAWGN (dash-dotted).

for a channel of capacity C. The functions $F_0(\rho; C)$ and $E_r(R; C)$ are defined similarly. Using the capacity expressions for the BEC and BSC we obtain

$$F_0^{\text{bec}}(\rho; C) = \left(\frac{1}{2^\rho} - 1\right) C + 1 \tag{11}$$

$$F_0^{\text{bsc}}(\rho; C) = \frac{1}{2^\rho}\left(\left(h^{-1}(1-C)\right)^{\frac{1}{1+\rho}} + \left(1 - h^{-1}(1-C)\right)^{\frac{1}{1+\rho}}\right)^{1+\rho} \tag{12}$$

where $h(p) = -p\log p - (1-p)\log(1-p)$ is the binary entropy function and $h^{-1}(x)$ denotes its inverse.

II. Main Result

The main results of this paper are Theorem 2.2, Corollary 2.1, and Theorem 2.3 and their proofs rely on the following 2 results.

Theorem 2.1 (BSC Channel Decomposition [4], [5]): Every binary-input symmetric memoryless channel can be decomposed into subchannels that are BSCs.

In this decomposition into BSC subchannels, each channel output Y is associated with an index $A = f(Y)$, which depends exclusively on the channel output. Denoting by $P_A(a)$ the probability mass or density function of subchannel a, and by $\mathcal{Y}(a)$ the corresponding binary output alphabet of BSC index a, we have that $P_{Y|X}(y|x) = P_{Y,A|X}(y, a|x) = P_{Y|X,A}(y|x, a)P_A(a)$ [4], [5]. Similarly, let $C(A) = I(X; Y|A)$ denote the capacity of BSC subchannel A. Note that $\mathbb{E}[C(A)] = C$ [4], [5].

Lemma 2.1: The function $F_0^{\text{bsc}}(\rho; C)$ is concave (convex-$\cap$) in $C \in [0, 1]$ for any $0 \le \rho \le 1$.

Proof: In order to show the result, one needs to evaluate the second derivative of $F_0^{\text{bsc}}(\rho; C)$ with respect to C and check the sign. This is a tedious and lengthy calculation, and details are omitted for the sake of readability. ■

Theorem 2.2: Consider an arbitrary binary-input symmetric memoryless channel. Then, the function $F_0(\rho; C)$ for $0 \le \rho \le 1$ and $0 \le C \le 1$ can be upper and lower bounded as follows

$$F_0^{\text{bec}}(\rho; C) \le F_0(\rho; C) \le F_0^{\text{bsc}}(\rho; C). \tag{13}$$

Proof: The proof is built around the decomposition of binary-input memoryless symmetric channels into BSCs (Theorem 2.1). Assume such a decomposition. Then, since $P_{Y|X}(y|x) = P_{Y|X,A}(y|x, a)P_A(a)$ [4], [5] we have that

$$F_0(\rho; C) = \mathbb{E}\left[\left(\sum_{x' \in \mathcal{X}} P_X(x') \frac{P_{Y|X}(Y|x')^{\frac{1}{1+\rho}}}{P_{Y|X}(Y|X)^{\frac{1}{1+\rho}}}\right)^\rho\right] \tag{14}$$

$$= \mathbb{E}\left[\mathbb{E}\left[\left(\sum_{x' \in \mathcal{X}} P_X(x') \frac{P_{Y|X,A}(Y|x', A)^{\frac{1}{1+\rho}}}{P_{Y|X,A}(Y|X, A)^{\frac{1}{1+\rho}}}\right)^\rho \middle| A\right]\right] \tag{15}$$

$$= \mathbb{E}\left[F_0^{\text{bsc}}(\rho; C(A))\right]. \tag{16}$$

From Lemma 2.1, the function $F_0^{\text{bsc}}(\rho; C)$ is concave. Therefore, by applying Jensen's inequality, we obtain

$$F_0(\rho; C) = \mathbb{E}\Big[F_0^{\text{bsc}}(\rho; C(A))\Big] \tag{17}$$

$$\le F_0^{\text{bsc}}(\rho; \mathbb{E}[C(A)]) \tag{18}$$

$$= F_0^{\text{bsc}}(\rho; C). \tag{19}$$

The bound is obviously achieved if the channel is a BSC.

Since $F_0^{\text{bsc}}(\rho; C)$ is concave, we lower-bound it by a straight line joining the points $F_0^{\text{bsc}}(\rho; 0)$ ($C = 0$) and $F_0^{\text{bsc}}(\rho; 1)$ ($C = 1$) (see Figure 2), and then evaluate the expectation, i.e.,

$$F_0^{\text{bsc}}(\rho; C) \ge F_0^{\text{bsc}}(\rho; 0) + C\left(F_0^{\text{bsc}}(\rho; 1) - F_0^{\text{bsc}}(\rho; 0)\right) \tag{20}$$

$$= 1 + C(2^{-\rho} - 1) \tag{21}$$

$$= F_0^{\text{bec}}(\rho; C). \tag{22}$$

This bound is obviously achieved if the channel is a BEC. ■

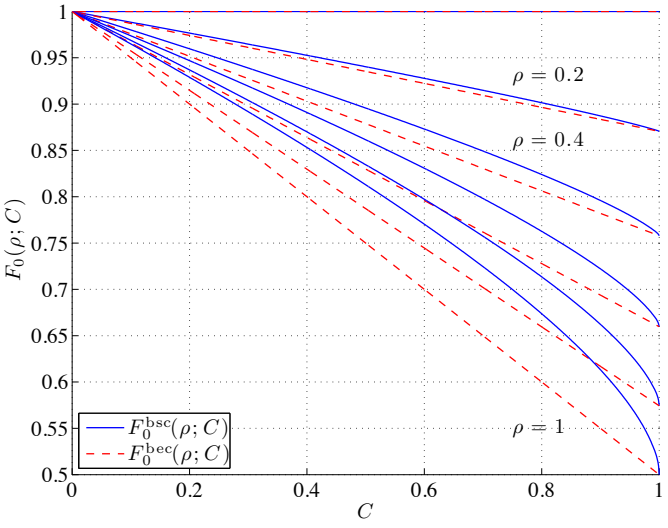

Fig. 2. Extremes of $F_0(\rho; C)$ for $\rho = 0, 0.2, 0.4, 0.6, 0.8, 1$.

The following results follow from Theorem 2.2.

Corollary 2.1: Consider an arbitrary binary-input symmetric memoryless channel. Then, the function $E_0(\rho; C)$ for $0 \leq \rho \leq 1$ and $0 \leq C \leq 1$ can be upper and lower bounded as follows

$$E_0^{\mathrm{bsc}}(\rho; C) \leq E_0(\rho; C) \leq E_0^{\mathrm{bec}}(\rho; C). \tag{23}$$

Furthermore, $E_0^{\mathrm{bsc}}(\rho; C)$ is convex (convex-$\cup$) in C.

Proof: The first claim follows directly from Theorem 2.2 and the fact that $b(x) = -\log x$ reverses the inequalities. The second claim follows convexity results on composed functions [6], since $b(x) = -\log x$ is convex and non-increasing, and $F_0^{\mathrm{bsc}}(\rho; C)$ is concave. ∎

Theorem 2.3 (Extremes of Error Exponents): Consider an arbitrary binary-input symmetric memoryless channel. Then, the random coding exponent $E_r(R; C)$ for $0 \leq C \leq 1$ can be upper and lower bounded as follows

$$E_r^{\mathrm{bsc}}(R; C) \leq E_r(R; C) \leq E_r^{\mathrm{bec}}(R; C). \tag{24}$$

Figure 1 illustrates the extremes of error exponents (Theorem 2.3). As we observe, the error exponents for the BIAWGN channel of the same capacity (with and without fading) fall between that of the BSC and that of the BEC. The error exponent of an arbitrary binary-input memoryless symmetric channel must lie in the shaded area.

III. Application: Extremes of Channel Dispersion

The channel dispersion was defined in [7] as

$$V = \lim_{P_e \to 0} \lim_{n \to \infty} n \left(\frac{C - R(n, P_e)}{Q^{-1}(P_e)} \right)^2 \tag{25}$$

The optimal rate for finite n and non-vanishing error probability can be approximated as [8], [7]

$$R(n, P_e) \approx C - \sqrt{\frac{V}{n}} Q^{-1}(P_e). \tag{26}$$

For discrete memoryless channels, Gallager's bound is tight at rates close to capacity [1]. Therefore, the channel dispersion can be obtained directly from the error exponent. In particular, one has that [9], [7]

$$V(C) = \frac{1}{E_r''(R = C; C)} = -E_0''(\rho = 0; C) \tag{27}$$

where we have parametrized V as a function of C, $E_r''(R; C) = \frac{\partial^2 E_r(R;C)}{\partial R^2}$ and $E_0''(\rho; C) = \frac{\partial^2 E_0(\rho;C)}{\partial \rho^2}$. The channel dispersion for the BEC and BSC is respectively given by [7],

$$V^{\mathrm{bec}}(C) = C(1 - C) \tag{28}$$

$$V^{\mathrm{bsc}}(C) = h^{-1}(1-C)\big(1 - h^{-1}(1-C)\big) \log^2 \frac{1 - h^{-1}(1-C)}{h^{-1}(1-C)} \tag{29}$$

We have the following result, illustrated in Figure 3.

Theorem 3.1 (Extremes of Channel Dispersion): Consider an arbitrary binary-input symmetric memoryless channel. Then, the channel dispersion $V(C)$ can be upper and lower bounded as follows

$$V^{\mathrm{bec}}(C) \leq V(C) \leq V^{\mathrm{bsc}}(C). \tag{30}$$

Proof: For an arbitrary channel with fixed C we have that $E_0(0; C) = 0$, and $E_0'(0; C) = \frac{\partial E_0(\rho;C)}{\partial \rho}\big|_{\rho=0} = C$. Therefore, a second-order Taylor expansion of $E_0(\rho; C)$ around $\rho = 0$ shows that $E_0''(0; C)$ will have the same extremes as $E_0(\rho; C)$. Hence, applying Corollary 2.1 to (27) yields the desired result. ∎

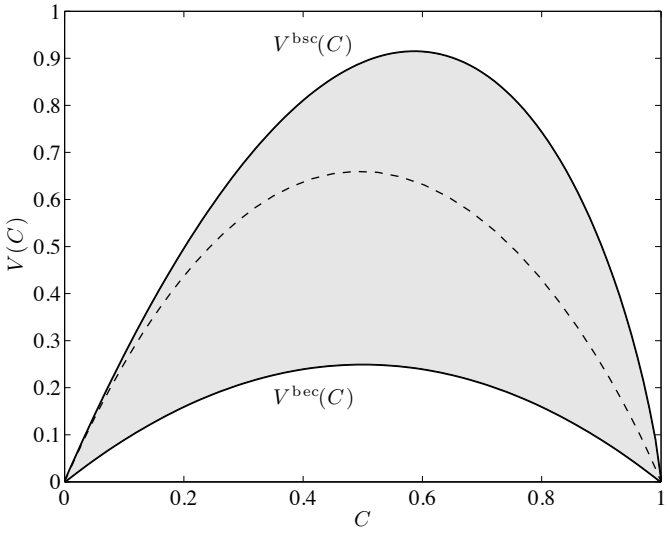

Fig. 3. Extremes of $V(C)$. The dashed line corresponds to the BIAWGN channel. The shaded area indicates the possible region.

References

[1] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, Inc. New York, NY, USA, 1968.

[2] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, 1948.

[3] F. Jelinek, *Probabilistic Information Theory*, McGraw-Hill New York, 1968.

[4] I. Land, *Reliability Information in Channel Decoding: Practical Aspects and Information Theoretical Bounds*, Ph.D. thesis, University of Kiel, Germany, 2004.

[5] I. Land and J. Huber, "Information combining," *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 3, pp. 227–330, 2006.

[6] S. P. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[7] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[8] V. Strassen, "Asymptotische Abschätzungen in Shannons Informationstheorie," in *Trans. Third Prague Conf. Information Theory, Statist. Decision Functions, Random Processes (Liblice, 1962)*, 1962, pp. 689–723.

[9] C.E. Shannon, "Certain results in coding theory for noisy channels," *Information and Control*, vol. 1, no. 1, pp. 6–25, 1957.