Extremes of Error Exponents

Albert Guillén i Fàbregas, *Senior Member, IEEE*, Ingmar Land, *Senior Member, IEEE*, and Alfonso Martinez, *Senior Member, IEEE*

***Abstract*—This paper determines the range of feasible values of standard error exponents for binary-input memoryless symmetric channels of fixed capacity C and shows that extremes are attained by the binary symmetric and the binary erasure channel. The proof technique also provides analogous extremes for other quantities related to Gallager's E_0 function, such as the cutoff rate, the Bhattacharyya parameter, and the channel dispersion.**

***Index Terms*—Bhattacharyya parameter, channel capacity, channel dispersion, cutoff rate, discrete memoryless channels, error exponents, error probability, random coding, symmetric channels.**

I. Introduction

In the context of coded communication, the channel coding theorem relates the error probability and the code rate, showing that there exist codes whose error probability tends to zero provided that the code rate is smaller than the channel capacity. For uncoded systems, the error probability and the channel capacity are also related. In particular, in [1]–[3], it is shown that given one of the two values, tight bounds on the other can be given for the family of binary-input memoryless and symmetric (BIMS) channels. Such channels are described by the channel transition probability $P_{Y|X}(y|x)$, where $x \in \{x_0, x_1\}$ and $y \in \mathcal{Y}$. We assume that the channel output alphabet $\mathcal{Y}$ has finite size, though our approach also holds for well-behaved channels with infinite alphabet size, like the binary-input additive white Gaussian noise (BIAWGN) channel. We adopt Gallager's definition of symmetric channel [4, p. 94], that is, a channel is said symmetric if 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. Both the binary erasure channel (BEC) and the binary symmetric channel (BSC) are symmetric.

More precisely, in [1]–[3], it is shown that the uncoded error probability of any BIMS channel with capacity C is upper-bounded by that of the BEC and lower-bounded by that of the BSC of the same capacity. Similar results have been found in [5] and [6] for the Bhattacharyya parameter, a simple upper bound to the uncoded error probability; here, only the extremal property of the BEC was proved. In the context of iterative decoding, analogous extremal properties of the BEC and BSC have been found [7], [8] for the building blocks of iterative decoders for low-density parity-check codes, namely variable-node and check-node decoders.

Upper and lower bounds to the error probability of good codes can be given in terms of error exponents, e.g., Gallager's random coding bound [4, Th. 5.6.3], the sphere-packing bound by Shannon *et al.* [9] and Arimoto's strong converse bound [10]. These exponents are expressed as optimization problems involving Gallager's E_0 function [4, Eq. 5.6.14]

$$E_0(\rho) \triangleq -\log F(\rho) \tag{1}$$

where

$$F(\rho) \triangleq \mathbb{E}\left[\left(\frac{\mathbb{E}\left[P_{Y|X}(Y|X')^{\frac{1}{1+\rho}}|Y\right]}{P_{Y|X}(Y|X)^{\frac{1}{1+\rho}}}\right)^{\rho}\right] \tag{2}$$

and the pair (X, Y) is distributed according to $P_X P_{Y|X}$. Here and throughout this paper, $\mathbb{E}[\cdot]$ denotes the expectation of a random variable and all logarithms are in base 2.

Equiprobable inputs maximize the E_0 function for BIMS channels [11, p. 203], and we henceforth assume such distribution, i.e., $P_X(x_0) = P_X(x_1) = \frac{1}{2}$.

In this paper, we characterize the feasible values of $E_0(\rho)$ for an arbitrary BIMS channel of fixed channel capacity C and show that the E_0 function is upper-bounded (respectively, lower-bounded) by that of the BEC (respectively, BSC) of the same capacity. Since the aforementioned exponents are expressed using the E_0 function, we are able to find their extremal values. In fact, our analysis leads to similar results for the cutoff rate, the Bhattacharyya parameter, the channel dispersion, and to a number of other extensions.

II. Feasible Pairs of Capacity C and $F(\rho)$ Function

The $F(\rho)$ functions for the BEC and BSC of erasure/crossover probability ε, respectively, denoted by $F^{\mathrm{bec}}(\rho)$ and $F^{\mathrm{bsc}}(\rho)$, are given by

$$F^{\mathrm{bec}}(\rho) \triangleq 2^{-\rho}(1-\varepsilon) + \varepsilon \tag{3}$$

$$F^{\mathrm{bsc}}(\rho) \triangleq 2^{-\rho}\left(\varepsilon^{\frac{1}{1+\rho}} + (1-\varepsilon)^{\frac{1}{1+\rho}}\right)^{1+\rho}. \tag{4}$$

Manuscript received June 28, 2012; revised November 19, 2012; accepted December 03, 2012. Date of publication December 10, 2012; date of current version March 13, 2013. This work was supported in part by the International Joint Project 2008/R2 of the Royal Society, in part by the Australian Research Council under ARC Discovery Grant DP0986089, and in part by the European Research Council under ERC grant agreement 259663. A. Martinez was supported in part by the Ministry of Economy and Competitiveness (Spain) under Grant RYC-2011-08150 and in part by the European Union's 7th Framework Programme (PEOPLE-2011-CIG) under Grant 303633. This paper was presented in part at the 2011 IEEE International Symposium on Information Theory.

A. Guillén i Fàbregas is with the Department of Information and Communication Technologies, Institució Catalana de Recerca i Estudis Avançats, Universitat Pompeu Fabra, 08018 Barcelona, Spain, and also with the Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, U.K. (e-mail: guillen@ieee.org).

I. Land is with the Institute for Telecommunications Research, University of South Australia, Adelaide, SA 5001, Australia (e-mail: ingmar.land@ieee.org).

A. Martinez is with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain (e-mail: alfonso.martinez@ieee.org).

Communicated by E. Arıkan, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2012.2233271

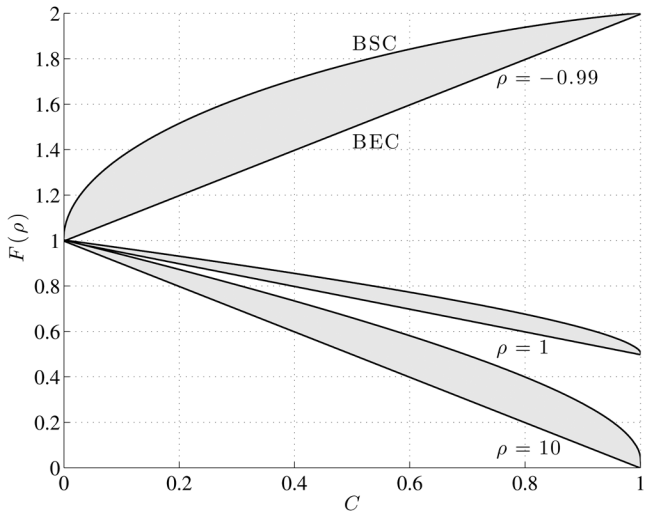

Fig. 1. Region of feasible points $(C, F(\rho))$ for $\rho = -0.99,\ 1,\ 10$. The upper curves correspond to the BSC and the lower straight lines to the BEC.

Using the capacity expressions for the BEC, $C^{\mathrm{bec}} \triangleq 1-\varepsilon$, and BSC, $C^{\mathrm{bsc}} \triangleq 1-h(\varepsilon)$, we can find the erasure/crossover probability corresponding to a given capacity C and parametrize the $F^{\mathrm{bec}}(\rho)$ and function $F^{\mathrm{bsc}}(\rho)$ as functions of C, namely

$$F^{\mathrm{bec}}(\rho; C) = 1 + \left(2^{-\rho} - 1\right) C \tag{5}$$

$$F^{\mathrm{bsc}}(\rho; C) = 2^{-\rho}\left(\left(h^{-1}(1-C)\right)^{\frac{1}{1+\rho}} + \left(1 - h^{-1}(1-C)\right)^{\frac{1}{1+\rho}}\right)^{1+\rho} \tag{6}$$

where $h(p) \triangleq -p\log p - (1-p)\log(1-p)$ is the binary entropy function, and $h^{-1}(x)$ denotes the inverse of $h(p)$ for $p \in [0, \frac{1}{2}]$. $C^{\mathrm{bec}}\big(F(\rho)\big), C^{\mathrm{bsc}}\big(F(\rho)\big)$ are, respectively, defined as the inverses of (5) and (6) with respect to C.

For BIMS channels, one has the bounds $0 \le C \le 1$ and $0 \le F(\rho) \le 1$ for $\rho \ge 0$ and $1 \le F(\rho) \le 2$ for $-1 < \rho \le 0$. This is a consequence of the facts that $F(\rho)$ is nonnegative and nonincreasing for $\rho > -1$ [4, App. 5B], that $\lim_{\rho\to-1} F(\rho) = 2$, and that $F(0) = 1$. It is, however, not apparent whether further limitations exist on the feasible pairs of capacity C and $F(\rho)$. Against this first impression, the next theorem tightly characterizes the set of possible pairs of capacity C and $F(\rho)$ function for any BIMS channel (see Fig. 1). In the next section, we apply this theorem and prove several analogous characterizations for other relevant quantities in the analysis of the error probability over BIMS channels.

Theorem 1: For any BIMS channel with capacity C and function $F(\rho)$ for $\rho > -1$, the following statements hold:

1) the function $F(\rho)$ of the channel satisfies

$$F^{\mathrm{bec}}(\rho; C) \le F(\rho) \le F^{\mathrm{bsc}}(\rho; C); \tag{7}$$

2) the capacity C of the channel satisfies

$$C^{\mathrm{bsc}}\big(F(\rho)\big) \le C \le C^{\mathrm{bec}}\big(F(\rho)\big), \quad -1 < \rho \le 0 \tag{8}$$

$$C^{\mathrm{bec}}\big(F(\rho)\big) \le C \le C^{\mathrm{bsc}}\big(F(\rho)\big), \quad \rho \ge 0. \tag{9}$$

The extremes in (7)–(9) are attained by the BEC and the BSC.

Furthermore, for a given pair $\big(C, F(\rho)\big)$ satisfying the inequalities in (7) or (8), there exists a BIMS channel with capacity C and function $F(\rho)$. Conversely, if the inequalities do not hold for the pair $\big(C, F(\rho)\big)$, there exists no such BIMS channel with capacity C and function $F(\rho)$.

A. Proof of Theorem 1

The proof is built around the idea that every BIMS channel admits a decomposition into subchannels that are BSCs. This decomposition follows directly from Gallager's definition of symmetric channels [4, p. 94] as used in this paper. A formal description may be found e.g., in [3] and [7]. Here, we deem identical the BEC with erasure probability 1 and the BSC with crossover probability $\frac{1}{2}$. In this decomposition, each channel output Y is associated with an index $A = f(Y)$ which is independent of the input and depends on the channel output only. We denote by $P_A(a)$ the probability mass or density function of subchannel a, and by $\mathcal{Y}(a)$ the corresponding binary-output alphabet of the BSC with index a. Assuming such a decomposition, and since $P_{Y|X}(y|x) = P_{Y|X,A}(y|x,a)P_A(a)$ [3], [7], we have

$$F(\rho) = \mathbb{E}\left[\left(\frac{\mathbb{E}\big[P_{Y|X}(Y|X')^{\frac{1}{1+\rho}}|Y\big]}{P_{Y|X}(Y|X)^{\frac{1}{1+\rho}}}\right)^{\rho}\right] \tag{10}$$

$$= \mathbb{E}\left[\mathbb{E}\left[\left(\frac{\mathbb{E}\big[P_{Y|X,A}(Y|X',A)^{\frac{1}{1+\rho}}|Y,A\big]}{P_{Y|X,A}(Y|X,A)^{\frac{1}{1+\rho}}}\right)^{\rho}\,\middle|\, A\right]\right] \tag{11}$$

$$= \mathbb{E}\left[F^{\mathrm{bsc}}(\rho; C(A))\right] \tag{12}$$

where $C(a)$ denotes the capacity of subchannel a.

The following lemma is proved in Appendix I.

Lemma 1: The function $F^{\mathrm{bsc}}(\rho; C)$ is concave in $C \in [0,1]$ for any $\rho > -1$, nondecreasing for $-1 < \rho \le 0$, and nonincreasing for $\rho \ge 0$.

Noting that $\mathbb{E}[C(A)] = C$, and given the concavity of the function $F^{\mathrm{bsc}}(\rho; C)$, we apply Jensen's inequality to obtain

$$F(\rho) = \mathbb{E}\Big[F^{\mathrm{bsc}}(\rho; C(A))\Big] \tag{13}$$

$$\le F^{\mathrm{bsc}}(\rho; \mathbb{E}[C(A)]) \tag{14}$$

$$= F^{\mathrm{bsc}}(\rho; C). \tag{15}$$

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

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

$$F^{\mathrm{bsc}}(\rho; C) \ge F^{\mathrm{bsc}}(\rho; 0) + C\left(F^{\mathrm{bsc}}(\rho; 1) - F^{\mathrm{bsc}}(\rho; 0)\right) \tag{16}$$

$$= 1 + C(2^{-\rho} - 1) \tag{17}$$

$$= F^{\mathrm{bec}}(\rho; C). \tag{18}$$

This bound is obviously achieved when the channel is a BEC, thus proving (7).

Equation (7) determines the boundaries of the region of feasible pairs $\big(C, F(\rho)\big)$. Since $F^{\mathrm{bsc}}(\rho; C)$ is concave and $F^{\mathrm{bec}}(\rho; C)$ is convex, the region of feasible pairs is convex. Moreover, the functions $F^{\mathrm{bsc}}(\rho; C)$ and $F^{\mathrm{bec}}(\rho; C)$ are nondecreasing for $-1 < \rho \leq 0$ and nonincreasing for $\rho \geq 0$. Fixing the value of $F(\rho)$, the convexity of the region implies (8).

We next prove that the region of feasible pairs $\big(C, F(\rho)\big)$ is connected by constructing a BIMS channel with corresponding capacity C and function $F(\rho)$. Consider a binary symmetric-erasure channel (BSEC) with input alphabet $\{x_0, x_1\}$, output alphabet $\{y_0, y_1, y_{\mathrm{e}}\}$, cross-over probability ε_{s} and erasure probability ε_{e}. Its transition probabilities are given by $P_{Y|X}(y_0|x_0) = P_{Y|X}(y_1|x_1) = 1-\varepsilon_{\mathrm{e}}-\varepsilon_{\mathrm{s}}, P_{Y|X}(y_0|x_1) = P_{Y|X}(y_1|x_0) = \varepsilon_{\mathrm{s}}$, and $P_{Y|X}(y_{\mathrm{e}}|x_0) = P_{Y|X}(y_{\mathrm{e}}|x_1) = \varepsilon_{\mathrm{e}}$. The capacity C^{bsec} and function $F^{\mathrm{bsec}}(\rho)$ are, respectively:

$$C^{\mathrm{bsec}} \triangleq (1-\varepsilon_{\mathrm{e}})\left(1 - h\left(\frac{\varepsilon_{\mathrm{s}}}{1-\varepsilon_{\mathrm{s}}-\varepsilon_{\mathrm{e}}}\right)\right) \tag{19}$$

$$F^{\mathrm{bsec}}(\rho) \triangleq 2^{-\rho}\big(\varepsilon_{\mathrm{s}}^{\frac{1}{1+\rho}} + (1-\varepsilon_{\mathrm{s}}-\varepsilon_{\mathrm{e}})^{\frac{1}{1+\rho}}\big)^{1+\rho} + \varepsilon_{\mathrm{e}}. \tag{20}$$

For fixed C, there exist several BSEC channels with capacity C, among them a BSC and a BEC. Each of them is characterized by a pair of probabilities ε_{s} and ε_{e}. The corresponding $F^{\mathrm{bsec}}(\rho)$ function is given by (20). Since the function $F^{\mathrm{bsec}}(\rho)$ is continuous in ε_{s} and ε_{e}, one can always find a BSEC with capacity C whose function $F^{\mathrm{bsec}}(\rho)$ coincides with the desired $F(\rho)$.

B. Applications

In the proof of Theorem 1, we exploited the fact that the region of feasible pairs $\big(C, F(\rho)\big)$ is convex and connected to characterize the extreme values of the capacity C or the function $F(\rho)$. In this section, we apply the theorem to provide extreme values for other relevant quantities in the error probability analysis of channel coding. A simple extension to channel parameters G given by $G(\rho) = g\big(F(\rho)\big)$, where $g(\cdot)$ is a monotonic continuous function, will prove convenient.

Theorem 2: Let the channel parameter G be given by $G(\rho) = g\big(F(\rho)\big)$, where $g(\cdot)$ is a monotonic strictly increasing continuous function. For any BIMS channel, we have that

1) the channel parameter $G(\rho)$ satisfies

$$g\big(F^{\mathrm{bec}}(\rho; C)\big) \leq G(\rho) \leq g\big(F^{\mathrm{bsc}}(\rho; C)\big); \tag{21}$$

2) the channel capacity C satisfies

$$C^{\mathrm{bsc}}\big(G(\rho)\big) \leq C \leq C^{\mathrm{bec}}\big(G(\rho)\big), \quad -1 < \rho \leq 0 \tag{22}$$

$$C^{\mathrm{bec}}\big(G(\rho)\big) \leq C \leq C^{\mathrm{bsc}}\big(G(\rho)\big), \quad \rho \geq 0. \tag{23}$$

Inequalities (21)–(23) are reversed if $g(\cdot)$ is monotonic, strictly decreasing and continuous.

Gallager's function: By letting $g(x) = -\log(x)$, the previous theorem readily gives the extremes of Gallager's function $E_0(\rho) = -\log F(\rho)$ for a fixed capacity, and the extremes of the capacity for a fixed E_0.

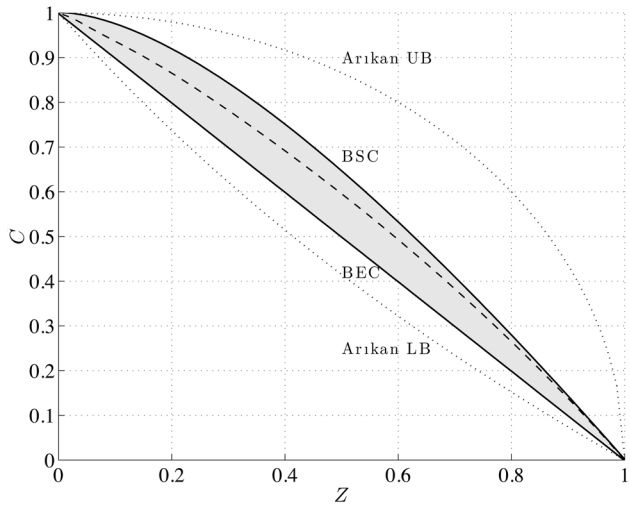

Fig. 2. Upper and lower bounds to the capacity C as a function of the Bhattacharyya parameter Z. Arıkan's upper and lower bounds [6, Eqs. (1), (2)] and the BIAWGN channel curve (dashed line) are also shown for reference.

Cutoff rate: A particular case of the E_0 function is the cutoff rate, given by $R_0 = E_0(1)$. Thus, the above result also gives the extremes of the cutoff rate.

Bhattacharyya parameter: A related quantity is the Bhattacharyya parameter Z, given by

$$Z = \sum_{y \in \mathcal{Y}} \sqrt{P_{Y|X}(y|x_0) P_{Y|X}(y|x_1)} = 2F(1) - 1. \tag{24}$$

The BSC/BEC have the largest/smallest possible Bhattacharyya parameter for BIMS channels of capacity C, interestingly giving the reverse extremes of the uncoded error probability [1]–[3]. This result recovers Sason's [5] and Arıkan's [6] bound for the BEC, and provides the extreme in the other direction attained by the BSC. Fig. 2 shows the bounds to C for a given value of Z from Theorem 2, as well as Arıkan's generic bounds for binary-input discrete memoryless channels [6, Eqs. (1), (2)], illustrating some improvement.

Random coding exponent: The random coding exponent $E_{\mathrm{r}}(R)$ [4, Sec. 5.6], given by

$$E_{\mathrm{r}}(R) = \max_{0 \leq \rho \leq 1} E_0(\rho) - \rho R, \tag{25}$$

provides an upper bound to the error probability of codes of rate R. This exponent involves a maximization of a function that, for fixed ρ falls under the conditions for applicability of Theorems 1 and 2. Therefore, the exponent $E_{\mathrm{r}}(R)$ satisfies

$$E_{\mathrm{r}}^{\mathrm{bsc}}(R; C) \leq E_{\mathrm{r}}(R) \leq E_{\mathrm{r}}^{\mathrm{bec}}(R; C). \tag{26}$$

Fig. 3 illustrates the extremes of random-coding error exponents $E_{\mathrm{r}}(R)$. The random-coding error exponent of an arbitrary BIMS channel must lie in the shaded area; two such examples are the BIAWGN channel of the same capacity (with and without fading).

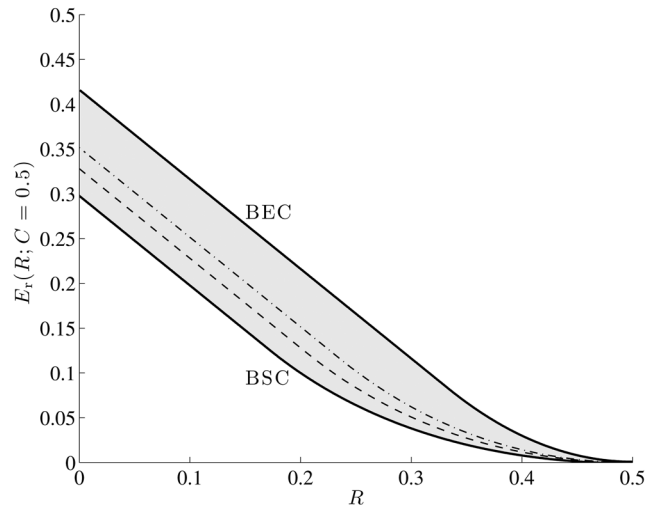

Fig. 3. Random coding error exponents of the BEC, BSC, BIAWGN (dashed), and Rayleigh fading BIAWGN (dash-dotted).

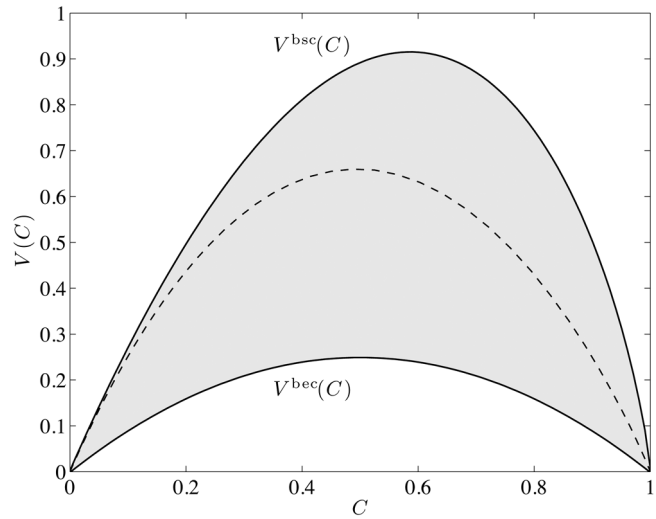

Fig. 4. Extremes of the channel dispersion $V(C)$.

Expurgated error exponent: For rates below the channel critical rate, the expurgated error exponent $E_{\rm ex}(R)$ [4, Sec. 5.7], given by

$$E_{\rm ex}(R) = \max_{\rho \geq 1} E_x(\rho) - \rho R, \tag{27}$$

provides a tighter estimate of the error probability of good codes than the random-coding exponent. The function $E_x(\rho)$ is expressed in terms of the Bhattacharyya parameter Z as

$$E_x(\rho) = -\rho \log \frac{1 + Z^{\frac{1}{\rho}}}{2}. \tag{28}$$

Theorem 2 provides the extremes of the expurgated exponent.

Strong converse exponent: In [10], Arimoto lower-bounded the error probability of block codes at rates above capacity in terms of the function $E_{\rm sc}(R)$ given by

$$E_{\rm sc}(R) = \sup_{-1 < \rho \leq 0} E_0(\rho) - \rho R. \tag{29}$$

Theorem 2 also provides the extremes of this exponent.

Sphere-packing exponent: The error probability of codes of rate R is lower-bounded by a bound that depends on the sphere-packing exponent [9]$E_{\rm sp}(R)$, given by

$$E_{\rm sp}(R) = \sup_{\rho > 0} E_0(\rho) - \rho R. \tag{30}$$

Again, Theorem 2 provides the extremes of this exponent.

Threshold-decoding error exponents: The exponent of random-coding bounds based on threshold decoding can also be expressed in closed form. Shannon [12] derived the exponent of Feinstein's bound to the error probability [13]. More generally, the exponent corresponding to a generalized form of Feinstein's bound [14] can be expressed as

$$E_{\rm r}^{\rm gfb}(R) = \sup_{\rho \geq 0} \frac{E_0(\rho) - \rho R}{1 + \rho}. \tag{31}$$

Theorem 2 directly gives the error exponent extremes for the generalized Feinstein's bound.

The exponent of the dependence-testing bound [15] is [14]

$$E_{\rm r}^{\rm dtb}(R) = \max_{0 \leq \rho \leq 1} E_0(\rho, s = 1) - \rho R \tag{32}$$

where $E_0(\rho, s) \triangleq -\log F(\rho, s)$, for $s \geq 0$, and

$$F(\rho, s) \triangleq \mathbb{E}\left[\left(\frac{\mathbb{E}\left[P_{Y|X}(Y|X')^{\rm s}|Y\right]}{P_{Y|X}(Y|X)^{\rm s}}\right)^{\rho}\right]. \tag{33}$$

Following similar and somewhat simpler steps to those in the proof of Lemma 1, one can prove that $F(\rho, s = 1)$, evaluated for a BSC with capacity C, is concave in C. Therefore, Theorem 2 holds and shows that the exponent of the DT bound has similar extreme values.

Channel dispersion: Recently, the Gaussian approximation to the error probability $P_{\rm e}$ of length-n codes at rates close to the capacity has received renewed attention. In this approximation, a key channel parameter is the dispersion V, which for BIMS channels [12], [15] is given by

$$V = \frac{1}{E_{\rm r}''(R = C)} = -E_0''(\rho = 0). \tag{34}$$

Moreover, it can be proved that one can choose either the $E_0(\rho)$ function or the simpler $E_0(\rho, 1)$ to compute the latter derivative, that is $E_0''(\rho = 0) = E_0''(\rho = 0, 1)$. As proved in Appendix I, the third derivative of $E_0(\rho, 1)$ at $\rho = 0$ is bounded for BIMS channels. Thus, a second-order Taylor expansion of $E_0(\rho, 1)$ around $\rho = 0$ shows that $E_0''(0, 1)$ has the same extremes as $E_0(\rho, 1)$. As illustration, Fig. 4 depicts the possible values of channel dispersion as a function of the capacity C of the BIMS channel. The dashed line, which lies within the shaded area indicating the feasible region of pairs capacity/dispersion, corresponds to the BIAWGN channel.

Error probability of specific codes: Our theorems may also be applied to specific codes $\mathcal{C}$ with a given distance spectrum by means of the Shulman–Feder bound [16] (see also [17]) given by

$$-\frac{1}{n}\log P_{\rm e}(\mathcal{C}) \geq E_{\rm r}\left(R+\frac{1}{n}\log\alpha_{\mathcal{C}}\right) \tag{35}$$

where $\alpha_{\mathcal{C}}$ is a function of the distance spectrum of the code that quantifies how far the distance spectrum of $\mathcal{C}$ is from that of the ensemble average.

Exact error probability: One might wonder whether our extremal results extend to the actual error probability. The answer is not immediately obvious. For uncoded transmission (a code of length $n=1$ and rate $R=1$) over a given BIMS channel of capacity C, the error probability P_b is upper- and lower-bounded by that of the BEC and the BSC, respectively, $h^{-1}(1-C) \leq P_b \leq \frac{1}{2}(1-C)$ [2]. In contrast, the extremes of the exponential bounds to the error probability, including the Bhattacharyya parameter, are reversed. This phenomenon suggests the existence of a pair (n, R_n) such that a crossing point occurs, in the sense that for rates above (respectively, below) R_n the extremes may be those of uncoded transmission (respectively, the error exponents).

Connection with Arıkan, Telatar, and Alsan: Unpublished work by Arıkan and Telatar [18] uncovered results of similar nature to those reported in this paper, showing that for channels with a fixed rate $R_\rho \triangleq \frac{\partial E_0(\rho)}{\partial \rho}$, for $0 \leq \rho \leq 1$, the random coding exponent satisfies

$$E_{\rm r}^{\rm bec}(R_\rho) \leq E_{\rm r}(R_\rho) \leq E_{\rm r}^{\rm bsc}(R_\rho). \tag{36}$$

For $\rho = 0$, we have that $R_\rho = C$ and we obtain the trivial result that $0 = E_{\rm r}^{\rm bec}(R_\rho) \leq E_{\rm r}(R_\rho) \leq E_{\rm r}^{\rm bsc}(R_\rho) = 0$. Instead, our results compare channels of a fixed capacity and provide the extremal values of the random-coding exponent and other quantities. The suitability of either of these two approaches to the problem may depend on the specific application. A more recent result by Alsan [19] recovers both Theorem 1 in this paper and the results in [18] as particular cases, for BIMS channels in the interval $0 \leq \rho \leq 1$.

Appendix I
Proof of Lemma 1

We aim at proving the concavity of the function

$$f(C) = \left(\varepsilon(C)^{\frac{1}{1+\rho}} + \left(1-\varepsilon(C)\right)^{\frac{1}{1+\rho}}\right)^{1+\rho} \tag{37}$$

where $\varepsilon(C)$ is itself a function of C, namely $\varepsilon = h^{-1}(1-C)$. Without loss of generality, we limit our attention to the interval $\varepsilon \in [0, 1/2]$. The function is concave if $\frac{d^2 f}{dC^2} \leq 0$.

Applying the chain rule of derivation, we have that

$$\frac{df}{dC} = \frac{df}{d\varepsilon}\frac{d\varepsilon}{dC} \tag{38}$$

$$\frac{d^2 f}{dC^2} = \frac{d^2 f}{d\varepsilon^2}\left(\frac{d\varepsilon}{dC}\right)^2 + \frac{df}{d\varepsilon}\frac{d^2\varepsilon}{dC^2}. \tag{39}$$

Direct computation gives

$$\frac{df}{d\varepsilon} = \left(\varepsilon^{\frac{1}{1+\rho}} + (1-\varepsilon)^{\frac{1}{1+\rho}}\right)^{\rho}\left(\varepsilon^{\frac{-\rho}{1+\rho}} - (1-\varepsilon)^{\frac{-\rho}{1+\rho}}\right) \tag{40}$$

$$\frac{d^2 f}{d\varepsilon^2} = -\frac{\rho}{1+\rho}\frac{\left(\varepsilon^{\frac{1}{1+\rho}} + (1-\varepsilon)^{\frac{1}{1+\rho}}\right)^{\rho-1}}{\left(\varepsilon(1-\varepsilon)\right)^{\frac{1+2\rho}{1+\rho}}}. \tag{41}$$

An application of the inverse function theorem yields

$$\frac{d\varepsilon}{dC} = \frac{1}{\log\frac{\varepsilon}{1-\varepsilon}} \tag{42}$$

$$\frac{d^2\varepsilon}{dC^2} = -\frac{1}{\varepsilon(1-\varepsilon)\left(\log\frac{\varepsilon}{1-\varepsilon}\right)^3 \ln 2}. \tag{43}$$

The derivatives with respect to C are, therefore, given by

$$\frac{df}{dC} = \left(\varepsilon^{\frac{1}{1+\rho}} + (1-\varepsilon)^{\frac{1}{1+\rho}}\right)^{\rho}\left(\varepsilon^{\frac{-\rho}{1+\rho}} - (1-\varepsilon)^{\frac{-\rho}{1+\rho}}\right)\frac{1}{\log\frac{\varepsilon}{1-\varepsilon}} \tag{44}$$

$$\frac{d^2 f}{dC^2} = -\frac{\left(\varepsilon^{\frac{1}{1+\rho}} + (1-\varepsilon)^{\frac{1}{1+\rho}}\right)^{\rho-1}}{\left(\log\frac{\varepsilon}{1-\varepsilon}\right)^2\left(\varepsilon(1-\varepsilon)\right)^{\frac{1+2\rho}{1+\rho}}} \left(\frac{\rho}{1+\rho} + \frac{1-2\varepsilon+\varepsilon^{\frac{1}{1+\rho}}(1-\varepsilon)^{\frac{\rho}{1+\rho}} - \varepsilon^{\frac{\rho}{1+\rho}}(1-\varepsilon)^{\frac{1}{1+\rho}}}{\ln\left(\frac{\varepsilon}{1-\varepsilon}\right)}\right). \tag{45}$$

Since we have that $\frac{df}{dC} \geq 0$ for $-1 < \rho \leq 0$ and $\frac{df}{dC} \leq 0$ for $\rho \geq 0$, we conclude that $f(C)$ is increasing and decreasing in the respective ranges of ρ.

The term before the brackets

$$-\frac{\left(\varepsilon^{\frac{1}{1+\rho}} + (1-\varepsilon)^{\frac{1}{1+\rho}}\right)^{\rho-1}}{\left(\log\frac{\varepsilon}{1-\varepsilon}\right)^2\left(\varepsilon(1-\varepsilon)\right)^{\frac{1+2\rho}{1+\rho}}} \tag{46}$$

is always nonpositive for $-1 \leq \rho \leq \infty$. Therefore, it suffices to show that the function

$$g(\varepsilon,\rho) \triangleq \frac{\rho}{1+\rho} + \frac{1-2\varepsilon+\varepsilon^{\frac{1}{1+\rho}}(1-\varepsilon)^{\frac{\rho}{1+\rho}} - \varepsilon^{\frac{\rho}{1+\rho}}(1-\varepsilon)^{\frac{1}{1+\rho}}}{\ln\left(\frac{\varepsilon}{1-\varepsilon}\right)} \tag{47}$$

is nonnegative for $\varepsilon \in [0, \frac{1}{2}]$ and $-1 \leq \rho \leq \infty$.

Let $z \triangleq \frac{\varepsilon}{1-\varepsilon} \in [0,1]$. With this change of variables, we obtain

$$g(z,\rho) = \frac{\rho}{1+\rho} + \frac{1-z+z^{\frac{1}{1+\rho}} - z^{\frac{\rho}{1+\rho}}}{(1+z)\ln z}. \tag{48}$$

We wish to show that $g(z,\rho) \geq 0$. The partial derivative with respect to ρ is given by

$$\frac{\partial g(z,\rho)}{\partial \rho} = \frac{1}{(1+\rho)^2}\left(\frac{1+z-z^{\frac{1}{1+\rho}} - z^{\frac{\rho}{1+\rho}}}{1+z}\right) \tag{49}$$

$$\triangleq \frac{1}{(1+z)(1+\rho)^2}\, g_0(z,\rho). \tag{50}$$

We are interested in the sign of $g_0(z,\rho)$, whose derivative is in turn given by

$$\frac{\partial g_0(z,\rho)}{\partial \rho} = \frac{\ln z}{(1+\rho)^2}\left(z^{\frac{1}{1+\rho}} - z^{\frac{\rho}{1+\rho}}\right). \tag{51}$$

We readily see that

$$\frac{\partial g_0(z,\rho)}{\partial \rho} \geq 0, \quad \rho \in (-1,1] \tag{52}$$

$$\frac{\partial g_0(z,\rho)}{\partial \rho} \leq 0, \quad \rho \in [1,+\infty). \tag{53}$$

Summarizing, since $g(z,\rho)$ is continuous in ρ for $\rho > -1$, we have that

1) $\frac{\partial g(z,\rho)}{\partial \rho} \leq 0$ in $\rho \in (-1,0]$, since $g_0(z,\rho)$ is nondecreasing and $g_0(z,\rho) \leq g_0(z,0) = 0$;
2) $\frac{\partial g(z,\rho)}{\partial \rho} \geq 0$ in $\rho \in [0,1]$, since $g_0(z,\rho)$ is nondecreasing and $g_0(z,\rho) \geq g_0(z,0) = 0$;
3) $\frac{\partial g(z,\rho)}{\partial \rho} \geq 0$ in $\rho \in [1,\infty)$, since $g_0(z,\rho)$ is nonincreasing and $g_0(z,\rho) \geq \lim_{\rho\to\infty} g_0(z,\rho) = 0$.

The fact that $g(z,0) = 0$ concludes the proof.

Appendix II

We wish to prove that the partial derivative $\frac{\partial^3 E_0(\rho,s=1)}{\partial \rho^3}\big|_{\rho=0}$ is bounded. To this end, we first note that the function $E_0(\rho, s=1)$ can be expressed as

$$E_0(\rho, s=1) = -\log \mathbb{E}\left[2^{-\rho\, i(X;Y)}\right] \tag{54}$$

where $i(x;y)$ is the information density, defined as

$$i(x;y) \triangleq \log \frac{P_{Y|X}(y|x)}{P_Y(y)}. \tag{55}$$

The function $E_0(\rho, s=1)$ is a cumulant generating function. Its third derivative evaluated at $\rho = 0$ gives the third-order cumulant, that is the third-order central moment,

$$\left.\frac{\partial^3 E_0(\rho,s=1)}{\partial \rho^3}\right|_{\rho=0} = \mathbb{E}\Big[\big(i(X;Y) - I(X;Y)\big)^3\Big](\ln 2)^2. \tag{56}$$

The next result shows that the kth absolute moment of the information density is bounded.

Lemma 2: Consider a memoryless channel with discrete input alphabet $\mathcal{X}$ and arbitrary output alphabet $\mathcal{Y}$. Then, with equiprobable inputs, we have

$$\mathbb{E}\Big[\big|i(X;Y) - I(X;Y)\big|^k\Big] \leq \left(2\log|\mathcal{X}| + \frac{k}{\ln 2}\big(1 + |\mathcal{X}|^{\frac{1}{k}}\big)\right)^k. \tag{57}$$

Proof: We will make use of Minkowski's inequality $\|A+B\|_k \leq \|A\|_k + \|B\|_k$ where $\|A\|_k \triangleq (\mathbb{E}[|A|^k])^{\frac{1}{k}}$. Using the definition of $i(X,Y)$, we now have that

$$\begin{aligned}
&\|i(X;Y) - I(X;Y)\|_k \\
&\leq \left\| \log \frac{\sum_{x'} \frac{1}{|\mathcal{X}|} P_{Y|X}(Y|x')}{P_{Y|X}(Y|X)} \right\|_k + I(X;Y) && (58)\\
&\leq 2\log|\mathcal{X}| + \left\| \log \frac{\sum_{x'} P_{Y|X}(Y|x')}{P_{Y|X}(Y|X)} \right\|_k && (59)\\
&\leq 2\log|\mathcal{X}| + \frac{1}{\ln 2}\left\| k\left(\frac{\sum_{x'} P_{Y|X}(Y|x')}{P_{Y|X}(Y|X)}\right)^{\frac{1}{k}} - k \right\|_k && (60)\\
&\leq 2\log|\mathcal{X}| + \frac{k}{\ln 2} + \frac{k}{\ln 2}\left(\mathbb{E}\left[\frac{\sum_{x'} P_{Y|X}(Y|x')}{P_{Y|X}(Y|X)}\right]\right)^{\frac{1}{k}} && (61)\\
&\leq 2\log|\mathcal{X}| + \frac{k}{\ln 2} + \frac{k}{\ln 2}|\mathcal{X}|^{\frac{1}{k}} && (62)
\end{aligned}$$

where we have used that $\ln x \leq k(x^{\frac{1}{k}} - 1)$ [20, Eq. (4.1.37)]. ■

Using Lemma 2, we have that for BIMS channels

$$\left.\frac{\partial^3 E_0(\rho, s=1)}{\partial \rho^3}\right|_{\rho=0} \leq (\ln 2)^2\left(2 + \frac{3}{\ln 2}(1 + 2^{\frac{1}{3}})\right)^3. \tag{63}$$

Acknowledgment

The authors wish to thank Erdal Arıkan and Emre Telatar for providing [18].

References

[1] M. Hellman and J. Raviv, "Probability of error, equivocation, and the Chernoff bound," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 4, pp. 368–372, Jul. 1970.
[2] I. Land, "Reliability information in channel decoding: Practical aspects and information theoretical bounds," Ph.D. dissertation, Univ. Kiel, Germany, 2004.
[3] I. Land and J. Huber, "Information combining," *Found. Trends Commun. Inf. Theory*, vol. 3, no. 3, pp. 227–330, 2006.
[4] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[5] I. Sason, "On universal properties of capacity-approaching LDPC code ensembles," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 2956–2990, Jul. 2009.
[6] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.
[7] I. Land, S. Huettinger, P. A. Hoeher, and J. B. Huber, "Bounds on information combining," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 612–619, Feb. 2005.
[8] I. Sutskover, S. Shamai, and J. Ziv, "Extremes of information combining," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1313–1325, Apr. 2005.
[9] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Inf. Control*, vol. 10, no. 1, pp. 65–103, 1967.

[10] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 3, pp. 357–359, May 1973.
[11] F. Jelinek, *Probabilistic Information Theory*. New York: McGraw-Hill, 1968.
[12] C. E. Shannon, "Certain results in coding theory for noisy channels," *Inf. Control*, vol. 1, no. 1, pp. 6–25, 1957.
[13] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inf. Theory*, vol. 4, no. 4, pp. 2–22, 1954.
[14] A. Martinez and A. Guillén i Fàbregas, "Random-coding bounds for threshold decoders: Error exponent and saddlepoint approximation," presented at the IEEE Int. Symp. Inf. Theory, Saint Petersburg, Russia, Jul./Aug. 2011.
[15] 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.
[16] N. Shulman and M. Feder, "Random coding techniques for nonrandom codes," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2101–2104, Sep. 1999.
[17] I. Sason and S. Shamai, "Performance analysis of linear codes under maximum-likelihood decoding: A tutorial," *Found. Trends Commun. Inf. Theory*, vol. 3, no. 1–2, 2006.
[18] E. Arıkan and E. Telatar, "BEC and BSC are E_0 extremal," unpublished technical note, 2008.
[19] M. Alsan, "Extremality properties for Gallager's random coding exponent," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, Jul. 2012, pp. 2944–2948.
[20] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. New York: Dover, 1964.

Albert Guillén i Fàbregas (S'01–M'05–SM'09) was born in Barcelona, Catalunya, Spain, in 1974. In 1999 he received the Telecommunication Engineering Degree and the Electronics Engineering Degree from Universitat Politècnica de Catalunya and Politecnico di Torino, respectively, and the Ph.D. in Communication Systems from Ecole Polytechnique Fédérale de Lausanne (EPFL) in 2004.

Since 2011 he has been a Research Professor of the Institució Catalana de Recerca i Estudis Avançats (ICREA) hosted at the Department of Information and Communication Technologies, Universitat Pompeu Fabra. He is also an Adjunct Researcher at the Department of Engineering, University of Cambridge. He has held appointments at the New Jersey Institute of Technology, Telecom Italia, European Space Agency (ESA), Institut Eurécom, University of South Australia, University of Cambridge where he was a Reader and a Fellow of Trinity Hall, as well as visiting appointments at EPFL, Ecole Nationale des Télécommunications (Paris), Universitat Pompeu Fabra, University of South Australia, Centrum Wiskunde & Informatica and Texas A&M University in Qatar. His specific research interests are in the area of information theory, communication theory, coding theory, digital modulation and signal processing techniques.

Dr. Guillén i Fàbregas received the Starting Grant from the European Research Council, the Young Authors Award of the 2004 European Signal Processing Conference, the 2004 Best Doctoral Thesis Award from the Spanish Institution of Telecommunications Engineers, and a Research Fellowship of the Spanish Government to join ESA. He is a co-author of the monograph book *Bit-Interleaved Coded Modulation*. He is an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY and *Foundations and Trends in Communications and Information Theory*, Now Publishers, a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and was an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2007–2011).

Ingmar Land (S'00–M'04–SM'11) is Senior Research Fellow at the Institute for Telecommunications Research (ITR), University of South Australia. Before joining ITR in 2007, he was Assistant Professor for Communication Theory at Aalborg University, Denmark. He received his Dr-Ing in 2004 from the University of Kiel, Germany, and he studied for his Dipl.-Ing. at the University of Ulm, Germany, and at the University of Erlangen-Nürnberg, Germany.

Dr. Land's areas of research are information theory, coding theory and communication theory with application to cooperative communications, multi-user communication, satellite communications, distributive source coding and physical-layer security.

Dr. Land received the Fakultätspreis 2005 from the University of Kiel, Germany, for the best doctoral thesis at the Faculty of Engineering. He received the ITG Literaturpreis 2005 from the Verband der Elektrotechnik, Elektronik und Informationstechnik (VDE), Germany, for pioneering research in the field of information technology in 2005. And he received the Teacher of the Year 2005 Award from Aalborg University, Denmark, for excellence in teaching at the School for Electronics and Information Technology. Dr. Land is co-author of the monograph "Information Combining" published in Foundations and Trends in Communications and Information Theory.

Alfonso Martinez (SM'11) was born in Zaragoza, Spain, in October 1973. He is currently a Ramón y Cajal Research Fellow at Universitat Pompeu Fabra, Barcelona, Spain. He obtained his Telecommunications Engineering degree from the University of Zaragoza in 1997. In 1998–2003 he was a Systems Engineer at the research centre of the European Space Agency (ESA-ESTEC) in Noordwijk, The Netherlands. His work on APSK modulation was instrumental in the definition of the physical layer of DVB-S2. From 2003 to 2007 he was a Research and Teaching Assistant at Technische Universiteit Eindhoven, The Netherlands, where he conducted research on digital signal processing for MIMO optical systems and on optical communication theory. Between 2008 and 2010 he was a post-doctoral fellow with the Information-theoretic Learning Group at Centrum Wiskunde & Informatica (CWI), in Amsterdam, The Netherlands. In 2011 he was a Research Associate with the Signal Processing and Communications Lab at the Department of Engineering, University of Cambridge, Cambridge, U.K.

His research interests lie in the fields of information theory and coding, with emphasis on digital modulation and the analysis of mismatched decoding; in this area he has coauthored a monograph on *Bit-Interleaved Coded Modulation*. More generally, he is intrigued by the connections between information theory, optical communications, and physics, particularly by the links between classical and quantum information theory.