Random–Coding Bounds for Threshold Decoders: Error Exponent and Saddlepoint Approximation

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

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

***Abstract*— This paper considers random-coding bounds to the decoding error probability with threshold decoders. A slightly improved version of the dependence-testing bound is derived. A loosening of this bound generates a family of Feinstein-like bounds, which improve on Feinstein's original version. The error exponents of these bounds are determined and simple, yet accurate, saddlepoint approximations to the corresponding error probabilities are derived.**

I. Introduction

Recently, spurred by the construction of near-capacity-achieving codes, renewed attention has been paid to the error probability in the finite-length regime. In particular, Polyanskiy *et al.* [1] have derived a number of new results, such as the random-coding union (RCU) bound, the dependence-testing bound (DT), and the $\kappa\beta$ bound among others. A key quantity in their development is the information density, defined as

$$i(\boldsymbol{x},\boldsymbol{y}) = \log\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{P_{\boldsymbol{Y}}(\boldsymbol{y})} \qquad (1)$$

where $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$ is the channel transition probability and $\boldsymbol{x},\boldsymbol{y}$ are the channel input and output sequences, respectively.

Instead of building our analysis around the concept information density, we shall assume threshold decoders using the usual maximum-likelihood (ML) metric $q(\boldsymbol{x},\boldsymbol{y}) = P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$. More precisely, we consider the Feinstein-like decoders analyzed in [1], which examine sequentially all codewords v, and output the first codeword whose metric $q(\boldsymbol{x}(v),\boldsymbol{y})$ exceeds a pre-specified threshold. Our first contribution is the optimization of a codeword- and channel-output-dependent threshold $\gamma(v,\boldsymbol{y})$ to obtain a simple, slightly improved DT bound.

The DT bound may be loosened by choosing a suboptimal threshold to obtain a family of Feinstein-type bounds. In particular, the information density is replaced by a *generalized* information density $i_s(\boldsymbol{x},\boldsymbol{y})$, given by

$$i_s(\boldsymbol{x},\boldsymbol{y}) = \log\frac{q(\boldsymbol{x},\boldsymbol{y})^s}{\mathrm{E}[q(\boldsymbol{X}',\boldsymbol{y})^s]}, \qquad (2)$$

This work has been supported by the International Joint Project 2008/R2 of the Royal Society, the FP7 Network of Excellence in Wireless Communications (NEWCOM++), the Isaac Newton Trust and the European Research Council under ERC grant agreement 259663.

A. Martinez was with Centrum Wiskunde & Informatica, Amsterdam, The Netherlands when this work was completed.

where $s \geq 0$. As noticed recently [2], the cumulant generating function of this generalized information density is closely related to Gallager's $E_0(\rho,s)$ function [3]. Indeed, for an i.i.d. codebook and a memoryless channel, we have that

$$\kappa(\tau) = \log\mathrm{E}[e^{\tau i_s(\boldsymbol{X},\boldsymbol{Y})}] \qquad (3)$$

$$= -nE_0(-\tau,s). \qquad (4)$$

Our second contribution is the determination of the error exponent attained by these bounds in terms of Gallager's $E_0(\rho,s)$ function, thereby extending Shannon's analysis of Feinstein's bound [4]. Further, as the bounds may be expressed as a tail probability of a particular random variable related to the (generalized) information density, we approximate this probability by the saddlepoint (or Laplace) method. Essentially as easy to compute as the Gaussian approximation [1], this approximation turns out to be more accurate, and thus provides an efficient method to estimate the *effective* capacity for finite block length and non-zero error probability, in particular when combined with the RCU bound [2].

Notation: Random variables are denoted by capital letters and their realization by small letters. Sequences are identified by a boldface font. The probability of an event is denoted by $\Pr\{\cdot\}$ and the expectation operator is denoted by $\mathrm{E}[\cdot]$. Logarithms are in natural units and information rates in nats, except in the examples, where bits are used.

II. Upper Bounds to the Error Probability

We adopt the conventional setup in channel coding. First, and for a given information message v, with $v \in \{1,2,\dots,M\}$, the encoder outputs a codeword of length n $\boldsymbol{x}(v) \in \mathcal{X}^n$, where $\mathcal{X}$ is the symbol channel input alphabet. One could consider more general vector alphabets and the error probability analysis remains unchanged. The coding rate R is defined as $R \triangleq \frac{1}{n}\log M$. The corresponding channel output of length n, denoted by $\boldsymbol{y} \in \mathcal{Y}^n$, where $\mathcal{Y}$ is the symbol channel output alphabet. The output sequence is generated according to the probability transition $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$. Then, the decoder computes ML decoding metrics, i. e. $q(\boldsymbol{x},\boldsymbol{y}) = P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$, and outputs a message $\hat{v}$ according to the thresholding procedure to be specified later.

We study the probability that the decoder outputs a message different from the one sent, i. e. $\Pr\{\hat{V} \neq V\}$. Specifically, we consider the average (codeword) error probability $\bar{P}_e$ over the ensemble of (randomly selected) i.i.d. codewords.

We consider general channels with ML decoding metric, i. e. $q(\boldsymbol{x},\boldsymbol{y}) = P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$. In our numerical examples, we only consider memoryless channels, for which $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} P_{Y|X}(y_i|x_i)$, with $P_{Y|X}(y|x)$ being the symbol transition probability.

A. The Dependence-Testing Bound

The DT bound was recently derived by Polyanskiy *et al.* [1, Thm. 17] by using a threshold decoder which sequentially considers all messages, and outputs the first message whose metric exceeds a pre-determined threshold. We next improve (slightly) this bound by using a message-and output-dependent threshold $\gamma(v,\boldsymbol{y})$.

For fixed message, codebook, and channel output, an error is made if the corresponding metric does not exceed the threshold, $q(\boldsymbol{X}(i),\boldsymbol{Y}) \le \gamma(i,\boldsymbol{Y})$, or if there exists an alternative codeword with lower index and metric above the threshold, $q(\boldsymbol{X}(j),\boldsymbol{Y}) > \gamma(j,\boldsymbol{Y})$, with $j < i$. Applying the union bound and averaging over all messages and codebooks, we find that the average error probability $\bar{P}_e$ is upper bounded by

$$\bar{P}_e \le \frac{1}{M}\sum_{i=1}^{M}\Big(\Pr\big\{q(\boldsymbol{X}(i),\boldsymbol{Y}) \le \gamma(i,\boldsymbol{Y})\big\} + \\ + \sum_{j<i}\Pr\big\{q(\boldsymbol{X}(j),\boldsymbol{Y}) > \gamma(j,\boldsymbol{Y})|\boldsymbol{X}(i)\big\}\Big) \quad (5)$$

$$\le \mathrm{E}\left[\frac{1}{M}\sum_{i=1}^{M}\Big(\Pr\big\{q(\boldsymbol{X}(i),\boldsymbol{Y}) \le \gamma(i,\boldsymbol{Y})|\boldsymbol{Y}\big\} + \\ + (M-i)\Pr\big\{q(\boldsymbol{X}(i),\boldsymbol{Y}) > \gamma(i,\boldsymbol{Y})|\boldsymbol{Y}\big\}\Big)\right], \quad (6)$$

where we rearranged the summation over j and explicitly wrote the expectation over $\boldsymbol{Y}$. In Eq. (6), the summands are respectively distributed according to $P_{\boldsymbol{X}|\boldsymbol{Y}}$ and $P_{\boldsymbol{X}}$.

Since the inequalities in Eq. (6) are equivalent to

$$\frac{P_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})}{P_{\boldsymbol{X}}(\boldsymbol{x})} \lessgtr \frac{\gamma(i,\boldsymbol{y})}{q(\boldsymbol{x},\boldsymbol{y})}\frac{P_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})}{P_{\boldsymbol{X}}(\boldsymbol{x})}, \quad (7)$$

the threshold $\gamma(i,\boldsymbol{y})$ may be optimized as in [1], by noting that in a hypothesis test between $P_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})$ and $P_{\boldsymbol{X}}$, the optimum threshold satisfies

$$(M-i) = \frac{\gamma(i,\boldsymbol{y})}{q(\boldsymbol{x},\boldsymbol{y})}\frac{P_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})}{P_{\boldsymbol{X}}(\boldsymbol{x})} \quad (8)$$

$$= \frac{\gamma(i,\boldsymbol{y})}{q(\boldsymbol{x},\boldsymbol{y})}\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})}{P_{\boldsymbol{Y}}(\boldsymbol{y})} \quad (9)$$

$$= \frac{\gamma(i,\boldsymbol{y})}{P_{\boldsymbol{Y}}(\boldsymbol{y})}. \quad (10)$$

Hence, the choice $\gamma(i,\boldsymbol{y}) = (M-i)P_{\boldsymbol{Y}}(\boldsymbol{y})$ gives the tighest bound. Moreover, using the relation [5, Eq. (2.132)]

$$P\left\{\frac{dP}{dQ} \le \gamma'\right\} + \gamma' Q\left\{\frac{dP}{dQ} > \gamma'\right\} = \mathrm{E}_P\left[\min\left\{1, \gamma'\frac{dQ}{dP}\right\}\right], \quad (11)$$

we may compactly rewrite the bound in Eq. (6) as

$$\bar{P}_e \le \frac{1}{M}\sum_{i}\mathrm{E}\left[\min\left\{1, (M-i)\frac{P_{\boldsymbol{Y}}(\boldsymbol{Y})}{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})}\right\}\right]. \quad (12)$$

The expectation is done according to $P_{\boldsymbol{X}}P_{\boldsymbol{Y}|\boldsymbol{X}}$.

Furthermore, since $\frac{1}{M}\sum_i \gamma(i,\boldsymbol{y}) = \frac{M-1}{2}P_{\boldsymbol{Y}}(\boldsymbol{y})$, and $\min\{1, ax\}$ is concave in x, applying Jensen's inequality relaxes Eq. (12) to the form derived in [1],

$$\bar{P}_e \le \mathrm{dtb}(n,M) \triangleq \mathrm{E}\left[\min\left\{1, \frac{M-1}{2}\frac{P_{\boldsymbol{Y}}(\boldsymbol{Y})}{P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})}\right\}\right] \quad (13)$$

$$= \mathrm{E}\Big[e^{-\left(i(\boldsymbol{X},\boldsymbol{Y}) - \log\frac{M-1}{2}\right)^{+}}\Big], \quad (14)$$

where $(x)^+ \triangleq \max\{0, x\}$.

B. Generalized Feinstein's Bound

We next relax the DT bound by applying Markov's inequality (with $s > 0$) to the second summand of the expectation in Eq. (6), namely

$$\bar{P}_e \le \mathrm{E}\left[\frac{1}{M}\sum_{i=1}^{M}\Pr\big\{q(\boldsymbol{X}(i),\boldsymbol{Y}) \le \gamma(i,\boldsymbol{Y})|\boldsymbol{Y}\big\} + \\ + (M-i)\frac{\mathrm{E}\big[q(\boldsymbol{X},\boldsymbol{Y})^s|\boldsymbol{Y}\big]}{\gamma(i,\boldsymbol{Y})^s}\right]. \quad (15)$$

Now, the choice $\gamma' = (M-i)\dfrac{\mathrm{E}[q(\boldsymbol{X},\boldsymbol{Y})^s|\boldsymbol{Y}]}{\gamma(i,\boldsymbol{Y})^s}$ gives

$$\bar{P}_e \le \mathrm{E}\left[\Pr\left\{\frac{q(\boldsymbol{X},\boldsymbol{Y})^s}{\mathrm{E}[q(\boldsymbol{X}',\boldsymbol{Y})^s|\boldsymbol{Y}]} \le \frac{M-I}{\gamma'} \;\Big|\; \boldsymbol{Y}\right\} + \gamma'\right] \quad (16)$$

$$= \Pr\left\{i_s(\boldsymbol{X},\boldsymbol{Y}) \le \log\frac{M-I}{\gamma'}\right\} + \gamma', \quad (17)$$

where I is equiprobably distributed in the set $\{1, 2, \dots, M\}$.

The classical Feinstein's bound [6] is obtained by setting $s = 1$, upper bounding the term dependent of I, $M - I$, by M (which also gives a maximal error probability version of the bound), and minimizing over γ'. For later use, we define

$$\mathrm{fb}(n,M) \triangleq \min_{\gamma'>0}\ \Pr\left\{i_s(\boldsymbol{X},\boldsymbol{Y}) \le \log\frac{M}{\gamma'}\right\} + \gamma'. \quad (18)$$

Clearly, other choices of the thresholds $\gamma(i,\boldsymbol{y})$ may result in tighter bounds, with the DT bound in the previous section dominating all these variants.

III. Error Exponents

We wish to bound the channel reliability function by finding the exponents $E_{\mathrm{dtb}}(R)$ and $E_{\mathrm{fb}}(R)$, respectively given by

$$E_{\mathrm{dtb}}(R) \triangleq \lim_{n\to\infty} -\frac{1}{n}\log \mathrm{dtb}(n,M) \quad (19)$$

$$E_{\mathrm{fb}}(R) \triangleq \sup_{s}\lim_{n\to\infty} -\frac{1}{n}\log \mathrm{fb}(n,M). \quad (20)$$

We make use of large-deviations theory, which gives the rate of exponential decay, by exploiting the close connection between the bounds and the tail probability of the rv $i_s(\boldsymbol{X},\boldsymbol{Y})$.

A. Error Exponent of the DT Bound

The identity for non-negative random variables A [1, Eq. (77)],

$$\mathrm{E}\big[\min\{1, A\}\big] = \Pr\{A \geq U\}, \tag{21}$$

where U is a uniform $(0,1)$ random variable, allows us to rewrite the DT bound as

$$\mathrm{dtb}(n, M) = \Pr\{Z \geq 0\}, \quad Z \triangleq \log\frac{M-I}{U} - i(\boldsymbol{X}, \boldsymbol{Y}), \tag{22}$$

with I equiprobably distributed in the set $\{1, 2, \dots, M\}$.

The cumulant transform $\kappa_{n,M}(\tau)$ of Z is obtained as

$$\begin{aligned}\kappa_{n,M}(\tau) &\triangleq \log \mathrm{E}\left[e^{\tau\log\frac{M-I}{U} - \tau i_s(\boldsymbol{X},\boldsymbol{Y})}\right] \qquad (23)\\ &= \log\mathrm{E}[(M-I)^\tau] - \log(1-\tau)\\ &\quad + \log\mathrm{E}\left[\left(\frac{\mathrm{E}[q(\boldsymbol{X}', \boldsymbol{Y})|\boldsymbol{Y}]}{q(\boldsymbol{X},\boldsymbol{Y})}\right)^\tau\right]. \qquad (24)\end{aligned}$$

The second summand is due to the expectation over U; hence, for the cumulant transform to converge we need $\tau < 1$.

An application of the Gärtner-Ellis theorem [7, Sec. 2.3] to the tail probability of Z gives the exponent in Eq. (19) as

$$\begin{aligned}E_{\mathrm{dtb}}(R) &= \sup_{0<\tau<1}\lim_{n\to\infty} -\frac{1}{n}\kappa_{n,M}(\tau)\\ &= \sup_{0<\tau<1}\left(\lim_{n\to\infty} -\frac{1}{n}\log\mathrm{E}\left[\left(\frac{\mathrm{E}[q(\boldsymbol{X}', \boldsymbol{Y})|\boldsymbol{Y}]}{q(\boldsymbol{X},\boldsymbol{Y})}\right)^\tau\right] - \tau R\right).\end{aligned} \tag{25}$$

Here we have approximated the message-dependent threshold term with $M - I$, a Riemann sum, as an integral,

$$\begin{aligned}\sum_{i=1}^{M}\frac{1}{M}(M-i)^\tau &= M^\tau\left(\int_0^1 (1-x)^\tau\,\mathrm{d}x + \mathrm{O}(M^{-1})\right) \qquad (26)\\ &= M^\tau\big((1+\tau)^{-1} + \mathrm{O}(M^{-1})\big), \qquad (27)\end{aligned}$$

and evaluated the limit

$$\lim_{n\to\infty}\frac{1}{n}\big(\log\mathrm{E}[(M-I)^\tau] - \log(1-\tau)\big) = \tau R. \tag{28}$$

For an i.i.d. codebook with $P_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^n P_X(x_i)$ and memoryless channel and metric, we may follow Gallager [3] and introduce the function $E_0(\rho, s)$ given by

$$E_0(\rho, s) = -\log\mathrm{E}\left[\left(\frac{\mathrm{E}[q(X', Y)^s|Y]}{q(X,Y)^s}\right)^\rho\right], \tag{29}$$

and express the exponent (25) as

$$E_{\mathrm{dtb}}(R) = \sup_{0<\rho<1}\big\{E_0(\rho, 1) - \rho R\big\}. \tag{30}$$

Clearly, the exponent may not exceed Gallager's random coding exponent, since the latter allows for $s \neq 1$.

B. Error Exponent of the Feinstein Bound

For the sake of simplicity, we focus on the memoryless channel with iid codebook; the formulas extend to a more general situation with little difficulty. Concerning Feinstein's bound, we may with no real loss of generality write γ in Eq. (18) as $e^{-n\gamma}$, so that the same reasoning applied in the previous section gives

$$\begin{aligned}&\lim_{n\to\infty} -\frac{1}{n}\log\mathrm{fb}(n, M) =\\ &\quad= \max_{\gamma>0}\,\min\Big\{\gamma, \sup_{\tau\leq 0}\big\{\tau(R+\gamma) - \kappa(\tau, s)\big\}\Big\} \qquad (31)\\ &\quad= \max_{\gamma>0}\,\min\Big\{\gamma, \sup_{\rho\geq 0}\big\{E_0(\rho, s) - \rho(R+\gamma)\big\}\Big\}, \qquad (32)\end{aligned}$$

where $\kappa(\tau, s)$ is the cumulant transform of the generalized information density and we used that $\kappa(\tau, s) = -E_0(-\rho, s)$.

For fixed γ, we consider an increasing function of γ, namely $f_1(\gamma) = \gamma$, and a non-increasing function, $f_2(\gamma) = \sup_{\rho\geq 0}\big\{E_0(\rho, s) - \rho(R+\gamma)\big\}$. These functions cross at the point $\gamma^* = \sup_{\tau\geq 0} E_0(\rho, s) - \rho(R + \gamma^*)$. It is clear that for $\gamma \leq \gamma_*$, $\min(f_1, f_2) = f_1$, and that for $\gamma \geq \gamma_*$, the reverse holds $\min(f_1, f_2) = f_2$. It follows that the lowest possible value of the maximum of f_1 and f_2 is attained precisely at γ^*, point at which either exponent is

$$\gamma^* = \sup_{\rho\geq 0}\frac{E_0(\rho, s) - \rho R}{1+\rho}. \tag{33}$$

The exponent of Feinstein's Bound obtains by optimizing over ρ (and implicitly γ),

$$\lim_{n\to\infty} -\frac{1}{n}\log\mathrm{fb}(n, M) = \sup_{\rho\geq 0}\frac{E_0(\rho, s) - \rho R}{1+\rho}. \tag{34}$$

A similar result was obtained by Shannon for Feinstein's classical bound, with $s = 1$ [4].

Following Gallager [3], the optimum choice is $s = \frac{1}{1+\hat{\rho}}$. Defining a new variable $\rho' = \rho/(1+\rho)$ (or $\rho = \rho'/(1-\rho')$), with $0 \leq \rho' \leq 1$, we rewrite $E_{\mathrm{fb}}(R)$ as

$$E_{\mathrm{fb}}(R) = \sup_{0\leq\rho'\leq 1}(1-\rho')E_0\left(\frac{\rho'}{1-\rho'}, 1-\rho'\right) - \rho' R. \tag{35}$$

Lastly, we remark that, by construction, $E_{\mathrm{fb}}(R) \leq E_{\mathrm{dtb}}(R)$.

IV. Saddlepoint Approximations

A. Motivation

Chernoff-type bounds provide an estimate of the tail probability via the cumulant transform, namely $\Pr\{Z \geq \varepsilon\} \sim e^{\kappa(\hat{\tau}) - \hat{\tau}\varepsilon}$, with $\hat{\tau} = \arg\min_\tau\big\{\kappa(\tau) - \tau\varepsilon\big\}$. Yet, clearly, a more accurate estimate would be of the form $\Pr\{Z \geq \varepsilon\} \sim \alpha(\kappa, \hat{\tau}) \cdot e^{\kappa(\hat{\tau}) - \hat{\tau}\varepsilon}$. Saddlepoint approximations provide such estimates [8], and have recently been used to approximate a version of the random-coding union bound [2].

B. Approximation to the DT Bound

Using the identity in Eq. (21) we may express the DT bound in Eq. (12) the as the tail probability of a continuous random variable $Z = \log \frac{M-I}{U} - i(\boldsymbol{X}, \boldsymbol{Y})$, whose cumulant transform we denoted by $\kappa_{n,M}(\tau)$. The parameter τ is a complex number for the purpose of deriving the saddlepoint approximation.

As the cumulant transform is the Laplace transform of the probability density function $p_Z(z)$, the density function itself is expressable as an inverse Laplace transform [8], namely

$$p_Z(z) = \frac{1}{2\pi j} \int_{\hat{\tau}-j\infty}^{\hat{\tau}+j\infty} e^{\kappa_{n,M}(\tau)-\tau z}\, \mathrm{d}\tau, \tag{36}$$

where $\hat{\tau} < 1$ from the definition of $\kappa_{n,M}$. Since $\bar{P}_e$ is the tail above $\varepsilon = 0$, we compute it by integrating over $z \in [0, \infty)$. Changing the integration order, we get

$$\mathrm{dtb}(n, M) = \frac{1}{2\pi j} \int_{\hat{\tau}-j\infty}^{\hat{\tau}+j\infty} \int_0^\infty e^{\kappa_{n,M}(\tau)-\tau z}\, \mathrm{d}z\, \mathrm{d}\tau \tag{37}$$

$$= \frac{1}{2\pi j} \int_{\hat{\tau}-j\infty}^{\hat{\tau}+j\infty} e^{\kappa_{n,M}(\tau)} \left(\frac{e^{-\tau z}}{-\tau} \bigg|_0^\infty \right) \mathrm{d}\tau \tag{38}$$

$$= \frac{1}{2\pi j} \int_{\hat{\tau}-j\infty}^{\hat{\tau}+j\infty} e^{\kappa_{n,M}(\tau)} \frac{1}{\tau}\, \mathrm{d}\tau, \tag{39}$$

where $\hat{\tau} > 0$ to guarantee convergence. For memoryless channels, and substituting the form of $\kappa_{n,M}(\tau)$ we get

$$\mathrm{dtb}(n, M) \simeq \frac{1}{2\pi j} \int_{\hat{\rho}-j\infty}^{\hat{\rho}+j\infty} e^{n(\rho R - E_0(\rho,1))} \frac{1}{\rho(1-\rho^2)}\, \mathrm{d}\rho. \tag{40}$$

We next expand the exponent in the integrand as a Taylor series around $\hat{\rho} = \min(1, \hat{\rho}_0)$, with $\hat{\rho}$ given by the root of $\hat{E}_0'(\hat{\rho}_0, 1) = R = \frac{1}{n}\log M$ (it is safe to replace $M-1$ by M here). Neglecting terms of order higher than 2 and up to a common factor n, we get

$$\rho R - E_0(\rho, 1) \sim \hat{\rho} R - E_0(\hat{\rho}, 1) + \big(R - E_0'(\hat{\rho}, 1)\big)(\rho - \hat{\rho}) - \frac{1}{2} E_0''(\hat{\rho}, 1)(\rho - \hat{\rho})^2. \tag{41}$$

Let us define the following parameters V and W:

$$V \triangleq -E_0''(\hat{\rho}, 1), \quad W \triangleq R - E_0'(\hat{\rho}, 1). \tag{42}$$

We have $V \geq 0$ and in general $V > 0$. We also have $W = 0$ if $\hat{\rho} \leq 1$. Beyond a critical rate, however, $W \neq 0$.

We proceed further by replacing the exponent in the integrand of Eq. (40) by Eq. (41) and expanding the term $(\rho(1-\rho^2))^{-1}$ into a sum of partial fractions. In the standard saddlepoint approximation to the tail probability, which has only a term ρ^{-1}, this term is replaced by $\hat{\rho}^{-1}$, unless $\hat{\rho}$ is close to zero, in which case the term ρ^{-1} is kept and integrated over. By analogy, and taking into account that the range of ρ is limited to $(0, 1)$, we use the approximation $(1-\rho^2) = (1+\rho)(1-\rho) \simeq (1+\hat{\rho})(1-\rho)$, and decompose the resulting fraction into a sum of simple fractions:

$$\frac{1}{\rho(1-\rho^2)} \simeq \frac{1}{\rho(1+\hat{\rho})} + \frac{1}{(1-\rho)(1+\hat{\rho})}. \tag{43}$$

Carrying out the integrals, we thus obtain our desired saddlepoint approximation

$$\mathrm{dtb}(n, M) \simeq \alpha_{\mathrm{dtb}}(n, M) \cdot e^{-n(\hat{E}_0(\hat{\rho},1) - \hat{\rho} R)} \tag{44}$$

$$\alpha_{\mathrm{dtb}}(n, M) = \frac{1}{2(1+\hat{\rho})} \left(\mathrm{erfcx}_1\left(\hat{\rho}\sqrt{\frac{nV}{2}}, W\sqrt{\frac{n}{2V}} \right) + {} + \mathrm{erfcx}_1\left((1-\hat{\rho})\sqrt{\frac{nV}{2}}, -W\sqrt{\frac{n}{2V}} \right) \right), \tag{45}$$

where we used the function $\mathrm{erfcx}_1(x, y) \triangleq \mathrm{erfcx}(x - y)\exp(-y^2) = \mathrm{erfc}\,(x-y)\exp(x^2 - 2xy)$.

This analysis extends to the DT bound in Eq. (14). The approximation is as in Eq. (45), with $1/(1+\hat{\rho})$ replaced by $2^{-\hat{\rho}}$, and with $W = R - \hat{E}_0'(\hat{\rho}, s) - \frac{1}{n}\log 2$. An approximation of the gain of the DT bound in Eq. (12) with respect to that in Eq. (14) can be obtained from the respective saddlepoint approximations. In particular, the gain can be approximated as $\frac{2^{-\hat{\rho}}}{1+\hat{\rho}}$ which suggests a maximum gain of about 6%.

C. Approximation to the Feinstein Bound

Since Feinstein's bound is expressed as a function of the tail probability of the generalized information density $i_s(\boldsymbol{X}, \boldsymbol{Y})$, it is also amenable to the saddlepoint approximation. With the approximation that the information density can be modeled as a continuous random variable and substituting for the proper cumulant transform, the steps we applied to approximation the DT bound in the previous section give here

$$\Pr\left\{ i_s(\boldsymbol{X}, \boldsymbol{Y}) < \log \frac{M}{\gamma} \right\} \simeq \alpha_0\, \gamma^{-\hat{\rho}} e^{-n(\hat{E}_0(\hat{\rho},s) - \hat{\rho} R)} \tag{46}$$

$$\alpha_0 = \frac{1}{2}\mathrm{erfcx}_1\left(\hat{\rho}\sqrt{\frac{nV}{2}}, \frac{\sqrt{n}W}{\sqrt{2V}} \right), \tag{47}$$

where we have defined

$$\hat{\rho} = \arg\max_{\rho \geq 0} \frac{\hat{E}_0(\rho, s) - \rho R}{1+\rho} \tag{48}$$

$$V \triangleq -\hat{E}_0''(\hat{\rho}, s) \tag{49}$$

$$W \triangleq R - \hat{E}_0'(\hat{\rho}, s) - \frac{1}{n}\log\gamma. \tag{50}$$

As for the value of γ, a good choice will turn out to be

$$\gamma \triangleq \left(\frac{\hat{\rho}}{2}\mathrm{erfcx}\left(\hat{\rho}\sqrt{\frac{nV}{2}} \right) e^{-n(E_0(\hat{\rho},s) - \hat{\rho} R)} \right)^{\frac{1}{1+\hat{\rho}}}, \tag{51}$$

for which we have that

$$\mathrm{fb}(n, M) \simeq \alpha_{\mathrm{fb}}(n, M) \cdot e^{\frac{-n(\hat{E}_0(\hat{\rho},s) - \hat{\rho} R)}{1+\rho}} \tag{52}$$

$$\alpha_{\mathrm{fb}}(n, M) = \left(\frac{\hat{\rho}}{2}\mathrm{erfcx}\left(\hat{\rho}\sqrt{\frac{nV}{2}} \right) \right)^{\frac{1}{1+\hat{\rho}}} \times \times \left(1 + \frac{\mathrm{erfcx}_1\left(\hat{\rho}\sqrt{\frac{nV}{2}}, \frac{\sqrt{n}W}{\sqrt{2V}} \right)}{\hat{\rho}\,\mathrm{erfcx}\left(\hat{\rho}\sqrt{\frac{nV}{2}} \right)} \right). \tag{53}$$

As $n \to \infty$ we have that $-\frac{1}{n}\log\gamma \to \gamma^*$, where γ^*, whose value is given in Eq. (33), was used to find of the exponent in Sect. III-B. Both choices are therefore consistent. Moreover, assuming that $W \simeq 0$, we note that the approximation has the form $g(\gamma) \triangleq \alpha_0 \gamma^{-\hat{\rho}} e^{-n(\hat{E}_0(\hat{\rho},s)-\hat{\rho}R)} + \gamma$, for which the optimum value of γ can be found by setting $g'(\gamma) = 0$ and solving for γ. This gives the value of γ in Eq. (51), again suggesting this is a good choice.

V. Applications

In this section we apply the bounds and approximations to the binary-input complex AWGN channel with signal-to-noise ratio snr, whose corresponding Gallager function is given by

$$E_0^{\text{bpsk}}(\rho, s) = -\log \mathbb{E}\left[\left(\frac{1}{2}\left(1 + \mathrm{e}^{-4s(\mathsf{snr}+\sqrt{\mathsf{snr}}XZ)}\right)\right)^{\rho}\right], \tag{54}$$

where $X \in \{-1, 1\}$ and $Z \sim \mathcal{N}(0, 1)$.

Fig. 1 shows the values of the exponents for the various bounds considered in this paper and Gallager's exponent (which is also the exponent of the Markov-RCU bound analyzed in [2]). As expected, allowing for an optimum s leads to a significant improvement in exponent, both between the RCU and the DT bounds, and between the Feinstein bound in Eq. (18) and the classical Feinstein bound. Somewhat intriguingly, the improvement in exponent between the DT bound and the generalized Feinstein bound is rather small.

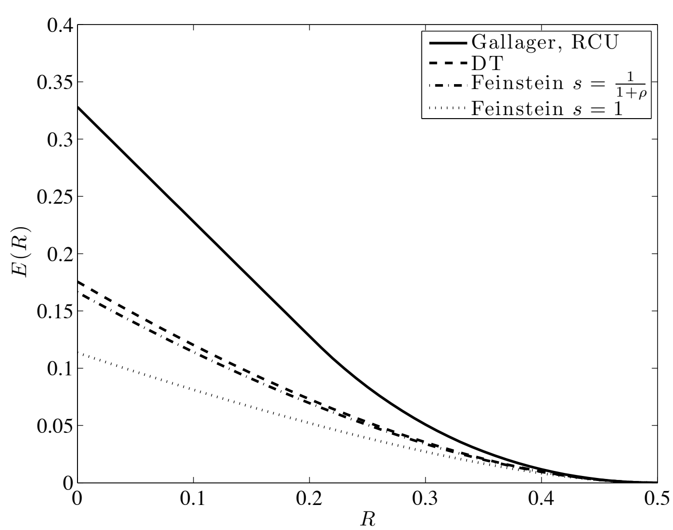

Fig. 1. Exponents for the binary-input AWGN channel; $\mathsf{snr} = -2.823$ dB.

Fig. 2 compares the proposed saddlepoint approximations to the DT and Feinstein bounds with $s = 1$ and $s = \frac{1}{1+\hat{\rho}}$ and the RCU bound. For this channel, the bounds cannot be computed exactly, and therefore, the corresponding simulations are reported. First, we observe an excellent match between bounds and saddlepoint approximations. Overall, the DT incurs a significant loss with respect to the RCU bound, mainly due to the loss in error exponent; the RCU has Gallager's exponent since it uses $s = \frac{1}{1+\hat{\rho}}$, while the DT has a worse exponent due to using $s = 1$. Even though the Feinstein bound exponent with $s = \frac{1}{1+\hat{\rho}}$ is close to that of the DT bound, the error probability is significantly worse; obviously the standard Feinstein bound performs the worst. Nonetheless, all bounds are sufficient to show the achievability of the most general expression for channel capacity [9].

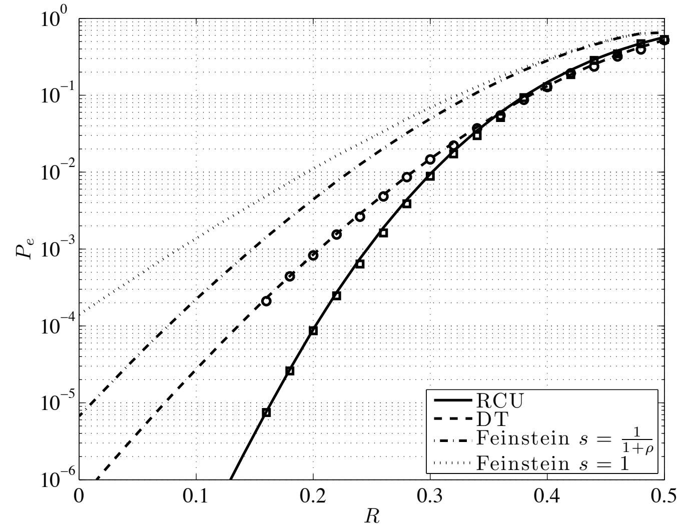

Fig. 2. Comparison of DT, Feinstein, and RCU bounds (simulation and saddlepoint approximations) for $n = 100$ and a binary-input AWGN channel with $\mathsf{snr} = -2.823$ dB.

VI. Conclusions

In this paper, we have derived a new version of the dependence-testing bound, and used it to find a family of Feinstein-type bounds which improve on the usual formulation of the Feinstein bound. We then computed the error exponents and saddlepoint approximations to these bounds. These saddlepoint approximations are a versatile tool that allow to accurately calculate the corresponding bounds for arbitrary discrete-input memoryless channels with a complexity similar to that of the error exponent or the Gaussian approximation.

References

[1] Y. Polyanskiy, H. V Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Th.*, vol. 45, no. 5, pp. 2307–2359, May 2010.

[2] A. Martinez and A Guillén i Fàbregas, "Saddlepoint Approximation of Random-Coding Bounds," in *Proc. Information Theory and Applications (ITA) Workshop, San Diego (USA)*, Feb 2011.

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

[4] C. E. Shannon, "Certain Results in Coding Theory for Noisy Channels," *Inf. Control*, pp. 6–25, 1957.

[5] Y. Polyanskiy, *Channel coding: non-asymptotic fundamental limits*, Ph.D. thesis, Princeton University, 2010.

[6] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inf. Theory*, vol. 4, no. 4, pp. 2–22, 1954.

[7] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Springer, 1998.

[8] J. L. Jensen, *Saddlepoint Approximations*, Oxford University Press, USA, 1995.

[9] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, 1994.