A Complex-Integration Approach to the Saddlepoint Approximation for Random–Coding Bounds

Alfonso Martinez
Universitat Pompeu Fabra (UPF)
alfonso.martinez@ieee.org

Jonathan Scarlett
University of Cambridge
jms265@cam.ac.uk

Marco Dalai
Università degli Studi di Brescia
marco.dalai@unibs.it

Albert Guillén i Fàbregas
ICREA & UPF
University of Cambridge
guillen@ieee.org

Abstract—This paper derives a saddlepoint approximation for the random-coding bound to the error probability of channel coding by using complex-integration techniques. The approximation is given by a sum of two terms: one with Gallager's exponent, and a second one with Arimoto's strong converse exponent (above capacity) or the sphere-packing exponent (below the critical rate).

I. INTRODUCTION

Two of the fundamental problems in channel coding are finding the error probability attained by a given code and characterizing the largest achievable rate compatible with vanishing error probability. The first problem is more relevant for practical transmission at finite blocklengths, while the second focuses in the asymptotic regime of large blocklengths. Since the latter regime allows for the use of analytic tools such as random coding, it had traditionally been favored in information theory. In the past years, however, spurred by the construction of near-capacity achieving codes and the interest in short-duration wireless communications, renewed attention has been paid to Strassen's Gaussian approximation to the *effective* channel capacity [1]. This approximation has the form

$$R_n(\epsilon) = C - \sqrt{\frac{V}{n}} \mathsf{Q}^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right), \qquad (1)$$

where $R_n(\epsilon)$ is the effective capacity for finite blocklength n and fixed error probability ϵ, C is the capacity, V the channel dispersion, $\mathsf{Q}(x)$ the Gaussian tail probability function, and the term $o\left(\frac{1}{\sqrt{n}}\right)$ is asymptotically negligible [1], [2], [3]. In terms of the error probability, Eq. (1) translates into

$$\epsilon \approx \mathsf{Q}\left(\frac{\sqrt{n}(C - R_n)}{\sqrt{V}}\right). \qquad (2)$$

While (1) accurately estimates the asymptotic behaviour of R_n for fixed ϵ, Eq. (2) is less precise for fixed rate R. Significantly, Eq. (2) does not have the correct exponential decay in n, which is given for rates above the critical rate by

$$\epsilon \leq e^{-n(E_0(\rho,s) - \hat{\rho}R)}, \qquad (3)$$

where $E_0(\rho, s)$ is Gallager's function and ρ and s are selected to minimize the exponent [4]. The saddlepoint approximation provides an alternative refined estimate to (2) and (3), namely

$$\epsilon \lesssim \alpha_n \, e^{-n(E_0(\rho,s) - \hat{\rho}R_n)} \bigl(1 + o(1)\bigr), \qquad (4)$$

for some coefficient α_n [5]. The aproximation in (4) has the correct exponential decay in n for fixed R, i. e. that of (3), and recovers the Gaussian approximation as R_n tends to the capacity and ϵ is fixed. Overall, it yields an efficient method to estimate the effective capacity $R_n(\epsilon)$ for finite n and ε.

The goal of this paper is to derive a refined form of (4) for rates beyond the capacity and the critical rate. We start in Sect. II by reviewing the refined random-coding union bound to the average error probability, recently proposed by Polyanskiy *et al.* [2] and derive a more tractable, weakened form by applying Markov's inequality. Then, in Sect. III, we determine the rate of exponential decay of this bound with the blocklength in terms of Gallager's $E_0(\rho, s)$ function, as in (3). Then, Sect. IV, the core of the paper, uses complex-integration techniques to derive a novel refined form of (4). In contrast with [5], where general channels and input distributions are considered, we limit our attention to continuous channels.

Notation: Random variables are represented by capital letters and their realizations by small letters. Sequences are identified by boldface font and their components by a subindex, e. .g. $\boldsymbol{x}$ and x_i. We denote the probability of an event by $\Pr\{\cdot\}$ and the expectation by $\mathrm{E}[\cdot]$. We may use a subindex in the event probability or the expectation to explicitly refer to the relevant random variables, e. g. $\Pr_A\{\cdot\}$. Logarithms are in natural units and information rates in nats.

II. UPPER BOUNDS TO THE ERROR PROBABILITY

We consider coding over memoryless channels with input x, output y, and channel transition law $W(y|x)$.

Encoder: First, and for a given information message v, with $v \in \{1, 2, \ldots, M_n\}$, the encoder outputs a codeword $\mathbf{x}$ of length n, that is $\mathbf{x} = (x_1, \ldots, x_n)$. The coding rate R_n is defined as $R_n \triangleq \frac{1}{n} \log M_n$.

Channel: The corresponding channel output of length n, denoted by $\mathbf{y} = (y_1, \ldots, y_n)$, is obtained from the input sequence according to the channel transition probability $W^n(\mathbf{y}|\mathbf{x}) \triangleq \prod_{i=1}^{n} W(y_i|x_i)$.

Decoder: Finally, the decoder selects the message $\hat{v}$ with largest decoding metric $q^n(\mathbf{x}, \mathbf{y})$, where the metric is assumed to be maximum-likelihood, i. e. $\hat{v} = \arg\max_v W^n(\mathbf{y}|\mathbf{x}(v))$, and $\mathbf{x}(v)$ is the codeword associate to message v.

We study the probability that the decoder outputs a message different from the one sent; we denote this average error probability for a given code by ϵ and express it as

$$\epsilon \triangleq \sum_{v=1}^{M_n} \frac{1}{M_n} \Pr\{v \neq \hat{v}\}. \tag{5}$$

We determine achievable values of the error probability by studying ensembles of codebooks whose M_n codewords are independently selected according to a probability distribution $P^n(\mathbf{x}) = \prod_{i=1}^n P(x_i)$, for some $P(x)$. For an average error probability within the ensemble $\bar{\epsilon}$, there exists at least one code with M_n codewords such that $\epsilon \leq \bar{\epsilon}$. Similarly, standard expurgation arguments show the existence of codes with $\frac{1}{2}M_n$ codewords whose maximal error probability at most $2\bar{\epsilon}$.

The random coding union (RCU) bound to the average error probability under ML decoding [2] is given by

$$\epsilon \leq \mathrm{rcu}(n, M_n) \tag{6}$$
$$\triangleq \mathrm{E}_{\mathbf{X}, \mathbf{Y}}\Big[\min\Big\{1, (M_n - 1)\Pr_{\bar{\mathbf{X}}}\{q^n(\bar{\mathbf{X}}, \mathbf{Y}) \geq q^n(\mathbf{X}, \mathbf{Y})\}\Big\}\Big], \tag{7}$$

where the random variables have the respective distributions $(\mathbf{X}, \mathbf{Y}) \sim P^n(\mathbf{x})W^n(\mathbf{y}|\mathbf{x})$ and $\bar{\mathbf{X}} \sim P^n(\bar{\mathbf{x}})$.

We may upperbound the inner probability in (7) by using Markov's inequality, $\Pr\{A \geq \varepsilon\} \leq \frac{1}{\varepsilon}\mathrm{E}[A]$ for non-negative A; and in fact a tighter bound is obtained by applying Markov's inequality to $\Pr\{A^s \geq \varepsilon^s\}$, for some $s \geq 0$. This gives

$$\epsilon \leq \mathrm{rcu}_s(n, M_n) \tag{8}$$
$$\triangleq \mathrm{E}_{\mathbf{X}, \mathbf{Y}}\left[\min\left\{1, (M_n - 1)\frac{\mathrm{E}_{\bar{\mathbf{X}}}[q^n(\bar{\mathbf{X}}, \mathbf{Y})^s]}{q^n(\mathbf{X}, \mathbf{Y})^s}\right\}\right]. \tag{9}$$

Moreover, exploiting the identity[1] where we let U be uniformly distributed in $(0, 1)$. we may rewrite $\mathrm{rcu}_s(n, M_n)$ in (9) as

$$\mathrm{rcu}_s(n, M_n) = \Pr_{\mathbf{X}, \mathbf{Y}, U}\left\{\frac{M_n - 1}{U}\frac{\mathrm{E}_{\bar{\mathbf{X}}}[q^n(\bar{\mathbf{X}}, \mathbf{Y})^s]}{q^n(\mathbf{X}, \mathbf{Y})^s} \geq 1\right\} \tag{15}$$
$$= \Pr_{\mathbf{X}, \mathbf{Y}, U}\{Z_n \geq 0\}, \tag{16}$$

[1] Let us define $W \triangleq \min(1, A)$ and let $F_W(w)$ denote its cdf. It holds that $F_W(w) = F_A(w)$ for $w \leq 1$ The expected value of W is therefore

$$\mathrm{E}[W] = \int_0^1 w \, dF_W(w) \tag{10}$$
$$= wF_W(w)|_0^1 - \int_0^1 F_W(w) \, dw \tag{11}$$
$$= 1 - \int_0^1 F_A(w) \, dw \tag{12}$$
$$= 1 - \int_0^1 \big(1 - \Pr_A\{A > w\}\big) \, dw \tag{13}$$
$$= \Pr_{A,U}\{A \geq U\}, \tag{14}$$

where we took logarithms and let $Z_n \triangleq \log \frac{M_n - 1}{U} - i_s^n(\mathbf{X}, \mathbf{Y})$. Here, $i_s^n(\mathbf{X}, \mathbf{Y})$ denotes a generalized information density,

$$i_s^n(\mathbf{x}, \mathbf{y}) = \log \frac{q(\mathbf{x}, \mathbf{y})^s}{\mathrm{E}_{\bar{\mathbf{X}}}[q(\bar{\mathbf{X}}, \mathbf{y})^s]}, \tag{17}$$

for $s \geq 0$. We have $i_s^n(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n i_s(x_i, y_i)$, where $i_s(x, y)$ is the corresponding symbol information density,

$$i_s(x, y) = \log \frac{q(x, y)^s}{\mathrm{E}_{\bar{X}}[q(\bar{X}, y)^s]}, \tag{18}$$

with $q(x, y)$ a symbol decoding metric. We focus on Maximum-Likelihood decoding, i. e. $q(x, y) = W(y|x)$, although the analysis holds in more generality.

Our analysis stems from this characterization of rcu_s as the tail probability of a random variable Z_n. As is usual in this context, the cumulant generating function (cgf) of Z_n is of critical importance. In this context, we observe that the cgf of $i_s(X, Y)$ is closely related to Gallager's $E_0(\rho, s)$ function for distribution P and non-negative s [4], which is given by:

$$E_0(\rho, s) = -\log \mathrm{E}_{X,Y}\left[\left(\frac{\mathrm{E}_{\bar{X}}[q(\bar{X}, Y)^s]}{q(X, Y)^s}\right)^\rho\right]. \tag{19}$$

Indeed, if we express the codeword metric as $q^n(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^n q(x, y)$, we then have the relationship

$$\log \mathrm{E}[e^{-\rho i_s^n(\mathbf{X}, \mathbf{Y})}] = -nE_0(\rho, s). \tag{20}$$

III. EXPONENTIAL DECAY OF THE ERROR PROBABILITY

In this section we briefly deal with the rate of exponential decay and review how to recover the bound (3). The starting point is the Chernoff bound to the tail probability of a random variable Z_n, $\Pr\{Z_n \geq 0\}$. The bound is given by

$$\Pr\{Z_n \geq 0\} \leq \inf_{\rho \geq 0} \mathrm{E}[e^{\rho Z_n}] = \inf_{\rho \geq 0}\{e^{\kappa_n(\rho)}\}, \tag{21}$$

with $\kappa_n(\rho) \triangleq \log \mathrm{E}[e^{\rho Z_n}]$. It follows that the rate of exponential decay of the probability $\Pr\{Z_n \geq 0\}$ is bounded as

$$\lim_{n \to \infty} -\frac{1}{n} \log \Pr\{Z_n \geq 0\} \geq \lim_{n \to \infty} \frac{1}{n} \sup_{\rho \geq 0}\{-\kappa_n(\rho)\}. \tag{22}$$

We use this identity to find the rate of exponential decay of the bound rcu_s $E_s(R)$ for a fixed rate R such that $M_n = e^{nR}$:

$$E_s(R) \triangleq \lim_{n \to \infty} -\frac{1}{n} \log \mathrm{rcu}_s(n, M_n) \tag{23}$$
$$= \lim_{n \to \infty} -\frac{1}{n} \log \Pr\left\{\log \frac{M_n - 1}{U} - i_s^n(\mathbf{X}, \mathbf{Y}) \geq 0\right\}. \tag{24}$$

For memoryless channels and codebooks generated with a distribution $P(x)$, we can evaluate the function $\kappa_{n,s}(\rho)$ as

$$\kappa_{n,s}(\rho) \triangleq \log \mathrm{E}_{\mathbf{X}, \mathbf{Y}, U}[e^{\rho \log \frac{M_n - 1}{U} - \rho i_s^n(\mathbf{X}, \mathbf{Y})}] \tag{25}$$
$$= \rho \log(M_n - 1) - \log(1 - \rho) - nE_0(\rho, s), \tag{26}$$

for $\rho < 1$. With this function, we lowerbound the limit in (24),

$$\lim_{n\to\infty} \sup_{0\le\rho<1} -\frac{1}{n}\big(\rho\log(M_n-1) - \log(1-\rho) - nE_0(\rho,s)\big) \quad (27)$$

$$= \sup_{0\le\rho<1}\big\{E_0(\rho,s) - \rho R\big\} \quad (28)$$

$$= \sup_{0\le\rho\le1}\big\{E_0(\rho,s) - \rho R\big\}, \quad (29)$$

where (28) follows from the fact that the contribution from $\log(1-\rho)$ vanishes asymptotically and a definition of the rate R as $R = \lim_{n\to\infty} R_n$; in (29) we used continuity of the E_0 function to extend the optimization range to include $\rho = 1$. Finally, optimization over the remaining parameter[2] s recovers Gallager's random-coding exponent $E_r(R,P)$,

$$E_r(R,P) = \sup_{s\ge0,0\le\rho\le1}\big\{E_0(\rho,s) - \rho R\big\}. \quad (30)$$

At rates below the mutual information $I(P)$, the exponent in (30) is positive and transmission of information with vanishing error probability can be achieved. Another important rate is the critical rate $R^*(P)$, which we define as the largest rate for which the random-coding exponent is achieved at $\hat\rho = 1$.

IV. SADDLEPOINT APPROXIMATION FOR THE ERROR PROBABILITY

While the parameter ρ is real-valued in the Chernoff bound, it may best be seen as a complex number for the purpose of deriving the saddlepoint approximation. The cumulant transform is then the Laplace transform of the probability density function $p_{Z_n}(z)$, and the density function itself may be computed as an inverse Laplace transform [6], [7], namely

$$p_{Z_n}(z) = \frac{1}{2\pi j}\int_{\rho_0-j\infty}^{\rho_0+j\infty} e^{\kappa_{n,s}(\rho) - \rho z}\,\mathrm{d}\rho, \quad (31)$$

where $\rho_0 < 1$ from the definition of $\kappa_{n,s}$. We assume that the information density $i_s(X,Y)$ has a density and that $e^{\kappa_{n,s}(\rho)}$ is absolutely integrable so that the inversion formula (31) applies. Among others, this rules out discrete channels, although extensions along the lines of [5] are possible.

Since rcu_s is given by the tail above $\varepsilon = 0$, we compute its value by integrating over $z \in [0,\infty)$,

$$\mathrm{rcu}_s(n,M_n) = \int_0^\infty \frac{1}{2\pi j}\int_{\rho_0-j\infty}^{\rho_0+j\infty} e^{\kappa_{n,s}(\rho) - \rho z}\,\mathrm{d}\rho\,\mathrm{d}z \quad (32)$$

$$= \frac{1}{2\pi j}\int_{\rho_0-j\infty}^{\rho_0+j\infty}\int_0^\infty e^{\kappa_{n,s}(\rho) - \rho z}\,\mathrm{d}z\,\mathrm{d}\rho \quad (33)$$

$$= \frac{1}{2\pi j}\int_{\rho_0-j\infty}^{\rho_0+j\infty} e^{\kappa_{n,s}(\rho)}\left(\frac{e^{-\rho z}}{-\rho}\bigg|_0^\infty\right)\mathrm{d}\rho \quad (34)$$

$$= \frac{1}{2\pi j}\int_{\rho_0-j\infty}^{\rho_0+j\infty} e^{\kappa_{n,s}(\rho)}\frac{\mathrm{d}\rho}{\rho}, \quad (35)$$

where we interchanged the integration order in (33) and required that $0 < \rho_0 < 1$ to guarantee convergence in (35). It will prove convenient to define $I_n(w)$, for real arbitrary w, as

$$I_n(w) \triangleq \frac{1}{2\pi j}\int_{w-j\infty}^{w+j\infty} e^{n(\rho R - E_0(\rho,s))}\frac{\mathrm{d}\rho}{\rho(1-\rho)}, \quad (36)$$

so that

$$\mathrm{rcu}_s(n,M_n) = I_n(\rho_0). \quad (37)$$

The critical points (poles and saddlepoints) of the integrand are of fundamental importance in our analysis. Eq. (36) has two poles, located at $\rho = 0$ and $\rho = 1$, and a saddlepoint at $\rho = \hat\rho$, a minimizer[3] of the exponent for fixed R and s,

$$\hat\rho \triangleq \arg\min_{-\infty<\rho<\infty}\big\{\rho R - E_0(\rho,s)\big\}. \quad (38)$$

Since the exponent is a convex function of ρ [6], the root $\hat\rho$ is given by the unique solution of the equation

$$E_0'(\hat\rho,s) = \frac{\partial E_0(\rho,s)}{\partial\rho}\bigg|_{\rho=\hat\rho} = R. \quad (39)$$

From here onwards, we let s denote the value that minimizes the exponent at the optimum $\hat\rho$. For each rate R and associated saddlepoint $\hat\rho$, we define a rate-dependent dispersion $V_{\hat\rho}$ as

$$V_{\hat\rho} \triangleq -\frac{\partial^2 E_0(\rho,s)}{\partial\rho^2}\bigg|_{\rho=\hat\rho} = -E_0''(\hat\rho,s). \quad (40)$$

At $R = I(P)$, we have $V_{\hat\rho} = V$, the channel dispersion. We assume that E_0 is strictly convex and the dispersion is therefore positive. This imposes no real limitation since the information density is constant otherwise and rcu_s is trivial.

As we saw in Sect. III, the saddlepoint satisfies $\hat\rho \in (0,1)$ for rates in the range $R^*(P) < R < I(P)$. Then, substituting $\kappa_{n,s}(\rho)$ in (26), we can set $\rho_0 = \hat\rho$ in (37) and get

$$\mathrm{rcu}_s(n,M_n) = I_n(\hat\rho). \quad (41)$$

If the rate R is such that $\hat\rho$ lies outside the interval of convergence of (35), Cauchy's residue theorem[4] allows to shift the integration axis to $\hat\rho$ at the cost of introducing additional terms in (41) [7, Ch. 26]. For simple poles ρ_k and Γ oriented counterclockwise in the complex plane, the theorem reads

$$\frac{1}{2\pi j}\oint_\Gamma f(\rho)\,\mathrm{d}\rho = \sum_{\rho_k}\lim_{\rho\to\rho_k}(\rho-\rho_k)f(\rho). \quad (42)$$

If $\hat\rho < 0$, we choose Γ as a rectangle oriented counterclockwise with vertices $\{\rho_0 \pm jT, \hat\rho \pm jT)\}$, for large real-valued T. For $f(\rho) = e^{\kappa_{n,s}(\rho)}\rho^{-1}$, we get

$$\frac{1}{2\pi j}\oint_\Gamma f(\rho)\,\mathrm{d}\rho = \lim_{\rho\to0}\rho\frac{e^{n(\rho R - E_0(\rho,s))}}{\rho(1-\rho)} \quad (43)$$

$$= 1. \quad (44)$$

The contributions from the horizontal sides of the rectangle vanish as $T \to \infty$, since $i_s(X,Y)$ is assumed to have a density and therefore $e^{n(zR - E_0(z,s))}$ decays fast enough at infinity by the Riemann-Lebesgue Lemma [6]. The remaining two sides on Γ correspond to $I_n(\rho_0)$ and $I_n(\hat\rho)$, and we therefore obtain

$$\lim_{T\to\infty}\frac{1}{2\pi j}\oint_\Gamma f(\rho)\,\mathrm{d}\rho = I_n(\rho_0) - I_n(\hat\rho). \quad (45)$$

[2]The function E_0 is maximized for the choice $\hat s = \frac{1}{1+\rho}$ for $\rho > -1$.

[3]This point is a saddlepoint because the function in the exponent has a maximum at $\hat\rho$ along the imaginary axis with real part $\hat\rho$.

[4]This theorem expresses the integral of an analytic function $f(\rho)$ over a closed curve Γ in terms of the residues at the poles ρ_k enclosed by Γ.

Equivalently, combining Eqs. (42), (44), and (45), we obtain

$$\mathrm{rcu_s}(n, M_n) = 1 + I_n(\hat{\rho}). \tag{46}$$

Analogously, if $\hat{\rho} > 1$, we get

$$I_n(\hat{\rho}) - I_n(\rho_0) = \lim_{\rho \to 1} (\rho - 1) \frac{e^{n(\rho R - E_0(\rho, s))}}{\rho(1 - \rho)} \tag{47}$$

$$= -e^{n(R - E_0(1, s))}, \tag{48}$$

instead of (44) and (45). And using again (42), we obtain

$$\mathrm{rcu_s}(n, M_n) = e^{n(R - E_0(1, s))} + I_n(\hat{\rho}). \tag{49}$$

To evaluate (36), we use the identity

$$\frac{1}{\rho(1 - \rho)} = \frac{1}{\rho} + \frac{1}{1 - \rho} \tag{50}$$

to express $I_n(\hat{\rho})$ as $I_n(\hat{\rho}) = I_{n,0}(\hat{\rho}) + I_{n,1}(\hat{\rho})$, with

$$I_{n,0}(w) \triangleq \frac{1}{2\pi j} \int_{w-j\infty}^{w+j\infty} e^{n(\rho R - E_0(\rho, s))} \frac{\mathrm{d}\rho}{\rho} \tag{51}$$

$$I_{n,1}(w) \triangleq \frac{1}{2\pi j} \int_{w-j\infty}^{w+j\infty} e^{n(\rho R - E_0(\rho, s))} \frac{\mathrm{d}\rho}{1 - \rho}. \tag{52}$$

These integrals are estimated by expanding the exponents in (51)–(52) as a Taylor series around $\hat{\rho}$, neglecting terms of order higher than 2, and estimating the error made by this truncation. We focus on (51), as the analysis of (52) is similar. The Taylor expansion of the exponent directly gives

$$\rho R - E_0(\rho, s) = \hat{\rho} R - E_0(\hat{\rho}, s) - \frac{1}{2} E_0''(\hat{\rho}, s)(\rho - \hat{\rho})^2 + r(\rho), \tag{53}$$

where we used (39) and $r(\rho)$ is a remainder-term function.

Within a proportionality factor $a_n \triangleq e^{n(\hat{\rho} R - E_0(\hat{\rho}, s))}$, we get

$$I_{n,0}(\hat{\rho}) = a_n \frac{1}{2\pi j} \int_{\hat{\rho}-j\infty}^{\hat{\rho}+j\infty} e^{-\frac{1}{2} n E_0''(\hat{\rho}, s)(\rho - \hat{\rho})^2} e^{nr(\rho)} \frac{\mathrm{d}\rho}{\rho} \tag{54}$$

$$= a_n \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\frac{1}{2} n E_0''(\hat{\rho}, s)\rho^2} e^{nr(\hat{\rho}+j\rho)} \frac{\mathrm{d}\rho}{\hat{\rho} + j\rho} \tag{55}$$

$$\approx a_n \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\frac{1}{2} n E_0''(\hat{\rho}, s)\rho^2} \frac{\mathrm{d}\rho}{\hat{\rho} + j\rho}, \tag{56}$$

after changing the integration variable from ρ to $\rho - \hat{\rho}$ in (55) and neglecting the remainder to obtain (56).

Next, we multiply and divide the integrand by $\hat{\rho} - j\rho$, find the contribution with imaginary part to vanish since it is an odd function of ρ, and evaluate the integral in (56) as [6]

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\frac{1}{2} n E_0''(\hat{\rho}, s)\rho^2} \frac{\mathrm{d}\rho}{\hat{\rho} + j\rho} = \hat{\rho} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\frac{1}{2} n E_0''(\hat{\rho}, s)\rho^2} \frac{\mathrm{d}\rho}{\hat{\rho}^2 + \rho^2} \tag{57}$$

$$= \bar{\mathsf{Q}}\big(n\hat{\rho} E_0''(\hat{\rho}, s)\big), \tag{58}$$

where $\bar{\mathsf{Q}}(x)$ is a modified Gaussian tail function, defined as $\bar{\mathsf{Q}}(x) \triangleq \mathrm{sign}(x)\frac{1}{2}\mathrm{erfc}\left(\frac{|x|}{\sqrt{2}}\right) e^{\frac{x^2}{2}}$.

Summarizing, we have

$$I_{n,0}(\hat{\rho}) \approx \bar{\mathsf{Q}}\big(n\hat{\rho} E_0''(\hat{\rho}, s)\big) e^{n(\hat{\rho} R - E_0(\hat{\rho}, s))}. \tag{59}$$

The analysis of $I_{n,1}(\hat{\rho})$ follows similar steps and yields

$$I_{n,1}(\hat{\rho}) \approx \bar{\mathsf{Q}}\big(n(1 - \hat{\rho}) E_0''(\hat{\rho}, s)\big) e^{n(\hat{\rho} R - E_0(\hat{\rho}, s))}. \tag{60}$$

The combination of (59) and (60) with (36), (46), and (49) yields our saddlepoint approximation to the random-coding refined union bound $\mathrm{rcu_s}(n, M_n)$ for fixed rate R,

$$\mathrm{rcu_s}(n, M_n) \approx \gamma_n + \alpha_n e^{-n(E_0(\hat{\rho}, s) - \hat{\rho} R)} \tag{61}$$

$$\gamma_n = \begin{cases} 1 & \hat{\rho} < 0 \\ 0 & 0 < \hat{\rho} < 1 \\ e^{n(R - E_0(1, s))} & \hat{\rho} > 1 \end{cases} \tag{62}$$

$$\alpha_n = \left(\bar{\mathsf{Q}}\big(\hat{\rho}\sqrt{nV_{\hat{\rho}}}\big) + \bar{\mathsf{Q}}\big((1 - \hat{\rho})\sqrt{nV_{\hat{\rho}}}\big)\right). \tag{63}$$

The approximation is continuous in $\hat{\rho}$. For $\hat{\rho} = 0$, i. e. $R = I(P)$, we determine its value by considering the limit $\hat{\rho} \to 0$, as the limits from above and below coincide. Similarly, for $R = R^*(P)$, we find the value at $\hat{\rho} = 1$ from the limit $\hat{\rho} \to 1$.

An alternative, slightly simpler form of the approximation is obtained by considering the asymptotic approximation to $\bar{\mathsf{Q}}$ for large absolute values of its parameter x:

$$\bar{\mathsf{Q}}(x) = \frac{1}{2\sqrt{\pi}x} \left(1 - \frac{1}{2x^2} + \frac{3}{4x^4} - \frac{15}{8x^6} + O(x^{-8})\right). \tag{64}$$

Keeping only the dominant term in (64), we rewrite (63) as

$$\alpha_n \approx \frac{1}{2\hat{\rho}(1 - \hat{\rho})\sqrt{\pi nV_{\hat{\rho}}}}. \tag{65}$$

This form of the coefficient makes it apparent that its sign is positive for $\hat{\rho} \in (0, 1)$ and negative if $\hat{\rho} < 0$ or $\hat{\rho} > 1$.

The exponent of the approximation coincides with that in Sect. III, as it should. In contrast to the analysis in [5], which was of the form (4), the approximation in (61) contains two summands in general. One of them, if non-zero, has a non-dominant exponent that coincides with that of Arimoto's strong converse [8] or that of the sphere-packing exponent [4, Ch. 5] for values of the rate beyond the mutual information $I(P)$ or the critical rate $R^*(P)$ respectively. Further investigation of this phenomenon, as well as full treatment of the lattice information densities and a refined appraisal of the error term made in the approximation itself, are left for future work.

REFERENCES

[1] V. Strassen, "Asymptotische Abschtzungen in Shannon's Informationstheorie," in *Trans. 3rd Prague Conf. on Inf. Theory*, 1962, pp. 689–723, [English Translation: http://www.math.wustl.edu/~luthy/strassen.pdf].

[2] 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.

[3] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.

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

[5] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations," *IEEE Trans. Inf. Theory*, pp. 2647–2666, May 2014.

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

[7] G. Doetsch, *Introduction to the Theory and Applications of the Laplace Transformation*. Springer, 1974.

[8] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 19, no. 3, pp. 357–359, May 1973.