Font-Segura, J., Vazquez-Vilar, G., Martínez, A., Guillén i Fàbregas, A. y Lancho, A. (2018). Saddlepoint approximations of lower and upper bounds to the error probability in channel coding. In *2018 52nd Annual Conference on Information Sciences and Systems (CISS).*

DOI: https://doi.org/10.1109/CISS.2018.8362304

Saddlepoint Approximations of Lower and Upper Bounds to the Error Probability in Channel Coding

Josep Font-Segura[1], Gonzalo Vazquez-Vilar[2], Alfonso Martinez[1],
Albert Guillén i Fàbregas[134], Alejandro Lancho[2]
[1]Universitat Pompeu Fabra, [2]Universidad Carlos III de Madrid, [3]ICREA, [4]University of Cambridge
Email: {josep.font,gvazquez,alfonso.martinez,guillen}@ieee.org, alancho@tsc.uc3m.es

(Invited Paper)

***Abstract*—Saddlepoint approximations of the meta-converse and random-coding union bounds are derived. These bounds accurately characterize the channel coding minimum error probability for symmetric memoryless channels in a wide range of system parameters. The proposed approximations are simple to compute and yield a unified analysis of both hypothesis-testing lower bounds and random-coding upper bounds.**

I. Introduction

In [1], Polyanskiy, Poor and Verdú derived new lower and upper bounds to the error probability in channel coding. Among other remarkable results, they showed that for transmission over a length-n channel $W^n(\cdot|\cdot)$, the error probability of the best code of length n and M codewords, $P_{\mathrm{e}}^{\star}(n,M)$, satisfies [1]

$$\mathrm{mc}(n,M) \le P_{\mathrm{e}}^{\star}(n,M) \le \mathrm{rcu}(n,M), \tag{1}$$

where $\mathrm{mc}(n,M)$ and $\mathrm{rcu}(n,M)$ are respectively the meta-converse and the random-coding union (RCU) bounds.

The lower bound to $P_{\mathrm{e}}^{\star}(n,M)$ in (1) is related to the error probability of a binary hypothesis testing problem. Let $\alpha_\beta(P,Q)$ denote the smallest type-I error probability among all tests discriminating between distributions P and Q, with a type-II error probability of at most β. Then, [1, Th. 27] establishes that $\mathrm{mc}(n,M)$ is given by

$$\mathrm{mc}(n,M) \triangleq \min_{P^n}\max_{Q^n}\Big\{\alpha_{\frac{1}{M}}\big(P^n\times W^n, P^n\times Q^n\big)\Big\}, \tag{2}$$

where the minimization is over all input distributions P^n, and the maximization is over a set of auxiliary, independent of the input, output distributions Q^n. The bound (2) is usually referred to as the meta-converse bound, since several previous converse bounds can be derived from it via relaxation.

The upper bound to $P_{\mathrm{e}}^{\star}(n,M)$ in (1) follows from applying the union bound to the random-coding error probability. For a random-coding ensemble defined by the probability distribution $P^n(\boldsymbol{x})$, $\mathrm{rcu}(n,M)$ is given by [1, Th. 16]

$$\mathrm{rcu}(n,M) \triangleq \mathrm{E}\big[\min\{1,(M-1)\mathrm{pep}(\boldsymbol{X},\boldsymbol{Y})\}\big], \tag{3}$$

where $(\boldsymbol{X},\boldsymbol{Y}) \sim P^n\times W^n$, and

$$\mathrm{pep}(\boldsymbol{x},\boldsymbol{y}) \triangleq \Pr\big[W^n(\boldsymbol{y}|\overline{\boldsymbol{X}}) \ge W^n(\boldsymbol{y}|\boldsymbol{x})\big], \tag{4}$$

is the pairwise error probability, with $\overline{\boldsymbol{X}} \sim P^n$.

The RCU and the meta-converse bounds accurately characterize the minimum error probability for a wide range of lengths n and rates $R \triangleq \frac{1}{n}\log M$. Unfortunately, evaluating (2) and (3) involves integrating tail probabilities of n-dimensional random variables, which is computationally hard even for simple channels and moderate values of n.

In this work, we use the saddlepoint approximation to approximate the RCU for i.i.d. random-coding ensembles and the meta-converse bound for symmetric channels. This technique is based on approximating the cumulant generating function of the random variable of interest by its second-order Taylor expansion, and obtaining the inverse Laplace integral of the resulting expression. This technique is usually referred to as saddlepoint approximation. The derived approximations are accurate, simple to compute and yield a unified analysis of both random-coding upper bounds and hypothesis-testing lower bounds for memoryless symmetric channels.

The results derived in this paper are related to the works [2]–[7]. In particular, [2] used saddlepoint methods to obtain approximations of random-coding bounds under mismatched decoding, [3] obtained refinements of the sub-exponential factor of the random-coding bound, and [6], [7] also found expansions of the random-coding error probability based on large deviation techniques. Laplace integration methods were used in [4] to approximate the meta-converse bound for the AWGN channel, and in [5] for parallel channels, the binary-input AWGN channel and the binary symmetric channel, for both converse and achievability bounds.

The derivations of the saddlepoint approximations of the RCU and meta-converse bounds are presented here in a tutorial format, and are intended for an audience not familiar with the saddlepoint technique. We first discuss the main steps to obtain a saddlepoint approximation of a tail probability of the sum of n random variables, and then study the tail probabilities needed to compute the meta-converse and the RCU bounds for symmetric memoryless channels. Finally, we discuss and numerically evaluate the exponential gap between the RCU and meta-converse bounds.

II. Saddlepoint Approximation

Let $(X_1,\ldots,X_n)$ be a sequence of real-valued non-lattice random variables. We wish to estimate the tail probability

This work has been funded in part by the European Research Council under ERC grant agreement 725411, by the Spanish Ministry of Economy and Competitiveness under grants TEC2013-41718-R, FJCI-2014-22747, IJCI-2015-27020, and TEC2016-78434-C3 (AEI/FEDER, EU), and by the Spanish Ministry of Education, Culture and Sport under grant FPU14/01274.

$\Pr\big[Z \geq \gamma_n\big]$ of the random variable $Z = \sum_{i=1}^{n} X_i$. The cumulant generating function of Z is defined as

$$\kappa(s) = \log \mathrm{E}\big[e^{sZ}\big], \tag{5}$$

with $s \in \mathcal{S}$, where $\mathcal{S} \subseteq \mathbb{C}$ is the region of convergence of $\kappa(s)$. For independent X_i, it is immediate that $\kappa(s)$ is the sum of the cumulant generating functions of each component X_i, i.e.,

$$\kappa(s) = \sum_{i=1}^{n} \log \mathrm{E}\big[e^{sX_i}\big]. \tag{6}$$

Using the inverse Laplace transformation [8], the probability density function of Z, $p(z)$, can be recovered from its cumulant generating function $\kappa(s)$ as

$$p(z) = \frac{1}{2\pi j}\int_{\hat{s}-j\infty}^{\hat{s}+j\infty} e^{\kappa(s)-sz}\,\mathrm{d}s, \tag{7}$$

where $j = \sqrt{-1}$, and $\hat{s}$ is a real-valued parameter that allows us to shift the integration line of (7).

The Taylor expansion of $\kappa(s)$ around $\hat{s}$ is given by

$$\kappa(s) = \kappa(\hat{s}) + \kappa'(\hat{s})(s-\hat{s}) + \frac{1}{2}\kappa''(\hat{s})(s-\hat{s})^2 + \varepsilon_\kappa(s-\hat{s}), \tag{8}$$

where $\varepsilon_\kappa(\cdot)$ collects the higher-order terms in the expansion,

$$\varepsilon_\kappa(t) = \sum_{\ell=3}^{\infty} \frac{\kappa^{(\ell)}(\hat{s})}{\ell!} t^\ell \tag{9}$$

and $\kappa^{(\ell)}(s)$ denote the ℓ-th order derivative of $\kappa(s)$.

Using the expansion (8) and the change of variable $s \leftrightarrow \sigma$, $s = \hat{s} + j\sigma$, from (7) we obtain

$$p(z) = \frac{e^{\kappa(\hat{s})-\hat{s}z}}{2\pi}\int_{-\infty}^{\infty} e^{j\sigma(\kappa'(\hat{s})-z)-\frac{1}{2}\kappa''(\hat{s})\sigma^2} e^{\varepsilon_\kappa(j\sigma)}\,\mathrm{d}\sigma. \tag{10}$$

Using the Taylor expansion of the exponential function, i.e., $e^x = \sum_{m=0}^{\infty}\frac{1}{m!}x^m$, the contribution of the remaining terms $\varepsilon_\kappa(\cdot)$ of the expansion (8) in equation (10) can be expressed as $e^{\varepsilon_\kappa(j\sigma)} = 1 + \tilde{\varepsilon}_\kappa(\sigma)$, where

$$\tilde{\varepsilon}_\kappa(\sigma) \triangleq \sum_{m=1}^{\infty}\frac{1}{m!}\left(\sum_{\ell=3}^{\infty}\frac{\kappa^{(\ell)}(\hat{s})}{\ell!}(j\sigma)^\ell\right)^m. \tag{11}$$

As a result, the probability density function (10) becomes

$$p(z) = e^{\kappa(\hat{s})-\hat{s}z}\cdot\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-j\sigma z}\varphi(\sigma)\,\mathrm{d}\sigma + \varepsilon_p(z)\right), \tag{12}$$

where $\varphi(\sigma)$ is the characteristic function of a normal distribution [9] with mean $\kappa'(\hat{s})$ and variance $\kappa''(\hat{s})$, i.e.,

$$\varphi(\sigma) = e^{j\kappa'(\hat{s})-\frac{1}{2}\kappa''(\hat{s})\sigma^2}, \tag{13}$$

and $\varepsilon_p(z)$ is an approximation error term given by

$$\varepsilon_p(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\sigma(\kappa'(\hat{s})-z)-\frac{1}{2}\kappa''(\hat{s})\sigma^2}\tilde{\varepsilon}_\kappa(\sigma)\,\mathrm{d}\sigma. \tag{14}$$

Since $\varphi(\sigma)$ is integrable in $\mathbb{R}$, solving (12) leads to the saddlepoint expansion of the probability density function of Z is given by

$$p(z) = e^{\kappa(\hat{s})-\hat{s}z}\cdot\left(\frac{1}{\sqrt{2\pi\kappa''(\hat{s})}}e^{-\frac{(z-\kappa'(\hat{s}))^2}{2\kappa''(\hat{s})}} + \varepsilon_p(z)\right). \tag{15}$$

For $\hat{s} \neq 0$, the tail probability $\Pr\big[Z \geq \gamma_n\big]$ is given by

$$\Pr\big[Z \geq \gamma_n\big] = \int_{\gamma_n}^{\infty} p(z)\,\mathrm{d}z \tag{16}$$

$$= \xi(\hat{s}) + \frac{1}{2\pi j}\int_{\hat{s}-j\infty}^{\hat{s}+j\infty}\frac{e^{\kappa(s)-s\gamma_n}}{s}\,\mathrm{d}s, \tag{17}$$

where $\xi(s) = \mathbb{1}\{s < 0\}$ and $\mathbb{1}\{\cdot\}$ is the indicator function. To obtain (17), we used $p(z)$ from equation (7), interchanged the integration order, and solved the integral with respect to z. We also used the Cauchy's residue theorem [10] to accommodate the case $\hat{s} < 0$ by introducing the additional term $\xi(\hat{s})$ due to the single pole located at $s = 0$. By using (8) and solving the integral in the second term of (17), we obtain that the tail probability of Z can be expanded as

$$\Pr\big[Z \geq \gamma_n\big] = \xi(\hat{s}) + e^{\kappa(\hat{s})-\hat{s}\gamma_n}\cdot\Bigg(e^{\hat{s}(\gamma_n-\kappa'(\hat{s}))+\frac{1}{2}\hat{s}^2\kappa''(\hat{s})} \cdot \mathrm{sign}(\hat{s})\cdot\frac{1}{2}\mathrm{erfc}\left(\frac{\gamma_n-\kappa'(\hat{s})+\hat{s}\kappa''(\hat{s})}{\sqrt{2\kappa''(\hat{s})}}\right) + \varepsilon\Bigg), \tag{18}$$

where $\mathrm{erfc}(x)$ is the complementary error function, and the approximation error term ε is given by

$$\varepsilon = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{j\sigma(\kappa'(\hat{s})-\gamma_n)-\frac{1}{2}\kappa''(\hat{s})\sigma^2}}{\hat{s}+j\sigma}\tilde{\varepsilon}_\kappa(\sigma)\,\mathrm{d}\sigma. \tag{19}$$

The characterization of the approximation errors of the expansions (15) and (18) involves the analysis of the terms in $\varepsilon_a(\sigma)$, and the integrations $\varepsilon_p(z)$ and ε.

For the choice of $\hat{s}$ satisfying $\kappa'(\hat{s}) = \gamma_n$, (18) becomes

$$\Pr\big[Z \geq \gamma_n\big] = \xi(\hat{s}) + e^{\kappa(\hat{s})-\hat{s}\kappa'(\hat{s})}\cdot\left(\Psi\left(\hat{s}\sqrt{\kappa''(\hat{s})}\right) + \varepsilon\right), \tag{20}$$

where $\Psi(x)$ is defined as

$$\Psi(x) \triangleq \frac{1}{2}\,\mathrm{erfc}\left(\frac{|x|}{\sqrt{2}}\right)e^{\frac{x^2}{2}}\,\mathrm{sign}(x), \tag{21}$$

where $\mathrm{erfc}(x)$ denotes the complementary error function and $\mathrm{sign}(x)$ is equal to 1 for $x \geq 0$ and -1 for $x < 0$.

When $(X_1, \ldots, X_n)$ is a sequence of i.i.d. random variables satisfying certain regularity conditions, this choice of $\hat{s}$ yields an approximation error ε with order $\mathcal{O}(n^{-1})$ [11, Ch. 2]. For the sake of clarity, and given the tutorial nature of this paper, in the next sections we shall only consider the approximations that follow from (15), (18) and (20) by neglecting the corresponding approximation error terms $\varepsilon_p(z)$ and ε.

III. Meta-Converse Bound

In this section, we consider the family of memoryless symmetric channels with continuous output alphabet, i.e., $W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} W(y_i|x_i)$, where $\Pr\big[W(Y|x) \geq \zeta\big]$, for $Y \sim W(\cdot|x)$, is independent of x. For certain i.i.d. auxiliary distributions $Q^n(\boldsymbol{y}) = \prod_{i=1}^{n} Q(y_i)$, the term in braces in (2) also becomes independent of P^n. More precisely, we consider an i.i.d. auxiliary distribution Q such that the tail probability $\Pr\big[W(Y|x) \geq \zeta Q(Y)\big]$ is independent of $x \in \mathcal{X}$. For symmetric memoryless channels, this condition is satisfied, e.g., when $Q = Q_{\text{ca}}$ is the capacity-achieving output distribution, or when $Q = Q_\rho$ is an exponent-achieving output distribution (see (38) for the precise definition of Q_ρ). For simplicity, we shall also assume that $W(\cdot|x)$ and Q share the same support.

For any choice of Q satisfying this symmetry condition, the meta-converse bound (2) is given by

$$\mathrm{mc}(n, M, Q) = \alpha_{\frac{1}{M}}\big(W^n(\cdot|\boldsymbol{x}), Q^n\big), \quad \forall \boldsymbol{x} \in \mathcal{X}^n. \tag{22}$$

We use the following characterization of the trade-off $\alpha_\beta(\cdot)$.

Lemma 1: The optimal error probability trade-off for testing between P and Q defined on $\mathcal{Y}$, is given by

$$\alpha_\beta\big(P, Q\big) = \max_{\eta \ge 0}\left\{\Pr\left[\frac{P(Y)}{Q(Y)} \le \eta\right] + \eta\left(\Pr\left[\frac{P(\overline{Y})}{Q(\overline{Y})} > \eta\right] - \beta\right)\right\} \tag{23}$$

where $Y \sim P$ and $\overline{Y} \sim Q$.

Defining the information density $\jmath(\boldsymbol{y})$ for any $\boldsymbol{x} \in \mathcal{X}^n$ as

$$\jmath(\boldsymbol{y}) \triangleq \log \frac{W^n(\boldsymbol{y}|\boldsymbol{x})}{Q^n(\boldsymbol{y})}, \tag{24}$$

and applying Lemma 1 with $\eta = e^{n\gamma}$, from (22) we obtain

$$\mathrm{mc}(n, M, Q) = \max_\gamma\left\{\Pr\big[\jmath(\boldsymbol{Y}) \le n\gamma\big] + e^{n\gamma}\left(\Pr\big[\jmath(\overline{\boldsymbol{Y}}) > n\gamma\big] - \frac{1}{M}\right)\right\}, \tag{25}$$

where $\boldsymbol{Y} \sim W^n(\cdot|\boldsymbol{x})$, and $\overline{\boldsymbol{Y}} \sim Q^n$.

A. Saddlepoint approximation

Let us define the random variables $Z_0 \triangleq \jmath(\boldsymbol{Y})$, where $\boldsymbol{Y} \sim W^n(\cdot|\boldsymbol{x})$, and $Z_1 \triangleq \jmath(\overline{\boldsymbol{Y}})$, where $\overline{\boldsymbol{Y}} \sim Q^n$, and let $\kappa_0(s)$ and $\kappa_1(s)$ be their respective cumulant generating functions, i.e.,

$$\kappa_0(s) = \log \mathrm{E}\big[e^{sZ_0}\big], \tag{26}$$

$$\kappa_1(s) = \log \mathrm{E}\big[e^{sZ_1}\big]. \tag{27}$$

Applying the expansion (20) to the (lower and upper) tail probabilities in (25), neglecting the remainder term ε, yields

$$\Pr\big[\jmath(\boldsymbol{Y}) \le \gamma\big] \simeq \xi(-s_0) + \Psi\Big(-s_0\sqrt{\kappa_0''(s_0)}\Big) \cdot e^{\kappa_0(s_0) - s_0\kappa_0'(s_0)}, \tag{28}$$

$$\Pr\big[\jmath(\overline{\boldsymbol{Y}}) > \gamma\big] \simeq \xi(s_1) + \Psi\Big(s_1\sqrt{\kappa_1''(s_1)}\Big) \cdot e^{\kappa_1(s_1) - s_1\kappa_1'(s_1)}, \tag{29}$$

where s_0 and s_1 satisfy $\kappa_0'(s_0) = \kappa_1'(s_1) = n\gamma$, and where we recall that $\xi(s) = \mathbb{1}\{s < 0\}$ and that $\Psi(\cdot)$ is defined in (21).

Both Z_0 and Z_1 are sums of i.i.d. random variables, and their cumulant generating functions are a shifted version of each other. Indeed, $\kappa_0(s)$ and $\kappa_1(s)$ are $\kappa_0(s) = n\kappa(s-1)$ and $\kappa_1(s) = n\kappa(s)$, where $\kappa(s)$ is

$$\kappa(s) = \log \int \frac{W(y|x)^s}{Q(y)^{s-1}}\,\mathrm{d}y. \tag{30}$$

Then, using the approximations (28) and (29) in equation (25), together with the relations $\kappa_0(s) = n\kappa(s-1)$ and $\kappa_1(s) = n\kappa(s)$, we obtain the following approximation of the meta-converse expression, $\mathrm{mc}(n, M, Q) \simeq \hat{\mathrm{mc}}(n, M, Q)$.

Approximation 1 (Meta-converse with fixed Q):

$$\hat{\mathrm{mc}}(n, M, Q) \triangleq \max_s\left\{c_n(s)e^{n(\kappa(s) + (1-s)\kappa'(s))} + \Big(\xi(s) - \tfrac{1}{M}\Big)e^{n\kappa'(s)}\right\} \tag{31}$$

where

$$c_n(s) \triangleq \xi(1-s) + \Psi\Big((1-s)\sqrt{n\kappa''(s)}\Big) + \Psi\Big(s\sqrt{n\kappa''(s)}\Big). \tag{32}$$

Obtaining a closed-form expression for the cumulant generating function $\kappa(s)$ and its first and second derivatives is difficult in general. By defining the functions $J_\ell(s)$, for $\ell = 0, 1, 2$, as

$$J_\ell(s) \triangleq \log \int \frac{W(y|x)^s}{Q(y)^{s-1}}\left(\log \frac{W(y|x)}{Q(y)}\right)^\ell \mathrm{d}y, \tag{33}$$

inspecting $\kappa(s)$ in (30), it follows that $\kappa(s) = \log J_0(s)$ and

$$\kappa'(s) = \frac{J_1(s)}{J_0(s)}, \tag{34}$$

$$\kappa''(s) = \frac{J_0(s)J_2(s) - J_1(s)^2}{J_0(s)^2}. \tag{35}$$

Hence, computing $\kappa(s)$, $\kappa'(s)$ and $\kappa''(s)$ involves solving one-dimensional integrals, in contrast to the n-dimensional integrals appearing in (25). The integrals in (33) can be numerically approximated with arbitrary precision.

B. Sphere-packing exponent

We now relate the meta-converse approximation (31) with the sphere-packing exponent. For a given input distribution P, the Gallager's E_0-function [12, Eq. (5.6.14)] is defined as

$$E_0(\rho, P) \triangleq -\log \int \left(\int P(x)W(y|x)^{\frac{1}{1+\rho}}\,\mathrm{d}x\right)^{1+\rho} \mathrm{d}y, \tag{36}$$

and we further define $E_0(\rho) \triangleq \max_P E_0(\rho, P)$.

The sphere-packing exponent, which is defined as

$$E_{\text{sp}}(R) \triangleq \sup_{\rho \ge 0}\big\{E_0(\rho) - \rho R\big\}, \tag{37}$$

is an upper bound to the reliability function in channel coding.

We consider a tilted output distribution that plays a crucial role in the derivation of the sphere-packing exponent $E_{\text{sp}}(R)$, as discussed in [13], [14]. We define $Q_\rho(y)$ as

$$Q_\rho(y) \triangleq \frac{1}{\mu(\rho)}\left(\int P(x)W(y|x)^{\frac{1}{1+\rho}}\,\mathrm{d}x\right)^{1+\rho}, \tag{38}$$

for some input distribution P, valid for $\rho \ge 0$, and where $\mu(\rho)$ is a normalizing factor. For the particular choice of $Q = Q_\rho$ with P uniform over the input alphabet, the saddlepoint approximation (31) maximized over $\rho \ge 0$ yields

$$\sup_{\rho \ge 0}\big\{\hat{\mathrm{mc}}(n, M, Q_\rho)\big\} = \sup_{\rho \ge 0, s}\left\{c_n(s)e^{n(\kappa(s) + (1-s)\kappa'(s))} + \Big(\xi(s) - \tfrac{1}{M}\Big)e^{n\kappa'(s)}\right\}. \tag{39}$$

In (39), we may let s be a function of ρ and still obtain a lower bound to $P_{\text{e}}^\star(n, M)$. Setting $s = \frac{1}{1+\rho}$, or equivalently

$\rho = \frac{1-s}{s}$ yields, after tedious but straightforward algebra, that $\kappa(s)$ and $\kappa'(s)$ are related to the E_0-function as

$$\kappa(s) = -sE_0\Big(\tfrac{1-s}{s}\Big), \tag{40}$$

$$\kappa'(s) = \frac{1}{s}E_0'\big(\tfrac{1-s}{s}\big) - E_0\big(\tfrac{1-s}{s}\big), \tag{41}$$

where $E_0'(\rho)$ denotes the derivative of $E_0(\rho)$ with respect to ρ. While (41) coincides with the derivative of the right-hand side of (40), this is not to be expected in general. Indeed, the derivative $\kappa'(s)$ is obtained assuming Q fixed and then substituting $Q = Q_\rho$ in the resulting expression, while this assignment is already implicit in the right-hand side of (40).

Substituting (40) and (41) in (39) with the change of variable $s \in [0,1] \leftrightarrow \rho = \frac{1-s}{s} \in [0,\infty)$, using that $M = e^{nR}$, we obtain the following approximation, $P_{\rm e}^\star(n,M) \gtrsim \hat{\rm mc}(n,M)$.

Approximation 2 (Sphere-packing bound):

$$\hat{\rm mc}(n,M) \triangleq \max_{\rho\geq 0}\bigg\{e^{-n(E_0(\rho)-\rho E_0'(\rho))}\cdot \cdot\bigg(\Psi\Big(\sqrt{nU(\rho)}\Big) + \Psi\Big(\rho\sqrt{nU(\rho)}\Big) - e^{-n(R-E_0'(\rho))}\bigg)\bigg\} \tag{42}$$

where $\Psi(\cdot)$ is defined in (21), and the variance term $U(\rho)$ is

$$U(\rho) \triangleq \frac{\kappa''\big(\frac{1}{1+\rho}\big)}{(1+\rho)^2}, \tag{43}$$

for $\kappa''(\cdot)$ given in (35)-(33) with auxiliary $Q = Q_\rho$.

We observe that (42) recovers the sphere-packing error exponent for the case of symmetric memoryless channels. More precisely, let us fix $\rho = \hat\rho$, where $\hat\rho$ is the unique solution to $E_0'(\hat\rho) = R - \delta$, for some $\delta > 0$. Then, (42) yields

$$P_{\rm e}^\star(n,M) \gtrsim \max_{\delta>0} e^{-n(E_0(\hat\rho)-\hat\rho(R-\delta))}\cdot \cdot\bigg(\Psi\Big(\sqrt{nU(\hat\rho)}\Big) + \Psi\Big(\hat\rho\sqrt{nU(\hat\rho)}\Big) - e^{-n\delta}\bigg). \tag{44}$$

Since $e^{-n\delta}$ vanishes for any $\delta > 0$ as $n \to \infty$, and since the functions $\Psi(\cdot)$ behave as $\mathcal{O}\big(n^{-\frac{1}{2}}\big)$, it follows that equation (44) recovers the sphere-packing exponent $E_{\rm sp}(R)$ by first letting $n \to \infty$, and then $\delta \to 0$.

IV. Random Coding Union Bound

In this section, we first derive an approximation to the pairwise error probability (4), and then use this approximation to find the saddlepoint approximation of the RCU bound (3).

A. Pairwise error probability

Let $Z(\boldsymbol{y}) = \log W^n(\boldsymbol{y}|\overline{\boldsymbol{X}})$. For memoryless channels $W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^n W(y_i|x_i)$, $Z(\boldsymbol{y})$ is the sum of n independent random variables with cumulant generating function

$$\omega(\tau) = \sum_{i=1}^n \log \mathrm{E}_P\big[W(y_i|\overline{X})^\tau\big]. \tag{45}$$

By performing a Taylor expansion of $\omega(\tau)$ around $\hat\tau$, disregarding the $\varepsilon_p(z)$ term in (15), the probability density function of $Z(\boldsymbol{y})$ can be approximated as

$$p(z) \simeq e^{\omega(\hat\tau)-\hat\tau z}\cdot\frac{1}{\sqrt{2\pi\omega''(\hat\tau)}}e^{-\frac{(z-\omega'(\hat\tau))^2}{2\omega''(\hat\tau)}}, \tag{46}$$

where $\omega'(\tau)$ and $\omega''(\tau)$ are the first and the second derivative of $\omega(\tau)$, respectively. Integrating (46) over $[\gamma_n(\boldsymbol{x},\boldsymbol{y}),\infty)$, where $\gamma_n(\boldsymbol{x},\boldsymbol{y}) = \log W^n(\boldsymbol{y}|\boldsymbol{x})$, we obtain that the pairwise error probability $\mathrm{pep}(\boldsymbol{x},\boldsymbol{y}) = \Pr\big[Z(\boldsymbol{y}) \geq \gamma_n(\boldsymbol{x},\boldsymbol{y})\big]$ can be approximated by

$$\mathrm{pep}(\boldsymbol{x},\boldsymbol{y}) \simeq \lambda_{\hat\tau}(\boldsymbol{x},\boldsymbol{y})e^{-i_{\hat\tau}(\boldsymbol{x};\boldsymbol{y})}, \tag{47}$$

where $i_{\hat\tau}(\boldsymbol{x};\boldsymbol{y})$ is a tilted information density given by

$$i_{\hat\tau}(\boldsymbol{x};\boldsymbol{y}) = \log\frac{W^n(\boldsymbol{y}|\boldsymbol{x})^{\hat\tau}}{\mathrm{E}\big[W^n(\boldsymbol{y}|\overline{\boldsymbol{X}})^{\hat\tau}\big]}, \tag{48}$$

and $\lambda_{\hat\tau}(\boldsymbol{x},\boldsymbol{y})$ is the pre-exponential factor

$$\lambda_{\hat\tau}(\boldsymbol{x},\boldsymbol{y}) = e^{\hat\tau(\log W^n(\boldsymbol{y}|\boldsymbol{x})-\omega'(\hat\tau))+\frac{1}{2}\hat\tau^2\omega''(\hat\tau)}\cdot \cdot\frac{1}{2}\mathrm{erfc}\left(\frac{\log W^n(\boldsymbol{y}|\boldsymbol{x})-\omega'(\hat\tau)+\hat\tau\omega''(\hat\tau)}{\sqrt{2\omega''(\hat\tau)}}\right). \tag{49}$$

We remark that the pairwise error probability (4) is a tail probability evaluated at a point $\gamma_n(\boldsymbol{x},\boldsymbol{y}) = \log W^n(\boldsymbol{y}|\boldsymbol{x})$ that depends on $\boldsymbol{x}$ and $\boldsymbol{y}$. Hence, the optimal parameter $\hat\tau$ would be chosen as the unique solution to $\omega'(\hat\tau) = \log W^n(\boldsymbol{y}|\boldsymbol{x})$. However, given the outer expectation over $\boldsymbol{X},\boldsymbol{Y}$ in (3), this would require one optimization for every $\boldsymbol{x}$ and $\boldsymbol{y}$. Instead, we let $\hat\tau > 0$ be fixed for every $\boldsymbol{x}$ and $\boldsymbol{y}$, at the cost of having $\log W^n(\boldsymbol{y}|\boldsymbol{x}) - \omega'(\hat\tau) \neq 0$ in the expression (49). As we will see, the effect of this will be negligible.

B. Random coding union bound

Using that $\mathrm{E}[\min\{1,Z\}] = \Pr[Z \geq U]$, where U is uniformly distributed in the $[0,1]$ interval, and defining the random variable $Z = \log(M-1) + \log\mathrm{pep}(\boldsymbol{X},\boldsymbol{Y}) - \log U$, we may write the RCU bound (3) as the tail probability $\mathrm{rcu}(n,M) = \Pr[Z \geq 0]$. Plugging the saddlepoint approximation of the pairwise error probability given in (48) and taking $\log(M-1) \simeq nR$, the cumulant generating function of Z is asymptotically given, according to (5), by

$$\chi(\rho) \simeq n\rho R + \log\mathrm{E}\big[\lambda_{\hat\tau}(\boldsymbol{X},\boldsymbol{Y})^\rho e^{-\rho i_{\hat\tau}(\boldsymbol{X};\boldsymbol{Y})}\big] - \log(1-\rho). \tag{50}$$

For a memoryless channel $W^n(\boldsymbol{y}|\boldsymbol{x})$ and i.i.d. input distribution $P^n(\boldsymbol{x})$, it is convenient to set $\hat\tau = \frac{1}{1+\hat\rho}$ and define the following tilted distribution

$$P_\rho^n(\boldsymbol{x})W_\rho^n(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{\nu_n}P^n(\boldsymbol{x})W^n(\boldsymbol{y}|\boldsymbol{x})e^{-\rho i_{\hat\tau}(\boldsymbol{x};\boldsymbol{y})}, \tag{51}$$

Hence, after some mathematical manipulations, we may express equation (50) as

$$\chi(\rho) \simeq n\rho R - nE_0(\rho,P) + \log\theta_n(\rho) - \log(1-\rho), \tag{52}$$

where $E_0(\rho,P)$ corresponds to Gallager's E_0 function defined in (36) and the term $\theta_n(\rho)$ is defined as

$$\theta_n(\rho) = \mathrm{E}_{P_\rho^n W_\rho^n}\Big[\lambda_{\hat\tau}(\boldsymbol{X},\boldsymbol{Y})^\rho\Big]. \tag{53}$$

Since $\mathrm{rcu}(n,M) = \Pr\big[Z \geq 0\big]$ with cumulant generating function (52), we use (17) to write the RCU bound as the following complex integration

$$\mathrm{rcu}(n,M) \simeq \frac{1}{2\pi j}\int_{\hat\rho-j\infty}^{\hat\rho+j\infty}\frac{e^{-n(E_0(\rho,P)-\rho R)}}{\rho(1-\rho)}\theta_n(\rho)\,\mathrm{d}\rho. \tag{54}$$

The saddlepoint approximation of the RCU bound involves expanding $\rho R - E_0(\rho, P)$ around $\rho = \hat{\rho}$, the unique solution of the equation

$$E_0'(\hat{\rho}, P) = R. \tag{55}$$

Hence,

$$\rho R - E_0(\rho, P) \simeq \hat{\rho} R - E_0(\hat{\rho}, P) + \frac{1}{2} V(\hat{\rho})(\rho - \hat{\rho})^2, \tag{56}$$

where $V(\hat{\rho})$ is the channel dispersion, i.e.,

$$V(\hat{\rho}) = -E_0''(\hat{\rho}, P). \tag{57}$$

The convergence of the complex integration (54) depends on the poles at $\rho = 0$ and $\rho = 1$. As discussed in [12, p. 142], the error exponent of the random-coding i.i.d. ensemble with distribution $P^n(\boldsymbol{x})$ is given by

$$E(R, P) = \min_{0 \le \rho \le 1} \rho R - E_0(\rho, P). \tag{58}$$

Clearly, a solution to (55) minimizes (58) only when $0 \le \hat{\rho} \le 1$, i.e., for rates R between the critical rate $R^*(P)$, defined as the rate for which $E(R, P)$ is achieved at $\hat{\rho} = 1$, and the mutual information $I(P)$, for which $\hat{\rho} = 0$. For this range of $\hat{\rho}$, the complex integration (54) converges, otherwise we need to shift the integration axis at a cost of introducing additional terms. Following the footsteps in [15, Sec. IV] based on the Cauchy's residue theorem [8], we obtain that

$$\mathrm{rcu}(n, M) \simeq \tilde{\xi}_n(\hat{\rho}) + \frac{1}{2\pi j} \int_{\hat{\rho} - j\infty}^{\hat{\rho} + j\infty} \frac{e^{-n(E_0(\rho, P) - \rho R)}}{\rho(1 - \rho)} \theta_n(\rho) \, \mathrm{d}\rho, \tag{59}$$

where $\tilde{\xi}_n(\hat{\rho})$ accounts for the contribution of the poles as

$$\tilde{\xi}_n(\hat{\rho}) = \begin{cases} 1 & \hat{\rho} < 0 \\ 0 & 0 \le \hat{\rho} \le 1 \\ e^{-n(E_0(1, P) - R)} \theta_n(1) & \hat{\rho} > 1. \end{cases} \tag{60}$$

Finally using the Taylor expansion (52), further approximating $\theta_n(\rho) \approx \theta_n(\hat{\rho})$, solving the complex integration (59) yields the saddlepoint approximation of $\mathrm{rcu}(n, M)$.

Approximation 3 (RCU bound):

$$\mathrm{rcu}(n, M) \simeq \tilde{\xi}_n(\hat{\rho}) + \psi_n(\hat{\rho}) e^{-n(E_0(\hat{\rho}, P) - \hat{\rho} R)}, \tag{61}$$

where the pre-exponential factor $\psi_n(\hat{\rho})$ is given by

$$\psi_n(\hat{\rho}) = \theta_n(\hat{\rho}) \cdot \Big(\Psi\big(\hat{\rho}\sqrt{nV(\hat{\rho})}\big) + \Psi\big((1 - \hat{\rho})\sqrt{nV(\hat{\rho})}\big) \Big). \tag{62}$$

The term $\theta_n(\hat{\rho})$ can be numerically computed from equations (53), (51) and (49). A simpler expression is obtained by further expanding the pre-exponential factor of the pairwise error probability $\lambda_{\hat{\tau}}(\boldsymbol{x}, \boldsymbol{y})$ in the right hand side of (49), and solving the expectation (53) with respect to the tilted distribution (51). By choosing $\hat{\tau} = \frac{1}{1+\hat{\rho}}$, we obtain

$$\theta_n(\hat{\rho}) \simeq \frac{1}{\sqrt{1 + \hat{\rho}}} \left(\frac{1 + \hat{\rho}}{\sqrt{2\pi n \overline{\omega}''(\hat{\rho})}} \right)^{\hat{\rho}}, \tag{63}$$

as $n \to \infty$, where $\overline{\omega}''(\hat{\rho})$ is an averaged variance given by

$$\overline{\omega}''(\hat{\rho}) = \int Q_{\hat{\rho}}(y) \left[\frac{\partial^2}{\partial \tau^2} \left(\log \int P(x) W(y|x)^{\tau} \, \mathrm{d}x \right) \bigg|_{\tau = \hat{\tau}} \right] \mathrm{d}y, \tag{64}$$

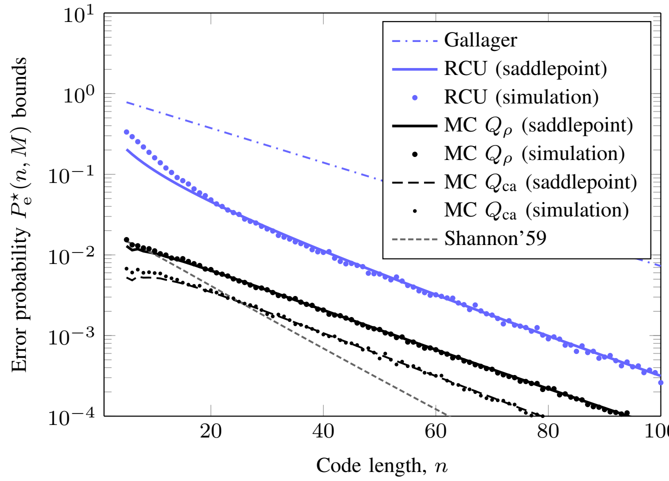

Fig. 1. Channel coding error probability bounds vs block-length n with parameters $R = 0.65$ bits per channel use and $\mathrm{snr} = 4.77$ dB.

and $Q_\rho(y)$ is the auxiliary distribution defined in (38).

We finally remark that since $\theta_n(\hat{\rho})$ in (63) behaves as $\mathcal{O}(n^{-\frac{\hat{\rho}}{2}})$, and that the functions $\Psi(\cdot)$ behave as $\mathcal{O}\big(n^{-\frac{1}{2}}\big)$, the RCU approximation (61) recovers the error exponent for i.i.d. random-coding ensembles.

V. Discussion

In the previous sections, we have seen that the saddlepoint approximation to the meta-converse bound (42) recovers the sphere-packing exponent (37), whereas the RCU approximation (61) does it for the error exponent (58) of the i.i.d. random-coding ensemble $P^n(x)$. We next study their pre-exponential factors in the region where the exponents coincide. To this end, we expand $\Psi(x)$ function defined in (21) as

$$\Psi(x) = \frac{1}{\sqrt{2\pi x}} \left(1 - \frac{1}{x^2} + \frac{3}{x^4} - \frac{15}{x^6} + \ldots \right). \tag{65}$$

For $x \to \infty$, we consider only the first term in the expansion. Then, for sufficiently large n, the meta-converse saddlepoint approximation (44) yields

$$P_e^\star(n, e^{nR}) \gtrsim \max_{\delta > 0} e^{-n(E_0(\hat{\rho}(\delta)) - \hat{\rho}(\delta)(R - \delta))} \cdot \cdot \left(\frac{1 + \hat{\rho}(\delta)}{\hat{\rho}(\delta)\sqrt{2\pi n U(\hat{\rho}(\delta))}} - e^{-n\delta} \right), \tag{66}$$

where $\hat{\rho}(\delta)$ is the solution to $E_0'(\rho) = R - \delta$, and the RCU saddlepoint approximation (61), yields

$$P_e^\star(n, e^{nR}) \lesssim \tilde{\xi}_n(\hat{\rho}) + \frac{\theta_n(\hat{\rho})}{\hat{\rho}(1 - \hat{\rho})\sqrt{2\pi n V(\hat{\rho})}} e^{-n(E_0(\hat{\rho}) - \hat{\rho} R)}, \tag{67}$$

where $\hat{\rho}$ satisfies $E_0'(\hat{\rho}) = R$. For rates between the critical rate and the mutual information, $\hat{\rho} \in (0, 1)$, $\tilde{\xi}_n(\hat{\rho}) = 0$, and the error exponents of the meta-converse bound and of the RCU bound asymptotically coincide as we let first $n \to \infty$ and then $\delta \to 0$. In contrast, for rates below the critical rate, since $\hat{\rho} > 1$, the error exponent of the RCU bound is dominated by the term $\tilde{\xi}_n(\hat{\rho})$ in (60), i.e., $E_0(1) - R$, whereas the sphere packing exponent remains $E_0(\hat{\rho}) - \hat{\rho} R$.

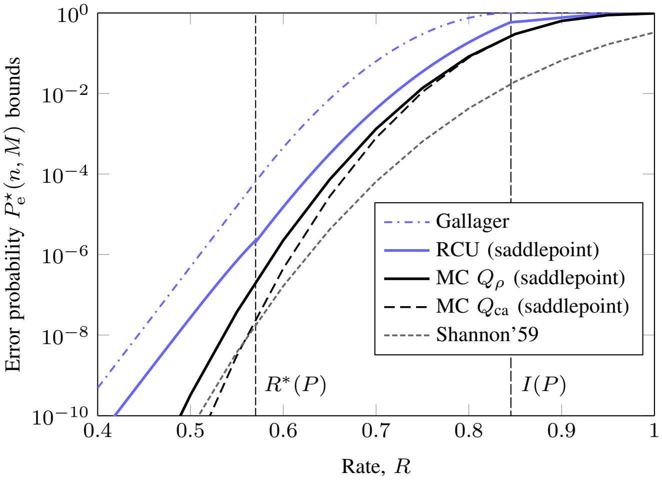

Fig. 2. Channel coding error probability bounds vs rate R in bits per channel use, with parameters $n = 100$ and $\mathrm{snr} = 4.77$ dB.

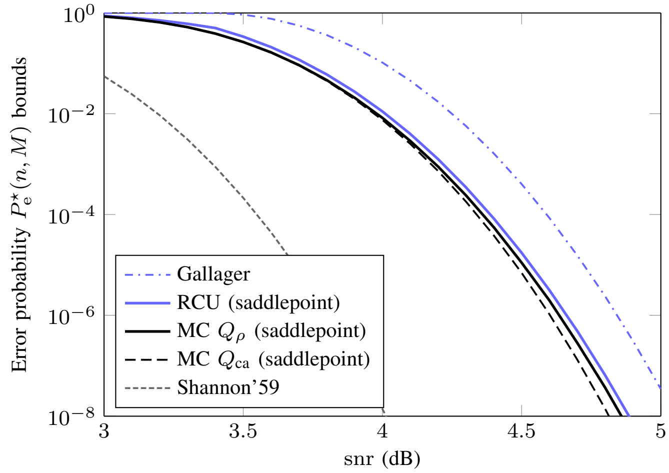

Fig. 3. Channel coding error probability bounds vs snr, with parameters $R = 0.75$ bits per channel use and $n = 1024$.

We numerically evaluate the derived approximations for the binary-input AWGN channel, with real-valued noise variance σ^2. For a signal energy E_s, codewords $\boldsymbol{x} \in \{-\sqrt{E_s}, +\sqrt{E_s}\}^n$ are distributed as $P^n(\boldsymbol{x}) = 2^{-n}$ for the random-coding upper bound. The signal-to-noise ratio is defined as $\mathrm{snr} = \frac{E_s}{\sigma^2}$, and $R = \frac{1}{n}\log_2(M)$ is the rate in bits per channel use. We consider the saddlepoint approximation of the RCU bound (61) with $\theta_n(\hat{\rho})$ given in (63), and that of the meta-converse (MC) (31) with the capacity-achieving output distribution $Q = Q_{ca}$, and optimized over exponent achieving auxiliary distributions $Q = Q_\rho$. As a reference, we also include the Gallager bound $P_e^\star(n, M) \leq e^{-nE(R,P)}$ and Shannon lower bound for the AWGN channel [16, Eq. (15)].

Fig. 1 depicts the approximations compared to a simulation of the actual RCU bound (obtained from (3) and (4)) and the corresponding MC bounds (25) via Monte Carlo integration methods. We can see that both the saddlepoint approximation of the RCU and that of the MC bounds are accurate for block-lengths as short as $n = 20$. In this scenario, the rate considered is above the critical rate of the channel, $R^*(P) = 0.5702$. Then, the error exponents of the RCU bound and MC bound with Q_ρ coincide. The MC with capacity-achieving auxiliary distribution Q_{ca} yields a weaker bound than that with Q_ρ. Furthermore, as it does not attain the sphere-packing exponent, the gap between the bounds exponentially increases with n.

Fig. 2 shows the error probability bounds versus the rate R. For reference, the critical rate and mutual information are also included. For $\mathrm{snr} = 4.77$ dB, these are $R^*(P) = 0.5702$ and $I(P) = 0.8453$ bits per channel use, respectively. We can see the the rate gap between the upper and lower bounds is approximately constant in the range R^* to I. However, below the critical rate, the bounds start to diverge from each other, as it could be expected from the error exponent analysis. Finally, Fig. 3 shows that, for $n = 1024$ and $R = 0.75$ bits per channel use, the RCU and MC bounds accurately characterize $P_e^\star(n, M)$ for the whole range of snr, as the relative gap is surprisingly small.

References

[1] 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.

[2] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2647–2666, May 2014.

[3] Y. Altuğ and A. B. Wagner, "Refinement of the random coding bound," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6005–6023, 2014.

[4] T. Erseghe, "On the evaluation of the Polyanskiy-Poor-Verdú converse bound for finite block-length coding in AWGN," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6578—-6590, Dec. 2015.

[5] ——, "Coding in the finite-blocklength regime: Bounds based on Laplace integrals and their asymptotic approximations," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6854—6883, Dec. 2016.

[6] J. Honda, "Exact asymptotics for the random coding error probability," in *IEEE Int. Symp. on Inf. Theory (ISIT)*, 2015, pp. 91–95.

[7] ——, "Comprehensive analysis on exact asymptotics of random coding error probability," *arXiv preprint 1707.04401*, 2017.

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

[9] W. Feller, *An Introduction to Probability Theory and its Applications*. New York, NY: Wiley, 1971, vol. 2.

[10] F. W. J. Olver, *Asymptotics and Special Functions*. New York, NY: Academic Press, 1971.

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

[12] R. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.

[13] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Information and Control*, vol. 10, no. 1, pp. 65–103, 1967.

[14] R. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 4, pp. 405–417, Jul. 1974.

[15] A. Martinez, J. Scarlett, M. Dalai, and A. Guillén i Fàbregas, "A complex-integration approach to the saddlepoint approximation for random–coding bounds," in *Int. Symp. Wireless Commun. Systems (ISWCS)*, 2014, pp. 618–621.

[16] C. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. Journal*, vol. 38, no. 3, pp. 611–656, May 1959.