Asymptotics of the Random Coding Union Bound

(Invited Paper)

Josep Font-Segura
Universitat Pompeu Fabra
josep.font@ieee.org

Alfonso Martinez
Universitat Pompeu Fabra
alfonso.martinez@ieee.org

Albert Guillén i Fàbregas
ICREA and Universitat Pompeu Fabra
University of Cambridge
guillen@ieee.org

***Abstract*—Saddlepoint approximations and expansions of the random coding union bound are derived for the i.i.d. random coding ensemble. Using the inverse Laplace transform of lattice and strongly non-lattice distributions, our results recover the random coding error exponent and refine the pre-exponential coefficient of the error probability. Explicit characterization of the terms are given for the binary symmetric channel and for the binary input AWGN channel.**

I. Introduction

Random coding [1], [2] is a central tool in information theory to prove the existence of codes with vanishing error probability P_{e} as the code length n tends to infinity. For a fixed coding rate R and i.i.d. random coding ensembles with input distribution $Q(x)$, the average error probability over the ensemble of codes is given by

$$P_{\mathrm{e}} = \alpha_n \cdot e^{-nE(R)} \tag{1}$$

with positive error exponent $E(R) > 0$ as long as the rate is below the mutual information $R < I(Q)$ [3].

The polynomial decay of α_n was first studied in [4] for strongly symmetric channels, and shown to be related to the slope of $E(R)$. Recently, more accurate characterizations of the term α_n were derived by [5]–[7]. In [5], large deviations techniques are used to upper bound α_n for discrete memoryless channels, whereas asymptotically tight expressions of α_n are derived in [7] for the exact random coding error probability using the saddlepoint method. Large deviations, saddlepoint and Laplace methods were also used to study coding rates for a fixed error probability in [8]–[10].

We consider the random coding union (RCU) bound to the error probability P_{e}, and note that it can be written as two nested tail probabilities. Inspired by our previous works [11]–[16], we use the saddlepoint method to approximate the RCU for i.i.d. random coding for both lattice and strongly non-lattice distributions. Our analysis differs from that of [5]–[7] in the way we optimize the parameters involved in the saddlepoint approximation.

The derived approximations and expansions of the RCU recover the random coding error exponent $E(R)$, refine the computation of α_n for finite length n based on the cumulant generating function of the two tail probabilities, and asymptotically retrieves the polynomial decay of α_n found in [5]–[7].

This work has been funded in part by the European Research Council under grant 725411, and by the Spanish Ministry of Economy and Competitiveness under grants FJCI-2014-22747 and TEC2016-78434-C3-1-R.

II. Random Coding Union Bound

We consider the i.i.d. random coding ensemble where M codewords of length n are independently generated according to the probability distribution $Q^n(\boldsymbol{x}) = \prod_{i=1}^n Q(x_i)$, and transmitted over a channel with transition probability $W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^n W(y_i|x_i)$. Let P_{e} be the average of the maximum likelihood decoding error probability over the ensemble of codes. Then, there exists a code of parameters (M, n) whose error probability is at most P_{e} [17, Th. 15]. Considering decoding ties as errors, and applying the union bound of the error events, P_{e} is weakened to the random coding union (RCU) [17, Th. 16], given by

$$\begin{aligned}\mathrm{rcu} = \mathrm{E}\big[\min\big\{1, (M-1)\\ \Pr\big[W^n(\boldsymbol{Y}|\overline{\boldsymbol{X}}) \geq W^n(\boldsymbol{Y}|\boldsymbol{X})|\boldsymbol{X},\boldsymbol{Y}\big]\big\}\big],\end{aligned} \tag{2}$$

where $(X, Y, \overline{X}) \sim Q \times W \times Q$. Further using the identity $\mathrm{E}[\min\{1, A\}] = \Pr[A \geq U]$, where U is uniform in the $[0, 1]$ interval and taking logarithms, the RCU (2) can be written as two nested tail probabilities, i.e.,

$$\mathrm{rcu} = \Pr\big[\Theta \geq \log U\big], \tag{3}$$

where Θ is the random variable

$$\Theta = \log(M-1) + \log \mathrm{pep}(\boldsymbol{X}, \boldsymbol{Y}), \tag{4}$$

with $\mathrm{pep}(\boldsymbol{x}, \boldsymbol{y})$ being the pairwise error (tail) probability

$$\mathrm{pep}(\boldsymbol{x}, \boldsymbol{y}) = \Pr\big[V(\boldsymbol{x}, \boldsymbol{y}) \geq 0\big], \tag{5}$$

and $V(\boldsymbol{x}, \boldsymbol{y})$ is a random variable defined as

$$V(\boldsymbol{x}, \boldsymbol{y}) = \log W^n(\boldsymbol{y}|\overline{\boldsymbol{X}}) - \log W^n(\boldsymbol{y}|\boldsymbol{x}). \tag{6}$$

For discrete memoryless channels, $V(\boldsymbol{x}, \boldsymbol{y})$ is the sum of n terms. As $n \to \infty$, we show in Sec. III.B and Sec. IV.B that Θ is also the sum of n terms, since (4) is asymptotically equivalent to the random variable $Z = nR - i_s(\boldsymbol{X}, \boldsymbol{Y})$, where $i_s(\boldsymbol{x}, \boldsymbol{y})$ is a tilted information density given by

$$i_s(\boldsymbol{x}, \boldsymbol{y}) = \log \frac{W^n(\boldsymbol{y}|\boldsymbol{x})^s}{\mathrm{E}\big[W^n(\boldsymbol{y}|\overline{\boldsymbol{X}})^s\big]}, \tag{7}$$

for $s > 0$. This motivates the use of the saddlepoint method to approximate the tail probabilities (3) and (5),

III. Lattice Case

For discrete input and output alphabets $\mathcal{X}^n$ and $\mathcal{Y}^n$, the random variables $V(\boldsymbol{x},\boldsymbol{y})$ and Z involved in the RCU may be lattice random variables. We say that X is a lattice random variable with distribution $p(x)$ and support set $\mathcal{L}=\{a+h\ell : \ell\in\mathbb{Z}\}$, where $h>0$ is the span and $a\in[0,h)$ is the offset, if $p(x)\geq 0$ for $x\in\mathcal{L}$, $p(x)=0$ for $x\notin\mathcal{L}$, and h is the largest value such that these conditions are satisfied.

A. Inverse Laplace Transform

Let $\boldsymbol{X}=(X_1,\ldots,X_n)$ be a sequence of random variables. We define the random variable $Z=\sum_{i=1}^n X_i$ with probability distribution $p(z)$. If Z is a lattice random variable with offset a and span h, the characteristic function $\varphi(t)$ [18] is given by

$$\varphi(t)=\mathrm{E}\big[e^{jt\cdot Z}\big]=\sum_{k=-\infty}^{+\infty} p_k\cdot e^{jt\cdot z_k}, \tag{8}$$

where $j=\sqrt{-1}$, and we defined the lattice point z_k as $z_k=a+hk$ and its probability mass as $p_k=p(z_k)$. Clearly, $|\varphi(t)|$ is $\frac{2\pi}{h}$-periodic. Then, the probability mass of Z at the lattice point ℓ can be recovered from $\varphi(t)$ by Fourier inversion over one period [19], i.e.,

$$p_\ell=\frac{h}{2\pi}\int_{-\frac{\pi}{h}}^{+\frac{\pi}{h}}\varphi(t)\cdot e^{-jt\cdot z_\ell}\,\mathrm{d}t. \tag{9}$$

Plugging (8) into (9) and interchanging integration and sum, the integrands are zero whenever $k\neq\ell$ and equal to $\frac{2\pi}{h}p_\ell$ only for $k=\ell$. Let $s=jt$. The inversion formula (9) can be written as the inverse Laplace transform [20]

$$p_\ell=\frac{h}{2\pi j}\int_{-j\frac{\pi}{h}}^{+j\frac{\pi}{h}} e^{\kappa(s)-sz_\ell}\,\mathrm{d}s, \tag{10}$$

where $\kappa(s)$ is the cumulant generating function defined as

$$\kappa(s)=\log\mathrm{E}\big[e^{s\cdot Z}\big]=\log\sum_{k=-\infty}^{+\infty}p_k\cdot e^{s\cdot z_k}. \tag{11}$$

Using the Cauchy's integral theorem [21], we may move the integration path to the line segment $\big(\hat{s}-j\frac{\pi}{h},\hat{s}+j\frac{\pi}{h}\big)$, as far as $\hat{s}\in\mathbb{R}$ is within the region of convergence of (11). With this new integration path location, we find a second order Taylor expansion of $\kappa(s)$ around $s=\hat{s}$, and extend the integration path to the whole line that crosses the real axis at $\hat{s}$, i.e., $\big(\hat{s}-j\infty,\hat{s}+j\infty\big)$, to approximate (10) as

$$p_\ell\approx\frac{h}{2\pi j}\int_{\hat{s}-j\infty}^{\hat{s}+j\infty} e^{\kappa(\hat{s})+\kappa'(\hat{s})(s-\hat{s})+\frac{1}{2}\kappa''(\hat{s})(s-\hat{s})^2-s\cdot z_\ell}\,\mathrm{d}s. \tag{12}$$

Making the change of variable $s=\hat{s}+jt$, we obtain

$$p_\ell\approx\frac{h\cdot e^{\kappa(\hat{s})-\hat{s}\cdot z_\ell}}{2\pi}\int_{-\infty}^{\infty}e^{jt\kappa'(\hat{s})-\frac{1}{2}\kappa''(\hat{s})t^2}\cdot e^{-jt\cdot z_\ell}\,\mathrm{d}t. \tag{13}$$

Since $e^{jt\kappa'(\hat{s})-\frac{1}{2}\kappa''(\hat{s})t^2}$ is the characteristic function of a Gaussian random variable with mean $\kappa'(\hat{s})$ and variance $\kappa''(\hat{s})$, the integration in (13) can be solved by Fourier inversion to obtain that p_ℓ may be approximated as

$$p_\ell\approx e^{\kappa(\hat{s})-\hat{s}\cdot z_\ell}\cdot\frac{h}{\sqrt{2\pi\kappa''(\hat{s})}}e^{-\frac{(z_\ell-\kappa'(\hat{s}))^2}{2\kappa''(\hat{s})}}. \tag{14}$$

The parameter $\hat{s}$ allows us to adjust the mean, variance and the exponential tilting of the approximation (14). A classical choice is $\kappa'(\hat{s})=z_0$ for every ℓ when (14) is used to approximate the tail probability $\Pr[Z\geq z_0]$. The point $\hat{s}$ is a saddlepoint and it is unique due to the convexity of $\kappa(s)$.

B. Pairwise Error Probability

Let us assume that the pair $Q(x)$, $W(y|x)$ is non-singular [5, Def. 1]. Then, the random variable $V(\boldsymbol{x},\boldsymbol{y})$ given in (6) has zero offset, and span g independent on $\boldsymbol{x}$ and $\boldsymbol{y}$. For clarity, we define $v_\ell=g\ell$, and denote its probability mass as $p_\ell(\boldsymbol{x},\boldsymbol{y})$. Then, equation (5) becomes

$$\mathrm{pep}(\boldsymbol{x},\boldsymbol{y})=\sum_{\ell=0}^{+\infty}p_\ell(\boldsymbol{x},\boldsymbol{y}), \tag{15}$$

The saddlepoint approximation of $p_\ell(\boldsymbol{x},\boldsymbol{y})$ is given by (14), with the cumulant generating function of $V(\boldsymbol{x},\boldsymbol{y})$, given by

$$\kappa_s(\boldsymbol{x},\boldsymbol{y})=\log\mathrm{E}\Bigg[\bigg(\frac{W^n(\boldsymbol{y}|\overline{\boldsymbol{X}})}{W^n(\boldsymbol{y}|\boldsymbol{x})}\bigg)^s\Bigg]. \tag{16}$$

Here, we place s as subindex to highlight the dependence on $\boldsymbol{x}$ and $\boldsymbol{y}$. Is is straightforward that (16) is related to the tilted information density (7) as $\kappa_s(\boldsymbol{x},\boldsymbol{y})=-i_s(\boldsymbol{x},\boldsymbol{y})$. For a given $\hat{s}$, plugging (14) into (15), we obtain that

$$\mathrm{pep}(\boldsymbol{x},\boldsymbol{y})\approx\gamma_{\hat{s}}(\boldsymbol{x},\boldsymbol{y})\cdot e^{-i_{\hat{s}}(\boldsymbol{x},\boldsymbol{y})}, \tag{17}$$

where the pre-exponential factor $\gamma_{\hat{s}}(\boldsymbol{x},\boldsymbol{y})$ is given by

$$\gamma_{\hat{s}}(\boldsymbol{x},\boldsymbol{y})=\frac{g}{\sqrt{2\pi\kappa''_{\hat{s}}(\boldsymbol{y})}}\sum_{\ell=0}^{+\infty}e^{-\hat{s}\cdot v_\ell-\frac{(v_\ell-\kappa'_{\hat{s}}(\boldsymbol{x},\boldsymbol{y}))^2}{2\kappa''_{\hat{s}}(\boldsymbol{y})}}. \tag{18}$$

We remark that $\kappa'_{\hat{s}}(\boldsymbol{x},\boldsymbol{y})$ is the first derivative of $\kappa_s(\boldsymbol{x},\boldsymbol{y})$ at $s=\hat{s}$, and that $\kappa''_{\hat{s}}(\boldsymbol{y})$ is its second derivative, also at $s=\hat{s}$, that only depends on $\boldsymbol{y}$. Since $\kappa_s(\boldsymbol{x},\boldsymbol{y})$ is the sum of n terms, we note that $\gamma_{\hat{s}}(\boldsymbol{x},\boldsymbol{y})$ decays as $\frac{1}{\sqrt{n}}$. Finally, as the tail probability (15) is evaluated at $V(\boldsymbol{x},\boldsymbol{y})=0$, the saddlepoint $\hat{s}$ satisfying $\kappa'_{\hat{s}}(\boldsymbol{x},\boldsymbol{y})=0$ clearly depends on $\boldsymbol{x}$ and $\boldsymbol{y}$.

C. Error Exponent

Let Θ be the random variable defined in (4). Using the approximation (17), we observe that Θ is asymptotically equivalent[1] to the random variable Z given by

$$Z=nR-i_{\hat{s}}(\boldsymbol{X},\boldsymbol{Y}), \tag{19}$$

where $\hat{s}$ is the saddlepoint of the pairwise error probability discussed in Sec. III-B. The Chernoff bound to the tail probability $\Pr[\Theta\geq\log U]$ allows to obtain the exponential decay of the RCU as the following upper bound

$$\lim_{n\to\infty}\frac{1}{n}\log\mathrm{rcu}\leq\inf_{0\leq\rho\leq 1}\lim_{n\to\infty}\frac{1}{n}\phi_n(\rho), \tag{20}$$

[1] A_n is asymptotically equivalent to B_n, $A_n\asymp B_n$, if $\lim_{n\to\infty}\frac{A_n}{B_n}=1$.

where $\phi_n(\rho)$ is the cumulant generating function of Θ. As discussed in [11, Sec. III], the convergence of $\Pr[\Theta \geq \log U]$ requires $0 \leq \rho \leq 1$. Since $\Theta \asymp Z$, $\phi_n(\rho)$ is asymptotically equivalent to the cumulant generating function of Z, given by

$$\chi_n(\rho) = n\rho R + \log \mathrm{E}\big[e^{-\rho i_{\hat{s}}(\boldsymbol{X},\boldsymbol{Y})}\big]. \tag{21}$$

For an i.i.d. codebook and a memoryless channel, the bound in (20) is optimized by setting $\hat{s} = \frac{1}{1+\rho}$. This differs from our previous work [6] where $\hat{s}$ is optimized independently of ρ, and also from the analysis in [7] that considers a sequence-dependent optimization parameter. Using this choice of $\hat{s}$, we recover the random coding error exponent, i.e., (21) becomes

$$\chi_n(\rho) = n\rho R - nE_0(\rho), \tag{22}$$

where $E_0(\rho)$ is the Gallager function [2, Eq. (5.6.14)]. Let $\hat{\rho}$ be the optimizer of the bound (20), or equivalently the unique solution to $E_0'(\rho) = R$. Then, the range $1 \geq \hat{\rho} \geq 0$ corresponds to the rate range $R_{\mathrm{c}}(Q) \leq R \leq I(Q)$, where $R_{\mathrm{c}}(Q)$ and $I(Q)$ are the critical rate and the mutual information, respectively, for the random coding ensemble $Q(x)$ [2, Eq. (5.6.30)].

D. Random Coding Union Bound

We now turn back to the refined asymptotics analysis to derive a saddlepoint approximation of the RCU. We let a and h as the offset and span of the random variable Z given by (19), and define the lattice points $z_\ell = a + h\ell$ for $\ell \in \mathbb{Z}$. It is convenient to write the RCU in terms of Θ given in (4), i.e.,

$$\mathrm{rcu} = \mathrm{E}\big[\mathbb{1}\{\Theta \geq 0\}\big] + \mathrm{E}\big[e^{\Theta}\mathbb{1}\{\Theta < 0\}\big], \tag{23}$$

The cumulant generating function of Θ, $\phi_n(\rho)$, and the cumulant generating function of Z, $\chi_n(\rho)$, satisfy

$$\phi_n(\rho) = \chi_n(\rho) + \pi_n(\rho), \tag{24}$$

where $\pi_n(\rho)$ is the term from the pre-exponential factor of the pairwise error probability. Using the pairwise error probability approximation (17), we obtain that $\pi_n(\rho)$ is approximated as

$$\pi_n(\rho) \approx \log \mathrm{E}\big[\gamma_{\hat{s}}(\boldsymbol{X}_\rho, \boldsymbol{Y}_\rho)^\rho\big], \tag{25}$$

with $\boldsymbol{X}_\rho$ and $\boldsymbol{Y}_\rho$ distributed according to

$$Q_\rho^n(\boldsymbol{x})W_\rho^n(\boldsymbol{y}|\boldsymbol{x}) \propto Q^n(\boldsymbol{x})W^n(\boldsymbol{y}|\boldsymbol{x})e^{-\rho i_{\hat{s}}(\boldsymbol{x},\boldsymbol{y})}. \tag{26}$$

Upon the optimization of the error exponent by $E_0(\hat{\rho}) = R$, our analysis suggests that the rate R determines a distribution of *typical* sequences $(\boldsymbol{x},\boldsymbol{y})$ through the saddlepoint $\hat{\rho}$, and in turn a set of *typical* sequences $\overline{\boldsymbol{x}}$ in the pairwise error probability through $\hat{s} = \frac{1}{1+\hat{\rho}}$. This allows to recover the error exponent, i.e., $\chi_n(\hat{\rho})$, and to study the contribution of the pre-exponential term of the pairwise error probability, i.e., $\pi_n(\hat{\rho})$, around the typical sequences $(\boldsymbol{x},\boldsymbol{y})$. After several manipulations omitted in this paper for the sake of space, it can be shown that the expectation (25) under the tilted probability distribution (26) is asymptotically equivalent to

$$\pi_n(\rho) \asymp \rho \log \psi_n(\rho) - \frac{1}{2}\log(1+\rho), \tag{27}$$

where the $\log(1+\rho)$ term comes from the quadratic exponential in (18), and $\psi_n(\rho)$ is given by

$$\psi_n(\hat{\rho}) = \frac{g}{\sqrt{2\pi n \overline{\kappa}''(\hat{\rho})}} \cdot \left(\frac{1}{1 - e^{-\hat{s}g}}\right) \tag{28}$$

where $\overline{\kappa}''(\hat{\rho})$ is given by

$$\overline{\kappa}''(\hat{\rho}) = \sum_y P_{\hat{\rho}}(y)\left[\frac{\partial^2}{\partial s^2}\Big(\log \sum_x Q(x)W(y|x)^s\Big)\bigg|_{s=\hat{s}}\right], \tag{29}$$

with $P_\rho(y)$ the single letter marginal distribution of (26).

Since Z is the sum of n independent terms, we note that $\chi_n(\rho)$ is linear with n, whereas $\pi_n(\rho)$ has order $\log n$. Hence, the saddlepoint approximation to the RCU involves finding the second order Taylor expansion of $\chi_n(\rho)$ around $\rho = \hat{\rho}$, i.e.,

$$\chi_n(\rho) \approx n\hat{\rho}R - nE_0(\hat{\rho}) - \frac{1}{2}nV(\hat{\rho})(\rho - \hat{\rho})^2 \tag{30}$$

where we used that $E_0'(\hat{\rho}) = R$. In (30), $V(\hat{\rho})$ is the variance of the tilted information density (7) given by

$$V(\hat{\rho}) = -E_0''(\hat{\rho}). \tag{31}$$

For $\hat{\rho} = 0$, i.e., for $R = I(Q)$, $V(\hat{\rho})$ coincides with the channel dispersion [17, Eq. (304)].

Using (24) and (27), the probability mass p_ℓ can be recovered from the cumulant generating function $\phi_n(\rho)$ as

$$p_\ell = \frac{h}{2\pi j}\int_{-j\frac{\pi}{h}}^{+j\frac{\pi}{h}} \frac{e^{\chi_n(\rho) - \rho(\theta_\ell - \log\psi_n(\rho))}}{\sqrt{1+\rho}}\,\mathrm{d}\rho. \tag{32}$$

Spelling out equation (23), and making the change of variable $\theta_\ell = z_\ell + \log\psi_n(\rho)$, we have that

$$\mathrm{rcu} = \sum_{\ell=\ell^\star}^{+\infty} p_\ell + \sum_{\ell=-\infty}^{\ell^\star-1} e^{\theta_\ell}\cdot p_\ell \tag{33}$$

$$= \sum_{\ell=\ell^\star}^{+\infty} p_\ell + \psi_n(\rho)\sum_{\ell=-\infty}^{\ell^\star-1} e^{z_\ell}\cdot p_\ell \tag{34}$$

where $\ell^\star$ is the smallest $\ell \in \mathbb{Z}$ such that $\theta_\ell \geq 0$. Placing the Taylor expansion (30) on p_ℓ, approximating $\psi_n(\rho) \approx \psi_n(\hat{\rho})$ and $\sqrt{1+\rho} \approx \sqrt{1+\hat{\rho}}$, solving the complex integration (32), and using the approximated p_ℓ in (34), we obtain the following approximation for the RCU.

Approximation 1 (Lattice case): For rates such that $R_{\mathrm{c}}(Q) < R < I(Q)$, let $0 < \hat{\rho} < 1$ be the unique solution to $E_0'(\hat{\rho}) = R$. Then, the RCU (2) can be approximated as

$$\mathrm{rcu} \approx \alpha_n \cdot e^{-n(E_0(\hat{\rho}) - \hat{\rho}R)}, \tag{35}$$

where

$$\alpha_n = \sum_{\ell=\ell^\star}^{+\infty} e^{-\hat{\rho}z_\ell}\cdot\frac{h\cdot e^{-\frac{z_\ell^2}{2nV(\hat{\rho})}}}{\sqrt{2\pi nV(\hat{\rho})(1+\hat{\rho})}} + \\ + \psi_n(\hat{\rho})\sum_{\ell=-\infty}^{\ell^\star-1} e^{(1-\hat{\rho})z_\ell}\cdot\frac{h\cdot e^{-\frac{z_\ell^2}{2nV(\hat{\rho})}}}{\sqrt{2\pi nV(\hat{\rho})(1+\hat{\rho})}}. \tag{36}$$

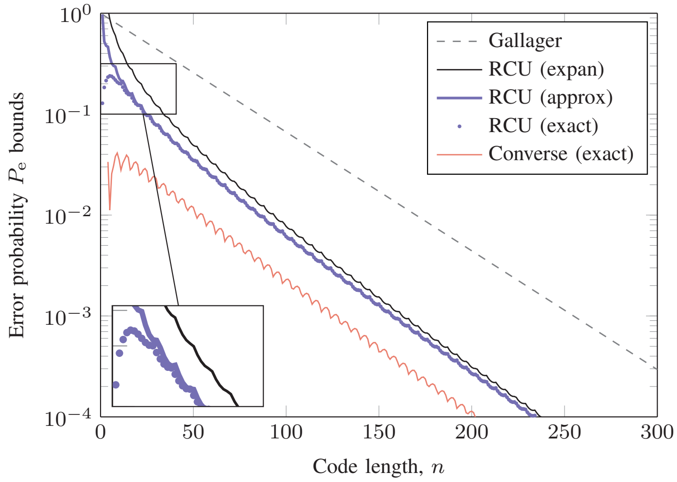

Fig. 1. BSC error probability bounds vs block-length n with $R = 0.1$ bits per channel use and crossover probability $\delta = 0.11$.

The computation of α_n as in (36) involves the infinite summation of quadratic terms. An alternative form of α_n involves finding its asymptotics neglecting the quadratic exponential terms of (36), and obtaining the following asymptotic result.

Expansion 1 (Lattice case): For rates such that $R_c(Q) < R < I(Q)$, as $n \to \infty$, we have

$$\alpha_n \asymp \frac{h \cdot \psi_n(\hat{\rho})^{\hat{\rho}}}{\sqrt{2\pi n V(\hat{\rho})(1+\hat{\rho})}} \cdot \left(e^{-\hat{\rho}\theta_{\ell^\star}} \left(\frac{1}{1-e^{-\hat{\rho}h}} \right) + \right.$$
$$\left. + e^{(1-\hat{\rho})\theta_{\ell^\star}} \left(\frac{e^{-(1-\hat{\rho})h}}{1-e^{-(1-\hat{\rho})h}} \right) \right), \quad (37)$$

where $\psi_n(\hat{\rho})$ is given in (28) and $\theta_{\ell^\star} = z_{\ell^\star} + \log \psi_n(\hat{\rho})$.

From (28) and (37), we observe that the pre-exponential term α_n of the RCU decays as $n^{-\frac{1+\hat{\rho}}{2}}$, matching the results of [4]–[7]. The improvement of (37) with respect to [6, Eq. (135)] is the refinement of $\pi_n(\rho)$ in (27) due to the dependence between $\hat{s}$ and $\hat{\rho}$.

E. Binary Symmetric Channel

We numerically evaluate our result (35) using α_n given in (36) and (37), respectively the RCU (approx) and the RCU (expan). We consider a binary symmetric channel (BSC) with crossover probability $\delta = 0.11$ and uniform input distribution $Q(x)$ at a rate $R = 0.3$ bits per channel use. As a reference, we include the Gallager bound [2, Eq. (5.6.18)], the exact RCU [17, Eq. (162)], and the converse bound [17, Eq. (137)].

We note that $\log W^n(\boldsymbol{y}|\boldsymbol{x}) = (n-t)\log(1-\delta) + t\log\delta$, where t is the Hamming distance between $\boldsymbol{x}$ and $\boldsymbol{y}$. As a result, the random variable $V(\boldsymbol{x},\boldsymbol{y})$ defined in (6) lies in a lattice of zero offset and span $g = \log\frac{1-\delta}{\delta}$, whereas Z in (19) has a non-zero offset and span $h = s \cdot g$. The explicit expressions of the terms in the RCU expansion (35) are related to the cumulants of Z and $V(\boldsymbol{x},\boldsymbol{y})$ given in (22) and (16), respectively.

Fig. 1 shows that both the RCU approximation and the RCU expansion capture the rippling effect of the exact RCU, and that the RCU approximation, with α_n in (36), accurately approximates the RCU even for small values of n.

IV. Strongly Non-Lattice Case

A. Inverse Laplace Transform

We turn into the case where an arbitrary random variable $Z = \sum_{i=1}^n X_i$ is strongly non-lattice [18], with probability distribution $p(z)$. In this case, the characteristic function of Z either satisfies $|\varphi(t)| < 1$ for all $t \neq 0$, or $p(z)$ is concentrated at a single point. Since $|\varphi(t)|$ is not periodic, Fourier inversion is now performed as

$$p(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi(t) \cdot e^{-jt\cdot z}. \quad (38)$$

Mimicking the developments for the lattice case, we write (38) as an inverse Laplace transform involving the cumulant generating function $\kappa(s)$ given by

$$\kappa(s) = \log \mathrm{E}[e^{s\cdot Z}] = \log \int_{-\infty}^{+\infty} p(z)\cdot e^{s\cdot z}, \quad (39)$$

move the integration path to $(\hat{s} - j\infty, \hat{s} + j\infty)$, Taylor expand $\kappa(s)$, and solve the complex integration to obtain

$$p(z) \approx e^{\kappa(\hat{s}) - \hat{s}z} \cdot \frac{1}{\sqrt{2\pi\kappa''(\hat{s})}} e^{-\frac{(z-\kappa'(\hat{s}))^2}{2\kappa''(\hat{s})}}. \quad (40)$$

B. Pairwise Error Probability

For a given $\boldsymbol{x}$ and $\boldsymbol{y}$, the pairwise error probability is the upper tail probability of the random variable $V(\boldsymbol{x},\boldsymbol{y})$ given in (6), with probability distribution $p(v)$, i.e.,

$$\mathrm{pep}(\boldsymbol{x},\boldsymbol{y}) = \int_0^{+\infty} p(v)\,\mathrm{d}v. \quad (41)$$

Applying the saddlepoint approximation (40) on $p(v)$ and using it on (41), we obtain

$$\mathrm{pep}(\boldsymbol{x},\boldsymbol{y}) \approx \gamma_{\hat{s}}(\boldsymbol{x},\boldsymbol{y}) \cdot e^{-i_{\hat{s}}(\boldsymbol{x},\boldsymbol{y})}, \quad (42)$$

where $i_{\hat{s}}(\boldsymbol{x},\boldsymbol{y})$ is the tilted information density (7), and for the strongly non-lattice case $\gamma_{\hat{s}}(\boldsymbol{x},\boldsymbol{y})$ is given by,

$$\gamma_{\hat{s}}(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{\sqrt{2\pi\kappa''_{\hat{s}}(\boldsymbol{y})}} \int_0^{+\infty} e^{-\hat{s}\cdot v - \frac{(v-\kappa'_{\hat{s}}(\boldsymbol{x},\boldsymbol{y}))^2}{2\kappa''_{\hat{s}}(\boldsymbol{y})}}\,\mathrm{d}v. \quad (43)$$

C. Random Coding Union Bound

The error exponent and cumulant generating function analysis we discussed in Sec. III-D also holds for the strongly non-lattice case, up to equation (31). Recalling that the RCU can be written as in (23), we now have that the counterpart of (34) is given by

$$\mathrm{rcu} = \int_{-\log\psi_n(\hat{\rho})}^{+\infty} p(z)\,\mathrm{d}z + \psi_n(\hat{\rho}) \int_{-\infty}^{-\log\psi_n(\hat{\rho})} e^z \cdot p(z)\,\mathrm{d}z, \quad (44)$$

where $p(\theta)$ and $p(z)$ are the probability density of Θ and Z, respectively, and now

$$\psi_n(\hat{\rho}) = \frac{1}{\sqrt{2\pi n\overline{\kappa}''(\hat{\rho})}} \cdot \frac{1}{\hat{s}}, \quad (45)$$

with $\overline{\kappa}''(\hat{\rho})$ is given by (29) substituting the summations by integrations. Applying (40) on $p(z)$ with the cumulant generating function (22) and Taylor expansion (30), we obtain the following approximation for the RCU.

Approximation 2 (Strongly non-lattice case): The RCU (2), with rates satisfying $R_{\rm c}(Q) < R < I(Q)$, is approximated as

$$\mathrm{rcu} \approx \alpha_n \cdot e^{-n(E_0(\hat{\rho})-\hat{\rho}R)}, \tag{46}$$

where $\hat{\rho}$ is the unique solution to $E_0'(\hat{\rho}) = R$, and

$$\alpha_n = \int_{-\log\psi_n(\hat{\rho})}^{+\infty} e^{-\hat{\rho}z_\ell} \cdot \frac{e^{-\frac{z_\ell^2}{2nV(\hat{\rho})}}}{\sqrt{2\pi nV(\hat{\rho})(1+\hat{\rho})}}\,\mathrm{d}z + \\ + \psi_n(\hat{\rho}) \int_{-\infty}^{-\log\psi_n(\hat{\rho})} e^{(1-\hat{\rho})z_\ell} \cdot \frac{e^{-\frac{z_\ell^2}{2nV(\hat{\rho})}}}{\sqrt{2\pi nV(\hat{\rho})(1+\hat{\rho})}}\,\mathrm{d}z. \tag{47}$$

The evaluation of α_n can be done efficiently using error function tables, though a simpler expression is obtained following the same reasoning as in the lattice case. Specifically, setting $\frac{1}{2nV(\hat{\rho})} = 0$ in (47), we obtain the following result.

Expansion 2 (Strongly non-lattice case): As $n \to \infty$, the asymptotics of the term α_n in (47) is given by

$$\alpha_n \asymp \frac{\psi_n(\hat{\rho})^{\hat{\rho}}}{\sqrt{2\pi nV(\hat{\rho})(1+\hat{\rho})}} \cdot \frac{1}{\hat{\rho}(1-\hat{\rho})} \tag{48}$$

valid for rates such that $R_{\rm c}(Q) < R < I(Q)$.

We note that the asymptotics of $\psi_n(\hat{\rho})$ and α_n in equations (45) and (48) can also be obtained from the asymptotics in the lattice case in equations (28) and (37) by taking the limits as the lattice spans tend to zero, i.e., $g \to 0$ and $h \to 0$.

D. Binary Input AWGN Channel

We finally illustrate the strongly non-lattice case with the binary input additive Gaussian noise (AWGN) channel, with uniform input distribution $Q(x)$, input alphabet $\{-\sqrt{\mathrm{snr}}, +\sqrt{\mathrm{snr}}\}$, and unit noise variance, for a signal-to-noise ratio of $\mathrm{snr} = 4$ dB. We also include the Gallager bound, a Monte Carlo simulation of the exact RCU (2), and the meta-converse bound [17, Th. 27] with the exponent achieving distribution using the method presented in [16].

The numerical example in Fig. 2 shows that both the approximated RCU and the expanded RCU computed through equations (47) and (48), respectively, are simple and accurate tools to estimate the RCU, especially in channels whose exact evaluation or numerical simulation is computationally hard.

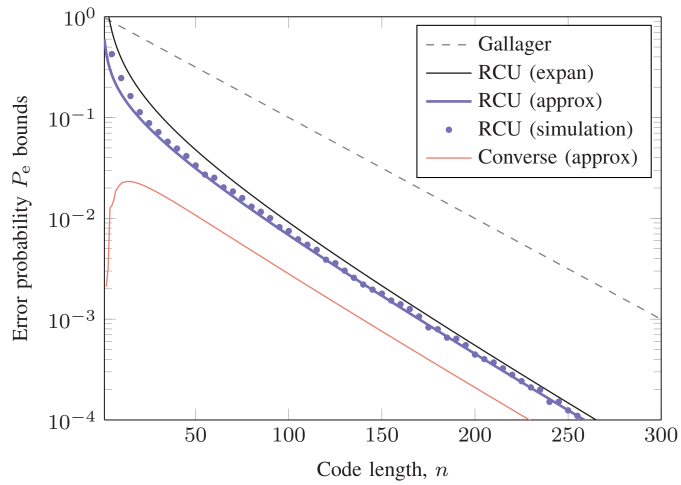

Fig. 2. Binary input AWGN channel error probability bounds vs block-length n with $R = 0.65$ bits per channel use and $\mathrm{snr} = 4$ dB.

References

[1] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. Journal*, vol. 27, pp. 379–423, May 1948.

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

[3] ——, "The random coding bound is tight for the average code," *IEEE Trans. Inf. Theory*, vol. 19, no. 2, pp. 244–246, Mar. 1973.

[4] R. L. Dobrushin, "Asymptotic estimates of the probability of error for transmission of messages over a discrete memoryless communication channel with a symmetric transition probability matrix," *Theory Probab. Appl.*, vol. 7, no. 3, pp. 270–300, 1962.

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

[6] 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, 2014.

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

[8] M. Tomamichel and V. Y. Tan, "Second-order coding rates for channels with state," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4427–4448, May 2014.

[9] T. Erseghe, "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.

[10] P. Moulin, "The log-volume of optimal codes for memoryless channels, asymptotically within a few nats," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2278–2313, 2017.

[11] A. Martinez and A. Guillén i Fàbregas, "Saddlepoint approximation of random–coding bounds." in *Information Theory and Applications (ITA) Workshop*, 2011, pp. 257–262.

[12] 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.

[13] J. Font-Segura, A. Martinez, and A. Guillén i Fàbregas, "Refined error probability approximations in quasi-static erasure channels," in *Int. Zurich Seminar on Commun. (IZS)*, Zurich, Switzerland, Mar. 2016.

[14] ——, "Asymptotics of the random-coding union bound in quasi-static fading channels," in *IEEE Inf. Theory Workshop (ITW)*, Cambridge, UK, Sep. 2016.

[15] ——, "Asymptotics of the error probability in quasi-static binary symmetric channels," in *IEEE Int. Symp. on Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017.

[16] J. Font-Segura, G. Vazquez-Vilar, A. Martinez, A. Guillén i Fàbregas, and A. Lancho, "Saddlepoint approximations in random coding," in *Conf. on Inf. Science and Sys. (CISS)*, Princeton, NJ, Mar. 2018.

[17] 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.

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

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

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

[21] F. W. J. Olver, *Asymptotics and Special Functions*. New York, NY: Academic Press, 1974.