Asymptotics of the Random Coding Error Probability for Constant-Composition Codes

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 to the error probability are derived for multiple-cost-constrained random coding ensembles where codewords satisfy a set of constraints. Constant-composition inputs over a binary symmetric channel are studied as a particular case. For codewords with equiprobable empirical distribution, the analysis recovers the same error exponent and pre-exponential polynomial decay as the uniform i.i.d. ensemble and provides an explicit formula for the loss in prefactor (third-order term) incurred by the constant-composition ensemble.**

I. Introduction

We consider random coding over a discrete memoryless channel $W(y|x)$ with input $x \in \{0, \dots, J-1\}$ and output $y \in \{0, \dots, K-1\}$. Input sequences (codewords) $\boldsymbol{x}$ and output sequences $\boldsymbol{y}$ of length n are probabilistically related by the channel law $W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^n W(y_i|x_i)$.

For a given product distribution $Q^n(\boldsymbol{x}) = Q(x_1) \cdots Q(x_n)$ over the input alphabet, in the multiple-cost-constrained random coding ensemble, the M codewords are randomly generated according to the probability distribution

$$P(\boldsymbol{x}) = \frac{1}{\mu_n} Q^n(\boldsymbol{x}) \mathbb{1}\{\boldsymbol{x} \in \mathcal{D}_1\} \cdots \mathbb{1}\{\boldsymbol{x} \in \mathcal{D}_L\}. \qquad (1)$$

Here, μ_n is a normalization factor and the indicator functions $\mathbb{1}\{\cdot\}$ enforce that codewords satisfy L constraints defined by cost functions $c_\ell(\boldsymbol{x})$ and sets $\mathcal{D}_\ell$ for $\ell = 1, \dots, L$, i.e.,

$$\mathcal{D}_\ell = \{\boldsymbol{x} : |c_\ell(\boldsymbol{x})| \leq a_\ell\}. \qquad (2)$$

The constant-composition ensemble is a particular case of the cost-constrained ensemble (1) when the constraint sets (2) are chosen to contain all codewords having a given empirical distribution [1, Ch. 2]. Both cost-constrained and constant-composition ensembles usually lead to smaller error probability than the i.i.d. ensemble [2]–[4]. However, for some cases [5] such as the constant composition code discussed in Sec. III there is a performance loss over the i.i.d. ensemble, where no cost constraints are imposed.

To assess the performance of the ensemble (1), we study the random coding union (RCU) bound to the average error probability over the ensemble [6, Th. 16], given by

$$\mathrm{rcu} = \sum_{\boldsymbol{x}\boldsymbol{y}} P(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) \min\bigl\{1, (M-1)\mathrm{pep}(\boldsymbol{x}, \boldsymbol{y})\bigr\}, \quad (3)$$

where the pairwise error probability is the probability that of randomly drawing a codeword $\overline{\boldsymbol{x}}$ with larger likelihood than the transmitted one, i.e.,

$$\mathrm{pep}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\overline{\boldsymbol{x}}} P(\overline{\boldsymbol{x}}) \mathbb{1}\bigl\{W^n(\boldsymbol{y}|\overline{\boldsymbol{x}}) \geq W^n(\boldsymbol{y}|\boldsymbol{x})\bigr\}. \qquad (4)$$

Since the average error probability over all codes randomly generated from (1) is upper bounded by the RCU, there must exist a code of length n and M codewords satisfying the set of conditions (2) whose error probability is at most (3).

For fixed code rate $R = \frac{1}{n} \log M$, refinements on the analysis of the decay of the vanishing error probability as $n \to \infty$ have received recent interest. For example, the works [7]–[10] used Edgeworth expansions, saddlepoint approximations, and large deviations methods to study the polynomial decay of the error probability exponential prefactor. Our previous results [11]–[13] differ from the aforementioned works in the treatment of the coupling between (3) and (4). In this work we assume that the channel $W(y|x)$ is non-singular [8, Def. 1] and derive an asymptotic closed-form approximation of (3) for the multiple-cost-constrained ensemble (1).

In Sec. III we apply our analysis to constant-composition codes with empirical distribution $\frac{1}{2}$ over the binary symmetric channel. While the error probability of this ensemble has the same exponential decay and pre-exponential polynomial decay as the uniform i.i.d. ensemble, it has a larger prefactor (third-order term or the constant in front of the exponential), and its error probability is thus worse than that of the i.i.d. ensemble.

II. Error Probability Asymptotics

A. Main Result

Let us assume that all the constants a_ℓ in the set defintions (2) do not depend on n, and that all the cost functions $c_\ell(\boldsymbol{x})$ have zero mean under the i.i.d. probability distribution $Q(x)$. Under maximum likelihood decoding, the RCU bound (3) for the multiple-cost-constrained ensemble (1) with fixed rate R is asymptotically approximated by

$$\hat{\mathrm{rcu}}_{\mathrm{cc}} = \alpha_{\mathrm{cc}}(\hat{\boldsymbol{\rho}})\, e^{-n(E_0^{\mathrm{cc}}(\hat{\boldsymbol{\rho}}) - \hat{\rho}_0 R)}, \qquad (5)$$

where $\boldsymbol{\rho} = (\rho_0, \dots, \rho_L)$ is a vector of parameters,

$$E_0^{\mathrm{cc}}(\boldsymbol{\rho}) = -\log \sum_y \Biggl(\sum_x Q(x) W(y|x)^{\frac{1}{1+\rho_0}} e^{\sum_{\ell=1}^L \rho_\ell c_\ell(x)}\Biggr)^{1+\rho_0} \qquad (6)$$

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 grant TEC2016-78434-C3-1-R.

is a multiple-cost-constrained Gallager function, the optimized $\hat{\boldsymbol{\rho}}$ is given by the solution of

$$\frac{\partial}{\partial \rho_0} E_0^{\text{cc}}(\boldsymbol{\rho})\Big|_{\rho_0=\hat{\rho}_0} = R \tag{7}$$

$$\frac{\partial}{\partial \rho_\ell} E_0^{\text{cc}}(\boldsymbol{\rho})\Big|_{\rho_\ell=\hat{\rho}_\ell} = 0, \tag{8}$$

for $\ell = 1, \ldots, L$, and $\alpha_{\text{cc}}(\hat{\boldsymbol{\rho}})$ is the prefactor

$$\alpha_{\text{cc}}(\hat{\boldsymbol{\rho}}) = \frac{1}{\mu_n} \sum_{m_0 \in \mathcal{F}_0} \sum_{m_1 \in \mathcal{F}_1} \cdots \sum_{m_L \in \mathcal{F}_L} q(\boldsymbol{\phi}_{\boldsymbol{m}}) + \\ + \frac{\psi_n(\hat{\boldsymbol{\rho}})}{\mu_n} \sum_{m_0 \in \mathcal{F}_0^c} \sum_{m_1 \in \mathcal{F}_1} \cdots \sum_{m_L \in \mathcal{F}_L} e^{\phi_{m_0}} q(\boldsymbol{\phi}_{\boldsymbol{m}}), \tag{9}$$

with $\mathcal{F}_0$ and $\mathcal{F}_\ell$ defined in (44) and (45). In (9), $\boldsymbol{\phi}_{\boldsymbol{m}}$ are the $(L+1)$–dimensional lattice points $\boldsymbol{\phi}_{\boldsymbol{m}}$ of the random variables

$$\Phi_0 = nR - i_{\boldsymbol{s}(\hat{\boldsymbol{\rho}})}(\boldsymbol{X}; \boldsymbol{Y}) \tag{10}$$

$$\Phi_\ell = c_\ell(\boldsymbol{X}), \tag{11}$$

where as usual capital letters indicate random variables, $i_{\boldsymbol{s}}(\boldsymbol{x}; \boldsymbol{y})$ is a generalized information density [6, Eq. (4)] for the multiple-cost-constrained ensemble

$$i_{\boldsymbol{s}}(\boldsymbol{x}; \boldsymbol{y}) = \log \frac{W^n(\boldsymbol{y}|\boldsymbol{x})^{s_0}}{\sum_{\overline{\boldsymbol{x}}} Q^n(\overline{\boldsymbol{x}}) W^n(\boldsymbol{y}|\overline{\boldsymbol{x}})^{s_0} e^{s_1 c_1(\overline{\boldsymbol{x}})} \cdots e^{s_L c_L(\overline{\boldsymbol{x}})}}, \tag{12}$$

with optimal tilting $\boldsymbol{s}(\hat{\boldsymbol{\rho}}) = (s_0, \ldots, s_L)$ related to $\hat{\boldsymbol{\rho}}$ as $s_0 = \frac{1}{1+\hat{\rho}_0}$ and $s_\ell = \hat{\rho}_\ell$, and $q(\boldsymbol{\phi}_{\boldsymbol{m}})$ is the saddlepoint approximation to the probability mass at the lattice point $\boldsymbol{\phi}_{\boldsymbol{m}}$, given by

$$q(\boldsymbol{\phi}_{\boldsymbol{m}}) = \Gamma \cdot \frac{e^{\xi_n(\hat{\boldsymbol{\rho}}) - \hat{\boldsymbol{\rho}}^T \boldsymbol{\phi}_{\boldsymbol{m}} - \frac{1}{2n} \boldsymbol{\phi}_{\boldsymbol{m}}^T \boldsymbol{V}_{\hat{\boldsymbol{\rho}}}^{-1} \boldsymbol{\phi}_{\boldsymbol{m}}}}{\sqrt{(2\pi \,|\det(n\boldsymbol{V}_{\hat{\boldsymbol{\rho}}})|}} \tag{13}$$

where $\boldsymbol{V}_{\hat{\boldsymbol{\rho}}}$ is the Hessian matrix of the multiple-cost-constrained Gallager function $E_0^{\text{cc}}(\boldsymbol{\rho})$ at $\boldsymbol{\rho} = \hat{\boldsymbol{\rho}}$. The details on the computation of the remaining parameters, i.e., $\psi_n(\hat{\boldsymbol{\rho}})$, $\xi_n(\hat{\boldsymbol{\rho}})$ and Γ can be found in equations (40), (42) and (47).

In Sec. II-B, we also show that the polynomial decay of $\alpha_{\text{cc}}(\hat{\boldsymbol{\rho}})$ is $n^{-(1+\hat{\rho}_0)/2}$ regardless the number of constraints L. This result matches the behaviour of the unconstrained case (see, e.g., the discussion in [8, Rem. 2]). We finally note that setting $L = 0$ recovers our saddlepoint approximation of the RCU bound for the unconstrained (i.i.d.) random coding ensemble $Q(x)$ reported in [12, Eq. (35)].

As anticipated in Sec. I, the multiple-cost-constrained Gallager function (6) may have a higher error exponent than that of the i.i.d. ensemble. However, for channels and cost functions such that $\hat{\rho}_\ell = 0$ for $\ell = 1, \ldots, L$, the function $E_0^{\text{cc}}(\boldsymbol{\rho})$ reduces to the i.i.d. $E_0(\rho)$ Gallager function [14], implying that no gain or loss is attained using the cost constrained ensemble over the i.i.d. ensemble in terms of error exponent. Since the prefactor term $\alpha_{\text{cc}}(\hat{\boldsymbol{\rho}})$ has the same polynomial decay as its i.i.d. counterpart $\alpha_{\text{iid}}(\hat{\rho}_0)$ [12, Eq. (36)], the difference in performance, a small loss in the binary-symmetric channel, is fully characterized by the analysis of the ratio of prefactors.

B. Proof

The method of proof extends the derivations in our previous work on unconstrained strongly non-lattice [11], unconstrained lattice [12] and cost-constrained strongly non-lattice [13] distributions to lattice distributions with multiple cost constraints.

We start by studying the pairwise error probability $\text{pep}(\boldsymbol{x}, \boldsymbol{y})$. For sake of clarity, we may leave the dependence on $\boldsymbol{x}$ and $\boldsymbol{y}$ implicit in some of the following derivations. Combining (1) and (4), we have that

$$\text{pep}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{\mu_n} \sum_{\overline{\boldsymbol{x}}} Q^n(\overline{\boldsymbol{x}}) \mathbb{1}\big\{\overline{\boldsymbol{x}} \in \mathcal{D}_1 \cap \cdots \cap \mathcal{D}_L \cap \mathcal{E}(\boldsymbol{x}, \boldsymbol{y})\big\} \tag{14}$$

where we defined the log-likelihood decoding error event

$$\mathcal{E}(\boldsymbol{x}, \boldsymbol{y}) = \big\{\overline{\boldsymbol{x}} \ : \log W^n(\boldsymbol{y}|\overline{\boldsymbol{x}}) \geq \log W^n(\boldsymbol{y}|\boldsymbol{x})\big\}. \tag{15}$$

Let $\boldsymbol{Z} = (Z_0, Z_1, \ldots, Z_L)$ be the random vector given by

$$Z_0 = \log W^n(\boldsymbol{y}|\overline{\boldsymbol{X}}) - \log W^n(\boldsymbol{y}|\boldsymbol{x}), \tag{16}$$

$$Z_\ell = c_\ell(\overline{\boldsymbol{X}}), \tag{17}$$

for $\ell = 1, \ldots, L$. Since $\overline{\boldsymbol{x}}$ belongs to a discrete alphabet, the random variables $Z_0, \ldots, Z_L$ may lie in a lattice. For each $\ell = 0, \ldots, L$, Z_ℓ is a lattice random variable with support set $\mathcal{L}_\ell = \{b_\ell + h_\ell m : m \in \mathbb{Z}\}$ of span h_ℓ and offset $b_\ell \in [0, h_\ell)$, if Z_ℓ has strictly zero probability mass outside $\mathcal{L}_\ell$, and h_ℓ is the largest value such that this condition is satisfied. As a consequence, $\boldsymbol{Z}$ lies in an $(L+1)$-dimensional lattice. Let $\boldsymbol{m} = (m_0, \ldots, m_L) \in \mathbb{Z}^{L+1}$. We denote by $p(\boldsymbol{z}_{\boldsymbol{m}})$ the probability mass of the lattice point $\boldsymbol{z}_{\boldsymbol{m}}$, i.e., the point $\boldsymbol{z} = (z_0, \ldots, z_L)$ where $z_\ell = b_\ell + h_\ell m_\ell$ for every $\ell = 0, \ldots, L$. Using the definition of the cost-constraint sets (2) and the error event set (15), equation (14) reads

$$\text{pep}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{\mu_n} \sum_{m_0 \in \mathcal{M}_0} \cdots \sum_{m_L \in \mathcal{M}_L} p(\boldsymbol{z}_{\boldsymbol{m}}) \tag{18}$$

where the summation sets $\mathcal{M}_0, \ldots, \mathcal{M}_L$ are given by

$$\mathcal{M}_0 = \{m \in \mathbb{Z} : b_0 + h_0 m \geq 0\} \tag{19}$$

$$\mathcal{M}_\ell = \{m \in \mathbb{Z} : |b_\ell + h_\ell m| \leq a_\ell\} \tag{20}$$

for $\ell = 1, \ldots, L$. The strongly lattice random variable $\boldsymbol{Z}$ has a periodic characteristic function [15], given by

$$\varphi(\boldsymbol{t}) = \sum_{m_0} \cdots \sum_{m_L} p(\boldsymbol{z}_{\boldsymbol{m}}) e^{j \boldsymbol{t}^T \boldsymbol{z}_{\boldsymbol{m}}} \tag{21}$$

where $j = \sqrt{-1}$, and for convenience, we defined the column vector $\boldsymbol{t} = (t_0, \ldots, t_L)$. We note that in (21), the summations are over all integers. By Fourier inversion [16], we have

$$p(\boldsymbol{z}_{\boldsymbol{m}}) = \frac{\Omega}{(2\pi j)^{L+1}} \int_{-\frac{\pi}{h_0}}^{+\frac{\pi}{h_0}} \mathrm{d}t_0 \cdots \int_{-\frac{\pi}{h_L}}^{+\frac{\pi}{h_L}} \mathrm{d}t_L \varphi(\boldsymbol{t})\, e^{-j \boldsymbol{t}^T \boldsymbol{z}_{\boldsymbol{m}}}, \tag{22}$$

where Ω is the volume of the lattice fundamental cell in the Fourier domain. For sake of simplicity we assume that the fundamental cell is the hypercube of hypervolume $\Omega = \prod_{\ell=0}^{L} h_\ell$.

Making the change of variable $\boldsymbol{s} = j\boldsymbol{t}$, the equation (22) can be written in terms of the inverse Laplace transformation [17]

$$p(\boldsymbol{z_m}) = \frac{\Omega}{(2\pi)^{L+1}} \int_{-j\frac{\pi}{h_0}}^{+j\frac{\pi}{h_0}} \mathrm{d}s_0 \cdots \int_{-j\frac{\pi}{h_L}}^{+j\frac{\pi}{h_L}} \mathrm{d}s_L \, e^{\kappa(\boldsymbol{s})-\boldsymbol{s}^T\boldsymbol{z_m}} \tag{23}$$

where $\kappa(\boldsymbol{s})$ is the cumulant generating function of the vector random variable $\boldsymbol{Z}$, defined as

$$\kappa(\boldsymbol{s}) = \log \sum_{m_0} \cdots \sum_{m_L} p(\boldsymbol{z_m}) e^{\boldsymbol{s}^T\boldsymbol{z_m}}. \tag{24}$$

Under the convergence conditions, the Cauchy's integral theorem [18] allows us to move the integration paths in (23) to the imaginary lines centered at a new point $\hat{\boldsymbol{s}} \in \mathbb{R}^{L+1}$. We now perform a Taylor expansion of $\kappa(\boldsymbol{s})$ around $\boldsymbol{s} = \hat{\boldsymbol{s}}$, i.e.,

$$\kappa(\boldsymbol{s}) \simeq \kappa(\hat{\boldsymbol{s}}) + (\boldsymbol{s}-\hat{\boldsymbol{s}})^T \boldsymbol{\kappa}'(\hat{\boldsymbol{s}}) + \frac{1}{2}(\boldsymbol{s}-\hat{\boldsymbol{s}})^T \boldsymbol{\kappa}''(\hat{\boldsymbol{s}})(\boldsymbol{s}-\hat{\boldsymbol{s}}), \tag{25}$$

where $\boldsymbol{\kappa}'(\boldsymbol{s})$ and $\boldsymbol{\kappa}''(\boldsymbol{s})$ are the gradient and the Hessian matrix of $\kappa(\boldsymbol{s})$, and extend the integration limits to $(\hat{s}_\ell - j\infty, \hat{s}_\ell + j\infty)$ for every ℓ. With these new integration limits, we further make the change of variable $\hat{\boldsymbol{s}} + j\boldsymbol{u} = \boldsymbol{s}$ and combine the expansion (25) with the inversion formula (23) to obtain that

$$p(\boldsymbol{z_m}) \simeq e^{\kappa(\hat{\boldsymbol{s}})-\hat{\boldsymbol{s}}^T\boldsymbol{z_m}} \frac{\Omega}{(2\pi)^{L+1}} \cdot \\ \cdot \int_{-\infty}^{+\infty} \mathrm{d}u_0 \cdots \int_{-\infty}^{+\infty} \mathrm{d}u_L \, e^{-j\boldsymbol{u}^T\boldsymbol{z_m}} \psi(\boldsymbol{u}) \tag{26}$$

where $\psi(\boldsymbol{u})$ is the characteristic function of an $(L+1)$–dimensional Gaussian random variable with vector mean $\boldsymbol{\kappa}'(\hat{\boldsymbol{s}})$ and covariance matrix $\boldsymbol{\kappa}''(\hat{\boldsymbol{s}})$.

The r.h.s. of (26) is only an asymptotic approximation of the probability mass $p(\boldsymbol{z_m})$ due to the Taylor expansion (25) and the extension of the integration limits to $\pm j\infty$. Solving (26), we obtain a saddlepoint approximation of the probability mass of the $(L+1)$–dimensional strongly lattice random variable

$$p(\boldsymbol{z_m}) \simeq \Omega \cdot e^{\kappa(\hat{\boldsymbol{s}})-\hat{\boldsymbol{s}}^T\boldsymbol{z_m}} \cdot \frac{e^{-\frac{1}{2}\boldsymbol{z}_{\boldsymbol{m}}^T \boldsymbol{\kappa}''(\hat{\boldsymbol{s}})^{-1}\boldsymbol{z_m}}}{\sqrt{(2\pi)^{(L+1)} \det \boldsymbol{\kappa}''(\hat{\boldsymbol{s}})}}. \tag{27}$$

The approximation (27) is a generalization of the saddlepoint approximation [16, Eq. (2.2.3)] to the sum of non i.i.d. terms. Contrary to our previous works [11]–[13] where $\hat{\boldsymbol{s}}$ was chosen constant for every channel input and channel output pair, here we considered only the values of $\hat{\boldsymbol{s}}$ such that $\boldsymbol{\kappa}'(\hat{\boldsymbol{s}}) = \boldsymbol{0}$, the unique saddle point that minimizes the cumulant generating function. When $\boldsymbol{Z}$ is the a sequence of i.i.d. terms, the choice of $\hat{\boldsymbol{s}}$ leads to approximation error terms of order $\mathcal{O}(n^{-1})$ that hold uniformly for $\hat{\boldsymbol{s}}$ within the region of convergence of (24).

Taking into account that $\boldsymbol{Z}$ is given by (16) and (17), we can write the cumulant generating function (24) as

$$\kappa(\boldsymbol{s}) = \log \sum_{\overline{\boldsymbol{x}}} Q^n(\overline{\boldsymbol{x}}) \left(\frac{W^n(\boldsymbol{y}|\overline{\boldsymbol{x}})}{W^n(\boldsymbol{y}|\boldsymbol{x})} \right)^{s_0} e^{s_1 c_1(\overline{\boldsymbol{x}})} \cdots e^{s_L c_L(\overline{\boldsymbol{x}})}, \tag{28}$$

giving rise to a linear growth with n of $\kappa(\boldsymbol{s})$ and $\kappa''(\boldsymbol{s})$ by extension. Hence, it is safe to further approximate the r.h.s. of (27) as $n \to \infty$ by

$$p(\boldsymbol{z_m}) \simeq \Omega \cdot \frac{e^{\kappa(\hat{\boldsymbol{s}})-\hat{\boldsymbol{s}}^T\boldsymbol{z_m}}}{\sqrt{(2\pi)^{(L+1)} \det \boldsymbol{\kappa}''(\hat{\boldsymbol{s}})}}. \tag{29}$$

Finally using the approximation (29) in (18) and bringing back the explicit dependence on $\boldsymbol{x}$ and $\boldsymbol{y}$, we obtain that

$$\mathrm{pep}(\boldsymbol{x}, \boldsymbol{y}) \simeq \gamma_{\hat{\boldsymbol{s}}}(\boldsymbol{x}, \boldsymbol{y}) \cdot e^{-i_{\hat{\boldsymbol{s}}}(\boldsymbol{x};\boldsymbol{y})}, \tag{30}$$

where $i_{\hat{\boldsymbol{s}}}(\boldsymbol{x};\boldsymbol{y})$ is the information density given by (12), the term $\gamma_{\hat{\boldsymbol{s}}}(\boldsymbol{x}, \boldsymbol{y})$ is a pre-exponential factor given by

$$\gamma_{\hat{\boldsymbol{s}}}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\Omega}{\mu_n} \cdot \frac{1}{\sqrt{(2\pi)^{(L+1)} \det \boldsymbol{\kappa}''(\hat{\boldsymbol{s}})}} \cdot \\ \cdot \sum_{m_0 \in \mathcal{M}_0} \cdots \sum_{m_L \in \mathcal{M}_L} e^{-\hat{\boldsymbol{s}}^T\boldsymbol{z_m}}, \tag{31}$$

and $\boldsymbol{z_m}$ are the lattice points of the random variable $\boldsymbol{Z}$ defined in (16) and (17). We remark that the terms $\gamma_{\hat{\boldsymbol{s}}}(\boldsymbol{x}, \boldsymbol{y})$ and $i_{\hat{\boldsymbol{s}}}(\boldsymbol{x};\boldsymbol{y})$ depend on $\boldsymbol{x}$ and $\boldsymbol{y}$ in various ways. For both terms, $\hat{\boldsymbol{s}} = \hat{\boldsymbol{s}}(\boldsymbol{x}, \boldsymbol{y})$ is the unique saddle point satisfying $\boldsymbol{\kappa}'(\hat{\boldsymbol{s}}) = 0$. The term $\boldsymbol{\kappa}''(\hat{\boldsymbol{s}})$ in (31) depends on $\boldsymbol{y}$ through the definition of $\kappa(\boldsymbol{s})$ in (28). Lastly, even though the offset b_0 depends on $\boldsymbol{x}$ and $\boldsymbol{y}$, we consider channels such as the binary symmetric channel for which $\mathcal{M}_0$ does not depend on $\boldsymbol{x}$ and $\boldsymbol{y}$.

As shown in [7, Prop. 1], under mild assumptions of the constraint sets (2), the normalizing factor μ_n decays as $n^{-L/2}$. Since the term $\det \boldsymbol{\kappa}''(\hat{\boldsymbol{s}})$ grows as n^{L+1}, the factor $\gamma_{\hat{\boldsymbol{s}}}(\boldsymbol{x}, \boldsymbol{y})$ is sub-exponential with n and asymptotically behaves as $n^{-1/2}$.

We now proceed to study the RCU itself. Let $\boldsymbol{\Theta} = (\Theta_0, \Theta_1, \ldots, \Theta_L)$ be $\Theta_\ell = c_\ell(\boldsymbol{X})$ for each $\ell = 1, \ldots, L$ and

$$\Theta_0 = \log(M-1) + \log \mathrm{pep}(\boldsymbol{X}, \boldsymbol{Y}). \tag{32}$$

We first split the RCU expression (3) into two sums

$$\mathrm{rcu} = \sum_{\boldsymbol{xy}} P(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) \mathbb{1}\{\Theta_0 \geq 0\} + \\ + \sum_{\boldsymbol{xy}} P(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) e^{\Theta_0} \mathbb{1}\{\Theta_0 < 0\}, \tag{33}$$

and use the multiple-cost-constrained distribution (1) to express (33) as (34) given at the bottom of the next page. The cumulant generating function of the underlying vector random variable $\boldsymbol{\Theta}$ is given by

$$\phi(\boldsymbol{\rho}) = \log \sum_{\boldsymbol{xy}} Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) (M-1)^{\rho_0} \mathrm{pep}(\boldsymbol{x}, \boldsymbol{y})^{\rho_0} \cdot \\ \cdot e^{s_1 c_1(\boldsymbol{x})} \cdots e^{s_L c_L(\boldsymbol{x})} \tag{35}$$

for the parameter vector $\boldsymbol{\rho} = (\rho_0, \ldots, \rho_L)$.

Following similar steps as in [12, Sec. III.D], which are omitted for sake of space, we use the saddlepoint approximation of the pairwise error probability (30) in the r.h.s. of (32) to express the cumulant generating function (35) as the sum of two functions, namely

$$\phi(\boldsymbol{\rho}) \simeq \chi_n(\boldsymbol{\rho}) + \pi_n(\boldsymbol{\rho}). \tag{36}$$

In (36), the function $\chi_n(\boldsymbol{\rho})$ accounts for the terms of $\boldsymbol{\Theta}$ that grow linearly with n, and hence contributes to the exponential part of the RCU bound, and is given by

$$\chi_n(\boldsymbol{\rho}) = \rho_0 nR - nE_0(\boldsymbol{\rho}) \tag{37}$$

where we used that $\log(M-1) \simeq nR$ and defined the multiple-cost-constrained Gallager function as (6). We also defined single-letter cost functions $c_\ell(x)$ such that

$$c_\ell(\boldsymbol{x}) = \sum_{i=1}^{n} c_\ell(x_i). \tag{38}$$

In contrast, the sub-linear terms in n that essentially arise from the pairwise error probability sub-exponential term (31) and from the implicit randomness of the pairwise-error-probability saddle point $\hat{\boldsymbol{s}}$ are collected in $\pi_n(\boldsymbol{\rho})$ as

$$\pi_n(\boldsymbol{\rho}) = \rho_0 \log \psi_n(\boldsymbol{\rho}) + \xi_n(\boldsymbol{\rho}). \tag{39}$$

Defining the vector $\boldsymbol{s}(\boldsymbol{\rho})$ as $s_0 = (1+\rho_0)^{-1}$ and $s_\ell = \rho_\ell$ for $\ell = 1, \dots, L$, the quantity $\psi_n(\boldsymbol{\rho})$ accounts for the average contribution of $\gamma_{\hat{\boldsymbol{s}}}(\boldsymbol{x}, \boldsymbol{y})$ into the RCU and is given by

$$\psi_n(\boldsymbol{\rho}) = \frac{\Omega}{\mu_n} \cdot \frac{1}{\sqrt{(2\pi)^{(L+1)}\, \overline{\det \boldsymbol{\kappa}''(\boldsymbol{s}(\boldsymbol{\rho}))}}} \cdot \cdot \sum_{m_0 \in \mathcal{M}_0} \cdots \sum_{m_L \in \mathcal{M}_L} e^{-\boldsymbol{s}(\boldsymbol{\rho})^T \boldsymbol{z}_{\boldsymbol{m}}}, \tag{40}$$

where $\overline{\det \boldsymbol{\kappa}''(\boldsymbol{s}(\boldsymbol{\rho}))}$ is the determinant of the Hessian matrix of the cumulant generating function (28) averaged over the tilted distribution

$$Q^n_{\boldsymbol{\rho}}(\boldsymbol{x}) W^n_{\boldsymbol{\rho}}(\boldsymbol{y}|\boldsymbol{x}) \propto Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) e^{-\rho_0 i_{\boldsymbol{s}(\boldsymbol{\rho})}(\boldsymbol{x},\boldsymbol{y}) + \sum_\ell \rho_\ell c_\ell(\boldsymbol{x})}. \tag{41}$$

On the other hand, the quantity $\xi_n(\boldsymbol{\rho})$ in (39) evaluates the deviation of the saddlepoint exponent $i_{\hat{\boldsymbol{s}}}(\boldsymbol{x}; \boldsymbol{y})$ from the Gallager exponent $i_{\boldsymbol{s}(\boldsymbol{\rho})}(\boldsymbol{x}; \boldsymbol{y})$ as

$$\xi_n(\boldsymbol{\rho}) = \log \sum_{\boldsymbol{x}\boldsymbol{y}} Q^n_{\boldsymbol{\rho}}(\boldsymbol{x}) W^n_{\boldsymbol{\rho}}(\boldsymbol{y}|\boldsymbol{x}) e^{\rho_0 (i_{\boldsymbol{s}(\boldsymbol{\rho})}(\boldsymbol{x};\boldsymbol{y}) - i_{\hat{\boldsymbol{s}}}(\boldsymbol{x};\boldsymbol{y}))}. \tag{42}$$

Equation (36) also implies that the random variable $\boldsymbol{\Theta}$ is exponentially equivalent to the random variable $\boldsymbol{\Phi} = (\Phi_0, \Phi_1, \dots, \Phi_L)$ as $n \to \infty$, where now Φ_ℓ are given by (10)–(11). Clearly, for discrete alphabets, $\boldsymbol{\Phi}$ may lie in a lattice. We define for each $\ell = 0, \dots, L$ the span and offset of Φ_ℓ as g_ℓ and d_ℓ respectively, and use the notation $\boldsymbol{\phi}_{\boldsymbol{m}}$ to represent the lattice point given by $\boldsymbol{m} = (m_0, \dots, m_L)$, i.e., the point $\boldsymbol{\phi}$ whose elements satisfy $\phi_\ell = d_\ell + m_\ell g_\ell$.

We observe in the expression (34) at the bottom of the page that the RCU bound is a probability measure of the random variable Θ_0, while the random variables Θ_ℓ only define the set over which the probability measure is computed. Noting the relation between Θ_0 and Φ_0 given in (32) and (10) respectively, the RCU (34) bound satisfies

$$\text{rcu} = \frac{1}{\mu_n} \sum_{m_0 \in \mathcal{F}_0} \sum_{m_1 \in \mathcal{F}_1} \cdots \sum_{m_L \in \mathcal{F}_L} p(\boldsymbol{\phi}_{\boldsymbol{m}}) + + \frac{\psi_n(\boldsymbol{\rho})}{\mu_n} \sum_{m_0 \in \mathcal{F}_0^c} \sum_{m_1 \in \mathcal{F}_1} \cdots \sum_{m_L \in \mathcal{F}_L} e^{\phi_{m_0}} p(\boldsymbol{\phi}_{\boldsymbol{m}}), \tag{43}$$

where the summation intervals are now given by

$$\mathcal{F}_0 = \{m \in \mathbb{Z} : d_0 + g_0 m + \log \psi_n(\boldsymbol{\rho}) \geq 0\} \tag{44}$$

$$\mathcal{F}_\ell = \{m \in \mathbb{Z} : |d_\ell + g_\ell m| \leq a_\ell\}, \tag{45}$$

$\mathcal{F}_0^c$ is the complement of the set $\mathcal{F}_0$, and with some abuse of notation we employed ϕ_{m_0} to denote the first component of $\boldsymbol{\phi}_{\boldsymbol{m}}$, i.e., the lattice point $d_0 + g_0 m_0$ of the lattice random variable Φ_0 given by (10). The probability mass $p(\boldsymbol{\phi}_{\boldsymbol{m}})$ can be written in terms of the cumulant generating function $\phi(\boldsymbol{\rho})$ using the inverse Laplace transformation, similarly to (23), as

$$p(\phi_{\boldsymbol{m}}) = \frac{\Gamma}{(2\pi)^{L+1}} \int_{-j\frac{\pi}{g_0}}^{+j\frac{\pi}{g_0}} \cdots \int_{-j\frac{\pi}{g_L}}^{+j\frac{\pi}{g_L}} \mathrm{d}\boldsymbol{\rho}\, e^{\xi_n(\boldsymbol{\rho}) + \chi_n(\boldsymbol{\rho}) - \boldsymbol{\rho}^T \boldsymbol{\phi}_{\boldsymbol{m}}} \tag{46}$$

where we used again the relation between the cumulant generating functions of Φ_0 and Θ_0 given in equations (36), (37) and (39), and defined Γ as

$$\Gamma = \prod_{\ell}^{L} g_\ell. \tag{47}$$

At this point, we perform the saddlepoint approximation of (46) by expanding the cumulant generating function $\chi_n(\boldsymbol{\rho})$ around the parameter $\boldsymbol{\rho} = \hat{\boldsymbol{\rho}}$ satisfying the set of conditions (7)–(8), for $\ell = 1, \dots, L$. Hence, plugging

$$\frac{1}{n} \chi_n(\boldsymbol{\rho}) \simeq \hat{\rho}_0 R - E_0^{\text{cc}}(\hat{\boldsymbol{\rho}}) + \frac{1}{2} (\boldsymbol{\rho} - \hat{\boldsymbol{\rho}})^T \boldsymbol{V}_{\hat{\boldsymbol{\rho}}} (\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}) \tag{48}$$

into (46), where $\boldsymbol{V}_{\hat{\boldsymbol{\rho}}}$ is the Hessian matrix of $-E_0^{\text{cc}}(\boldsymbol{\rho})$, solving the complex integration similarly to (26)–(27), and using this approximation on $p(\boldsymbol{\phi}_{\boldsymbol{m}})$ in (43), we finally obtain the approximation (5). Analogously to the discussion after equation (31), the polynomial decays of μ_n and $\det(n\boldsymbol{V}_{\hat{\boldsymbol{\rho}}})$ lead to a polynomial decay of $n^{-1/2}$ in $p(\boldsymbol{\phi}_{\boldsymbol{m}})$. Therefore, with similar arguments to those in [12, Eq. (37)], we obtain that the overall polynomial decay of $\alpha_{\text{cc}}(\hat{\boldsymbol{\rho}})$ is $n^{-(1+\hat{\rho}_0)/2}$.

$$\text{rcu} = \frac{1}{\mu_n} \sum_{\boldsymbol{x}\boldsymbol{y}} Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) \mathbb{1}\{\Theta_0 \geq 0\} \mathbb{1}\{|\Theta_\ell| \leq a_\ell, \forall \ell\} + \frac{1}{\mu_n} \sum_{\boldsymbol{x}\boldsymbol{y}} Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) e^{\Theta_0} \mathbb{1}\{\Theta_0 < 0\} \mathbb{1}\{|\Theta_\ell| \leq a_\ell, \forall \ell\} \tag{34}$$

$$i_{s_0 s_1}(\boldsymbol{x}; \boldsymbol{y}) = s_0 t_{\boldsymbol{x}\boldsymbol{y}} \log \delta + s_0 (n - t_{\boldsymbol{x}\boldsymbol{y}}) \log(1-\delta) - (n - t_{\boldsymbol{y}}) \log\left[\frac{1}{2}(1-\delta)^{s_0} e^{-\frac{s_1}{2}} + \frac{1}{2}\delta^{s_0} e^{\frac{s_1}{2}}\right] - t_{\boldsymbol{y}} \log\left[\frac{1}{2}\delta^{s_0} e^{-\frac{s_1}{2}} + \frac{1}{2}(1-\delta)^{s_0} e^{\frac{s_1}{2}}\right] \tag{48}$$

III. Binary Symmetric Channel

We evaluate the RCU bound for a binary symmetric channel (BSC) with crossover probability $\delta < \frac{1}{2}$. We study both an i.i.d. ensemble with $Q(x) = \frac{1}{2}$ and a constant-composition ensemble with empirical distribution $\hat{Q}(x) = \frac{1}{2}$. In the latter, codewords are uniformly drawn from the type class $\mathcal{T}^n(\hat{Q})$ [1], the set of all sequences of length n with $\hat{Q}(x) = \frac{1}{2}$, i.e, the same number of zeroes and ones.

The constant composition ensemble is a particular case of the multiple cost constrained ensemble (1) for $L = 1$, auxiliary distribution $Q(x) = \frac{1}{2}$, and cost function

$$c_1(\boldsymbol{x}) = \sum_{i=1}^{n}\left(x_i - \frac{1}{2}\right). \tag{49}$$

Clearly, the ensemble (1) with $a_1 = 1$ is uniform over the type class $\mathcal{T}^n(\hat{Q})$ with $\hat{Q}(x) = \frac{1}{2}$.

Computing the saddlepoint approximation (5) requires the evaluation of two important quantities: the Gallager function (6) and the information density (12), respectively given by

$$E_0^{\mathrm{cc}}(\rho_0,\rho_1) = -\log\left[\left(\frac{1}{2}(1-\delta)^{\frac{1}{1+\rho_0}}e^{-\frac{\rho_1}{2}} + \frac{1}{2}\delta^{\frac{1}{1+\rho_0}}e^{\frac{\rho_1}{2}}\right)^{1+\rho_0} + \left(\frac{1}{2}\delta^{\frac{1}{1+\rho_0}}e^{-\frac{\rho_1}{2}} + \frac{1}{2}(1-\delta)^{\frac{1}{1+\rho_0}}e^{\frac{\rho_1}{2}}\right)^{1+\rho_0}\right] \tag{50}$$

and (48) at the bottom of the page. For this channel, we defined $t_{\boldsymbol{y}}$ and $t_{\boldsymbol{xy}}$ as the Hamming weight of $\boldsymbol{y}$ and the Hamming distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively.

We depict the saddlepoint approximation of the RCU bound for both constant-composition $\hat{\mathrm{rcu}}_{\mathrm{cc}}$ (5) and i.i.d. $\hat{\mathrm{rcu}}_{\mathrm{iid}}$ [12, Eq. (35)] ensembles with code rate of $R_{\mathrm{b}} = 0.2$ bits per channel use and crossover probability $\delta = 0.11$ in Fig. 1. We have also included the Gallager bound [14, Eq. (5.6.18)], the metaconverse lower bound [6, Eq. (137)], and the exact expressions of the RCU bounds as reference. We observe that both ensembles lead to the same exponential decay, and that the saddlepoint expression (5) accurately approximates the RCU bound for both ensembles.

Fig. 1 also reveals a gap between the constant-composition and the i.i.d. RCU bounds. We study this loss by means of the ratio $\omega_n = \frac{\mathrm{rcu}_{\mathrm{cc}}}{\mathrm{rcu}_{\mathrm{iid}}}$. Removing the quadratic terms in $\alpha_{\mathrm{cc}}(\hat{\boldsymbol{\rho}})$ analogously to the steps in [12, Eq. (37)] and noting that $\mu_n \simeq \sqrt{2/(\pi n)}$, one can show after some mathematical manipulations that the ratio of saddlepoint approximations $\hat{\omega}_n = \hat{\mathrm{rcu}}_{\mathrm{cc}}/\hat{\mathrm{rcu}}_{\mathrm{iid}}$ is asymptotically equivalent to

$$\hat{\omega}_n \asymp \frac{\sqrt{1+\hat{\rho}_0}}{2}\left(\sqrt{\frac{n^2\overline{\kappa}''_{\mathrm{iid}}(\hat{s}_0)}{\det\boldsymbol{\kappa}''(\hat{s}_0,\hat{s}_1)}}\frac{1-e^{-g_0}}{1-e^{-2g_0}}\right)^{\hat{\rho}_0}\sqrt{\frac{V_{\mathrm{iid}}(\hat{\rho}_0)}{|\det(\boldsymbol{V}_{\hat{\boldsymbol{\rho}}})|}}, \tag{52}$$

where $\overline{\kappa}''_{\mathrm{iid}}(\hat{s}_0)$ and $V_{\mathrm{iid}}(\hat{\rho}_0)$ are i.i.d. terms respectively given by [12, Eq. (29)] and [12, Eq. (31)], and g_0 is the span of the tilted information density (12). For the example in Fig. 1, constant-composition codes with empirical distribution $\frac{1}{2}$ converge to an error probability that is around a factor 1.45793 times that of equiprobable i.i.d. codes.

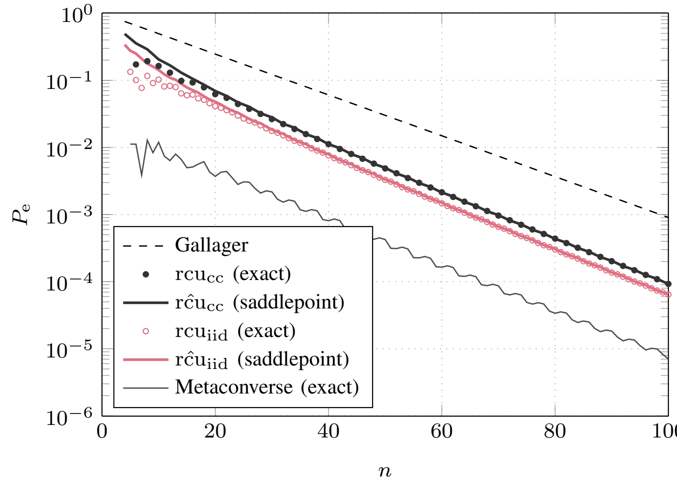

Fig. 1. Error probability vs code length n for code rate $R_{\mathrm{b}} = 0.2$ bits per channel use and BSC with crossover probability $\delta = 0.11$.

References

[1] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[2] R. Fano, *Transmission of Information: A Statistical Theory of Communications*. M.I.T. Press, 1961.

[3] A. Ganti, A. Lapidoth, and I. E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[4] S. Shamai and I. Sason, "Variations on the Gallager bounds, connections, and applications," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3029–3051, Dec. 2002.

[5] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Refinements of the third-order term in the fixed error asymptotics of constant-composition codes," in *IEEE Int. Symp. on Inf. Theory (ISIT)*, Hong Kong, 2015.

[6] 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.

[7] 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.

[8] Y. Altuğ and A. B. Wagner, "Refinement of the random coding bound," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6005–6023, 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] J. Honda, "Exact asymptotics for the random coding error probability," in *IEEE Int. Symp. on Inf. Theory (ISIT)*, Hong Kong, 2015, pp. 91–95.

[11] J. Font-Segura, G. Vazquez-Vilar, A. Martinez, A. Guillén i Fàbregas, and A. Lancho, "Saddlepoint approximations of lower and upper bounds to the error probability in channel coding," in *Conf. on Inf. Sys. and Sciences (CISS)*, Princeton, NJ, Mar. 2018.

[12] J. Font-Segura, A. Martinez, and A. Guillén i Fàbregas, "Asymptotics of the random coding bound," in *Int. Symp. on Inf. Theory and Its Applications (ISITA)*, Singapore, Oct. 2018.

[13] ——, "Saddlepoint approximation of the cost-constrained random coding error probability," in *IEEE Inf. Theory Workshop (ITW)*, Guangzhou, China, Nov. 2018.

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

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

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

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

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