Saddlepoint Approximation of the Cost-Constrained Random Coding Error Probability

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 pairwise error probability and to the random coding union bound are derived for the cost-constrained random coding ensemble. For the special case of the AWGN channel, an alternative expression to approximate the Shannon bound for optimal spherical codes is found.**

I. Introduction

Conceived by Shannon [1], the idea of random coding has been one of the main proof techniques in information theory. By generating a code ensemble whose codewords are i.i.d. distributed, the average performance of such ensemble guarantees the existence of a code with vanishing error probability as long as the code rate is smaller than the mutual information.

Many applications require codewords to satisfy a cost constraint, such as a maximum transmitted power [2, Ch. 7]. Under cost-constrained random coding, codewords are generated according to a cost conditioned distribution. This is similar to the idea of constant-composition codes where all codewords have the same empirical distribution [3]. Both cost-constrained and constant-composition ensembles may lead to performance gains over the i.i.d. ensemble [4]–[6].

In this work, we study saddlepoint approximations to the cost-constrained random coding error probability. For a fixed coding rate R below the channel capacity, the cost-constrained random coding error exponent $E(R)$ provides a first approximation of the error probability as $e^{-nE(R)}$ [2], where n is the code length. Saddlepoint approximations [7] aim at finding a more refined approximation of the form $\alpha_n e^{-nE(R)}$, where α_n is a subexponential factor. Such characterization has been recently addressed by [8]–[10] for the unconstrained case.

For the cost-constrained ensemble, we derive an estimate of α_n valid under some non-lattice conditions, and for rates above the critical rate. We later consider the particular case of shell input AWGN channel, and revisit the Shannon bounds on the performance of optimal spherical codes [11].

II. Cost-Constrained Random Coding

We consider a random coding ensemble in which codewords are drawn from a cost-constrained distribution $P^n(\boldsymbol{x})$ given by

$$P^n(\boldsymbol{x}) = \frac{1}{\mu_n}\prod_{i=1}^{n} Q(x_i)\mathbb{1}\left\{\boldsymbol{x} \in \mathcal{D}_n\right\}, \tag{1}$$

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.

where μ_n is a normalizing factor, $Q(x)$ is a given distribution, $\mathbb{1}\{\cdot\}$ is the indicator function, and $\mathcal{D}_n$ is a cost-constraint set

$$\mathcal{D}_n = \left\{\boldsymbol{x}\ :\ \delta_1 \le c^n(\boldsymbol{x}) \le \delta_2\right\}. \tag{2}$$

In (2), $c^n(\boldsymbol{x}) = \sum_{i=1}^{n} c(x_i)$ is a cost function satisfying $\mathbb{E}_{Q^n}[c^n(\boldsymbol{X})] = 0$, and we assume that δ_1 and δ_2 are constants independent of n. Since (1) is a probability distribution, it follows that $\mu_n = \mathbb{P}_{Q^n}[\boldsymbol{X} \in \mathcal{D}_n]$.

For every message m, equiprobably distributed over the set $\{1, \ldots, M\}$, a codeword $\boldsymbol{x}_m$ is independently generated according to (1). A codeword $\boldsymbol{x}_m$ is transmitted through a memoryless channel characterized by the transition probability $W^n(\boldsymbol{y}|\boldsymbol{x}_m)$. For a given codebook, let $\epsilon_n(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M)$ be the probability of having an error with maximum likelihood (ML) decoding. Then, random coding arguments prove the existence of a code with a vanishing error probability as good as, at least, the average error probability over the ensemble, denoted as $\epsilon_n(M) = \mathbb{E}_{P^n}[\epsilon_n(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M)]$.

III. Saddlepoint Approximations

Using the union bound, the random coding error probability $\epsilon_n(M)$ is upper bounded [12] by $\epsilon_n(M) \le \mathrm{rcu}(M)$, where the random coding union (RCU) bound is given by

$$\mathrm{rcu}(M) = \mathbb{E}_{P^n W^n}\big[\min\left\{1, (M-1)\mathrm{pep}(\boldsymbol{X}, \boldsymbol{Y})\right\}\big], \tag{3}$$

and the pairwise error probability $\mathrm{pep}(\boldsymbol{x}, \boldsymbol{y})$ is the probability that an independently generated codeword $\overline{\boldsymbol{x}}$ has a larger decoding metric than the transmitted codeword $\boldsymbol{x}$, i.e.,

$$\mathrm{pep}(\boldsymbol{x}, \boldsymbol{y}) = \mathbb{P}_{P^n}\big[W^n(\boldsymbol{y}|\overline{\boldsymbol{X}}) \ge W^n(\boldsymbol{y}|\boldsymbol{x})\big]. \tag{4}$$

We first provide a saddlepoint approximation to the pairwise error probability (4), and then use this approximation to find a second saddlepoint approximation to the RCU bound (3).

A. Pairwise Error Probability

The pairwise error probability (4) is the probability of the error event $\mathcal{E}_n(\boldsymbol{x}, \boldsymbol{y})$ given by

$$\mathcal{E}_n(\boldsymbol{x}, \boldsymbol{y}) = \left\{\overline{\boldsymbol{x}}\ : \log W^n(\boldsymbol{y}|\overline{\boldsymbol{x}}) \ge \log W^n(\boldsymbol{y}|\boldsymbol{x})\right\}, \tag{5}$$

for fixed $\boldsymbol{x}$ and $\boldsymbol{y}$. Then,

$$\mathrm{pep}(\boldsymbol{x}, \boldsymbol{y}) = \mathbb{E}_{P^n}\left[\mathbb{1}\left\{\overline{\boldsymbol{X}} \in \mathcal{E}_n(\boldsymbol{x}, \boldsymbol{y})\right\}\right] \tag{6}$$

$$= \frac{1}{\mu_n}\mathbb{E}_{Q^n}\left[\mathbb{1}\left\{\overline{\boldsymbol{X}} \in \mathcal{D}_n \cap \mathcal{E}_n(\boldsymbol{x}, \boldsymbol{y})\right\}\right], \tag{7}$$

where in (7) we have used the cost-constrained distribution (1)–(2). We define the random variables $Z(\boldsymbol{x},\boldsymbol{y}) = \log W^n(\boldsymbol{y}|\overline{\boldsymbol{X}}) - \log W^n(\boldsymbol{y}|\boldsymbol{x})$ and $V = c^n(\overline{\boldsymbol{X}})$. For a strongly non-lattice two-dimensional random variable $(Z(\boldsymbol{x},\boldsymbol{y}), V)$, and using the definitions of $\mathcal{D}_n$ and $\mathcal{E}_n(\boldsymbol{x},\boldsymbol{y})$, equation (7) can be written as

$$\mathrm{pep}(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{\mu_n}\int_0^\infty \mathrm{d}z \int_{\delta_1}^{\delta_2} \mathrm{d}v\, p(\boldsymbol{z}), \tag{8}$$

where we have defined the column vector $\boldsymbol{z} = (z, v)^T$ for convenience, and $p(\boldsymbol{z})$ is the jowiint probability density function of $Z(\boldsymbol{x},\boldsymbol{y}), V$. Using the inverse Laplace transformation [13], we may write $p(\boldsymbol{z})$ as

$$p(\boldsymbol{z}) = \left(\frac{1}{2\pi j}\right)^2 \int_{\hat{\tau}-j\infty}^{\hat{\tau}+j\infty} \mathrm{d}\tau \int_{\hat{\omega}-j\infty}^{\hat{\omega}+j\infty} \mathrm{d}\omega\, e^{\kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y}) - \boldsymbol{\tau}^T \boldsymbol{z}}, \tag{9}$$

where $\kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y})$ is the joint cumulant generating function of $Z(\boldsymbol{x},\boldsymbol{y}), V$, given by $\kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y}) = \log \mathbb{E}_{Q^n}\left[e^{\tau Z(\boldsymbol{x},\boldsymbol{y}) + \omega V}\right]$, and where $\boldsymbol{\tau} = (\tau, \omega)^T$. For our particular $Z(\boldsymbol{x},\boldsymbol{y})$ and V, $\kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y})$ is given by

$$\kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y}) = \log \mathbb{E}_{Q^n}\left[\left(\frac{W^n(\boldsymbol{y}|\overline{\boldsymbol{X}})}{W^n(\boldsymbol{y}|\boldsymbol{x})}\right)^\tau \cdot e^{\omega c^n(\overline{\boldsymbol{X}})}\right]. \tag{10}$$

We assume that $\hat{\boldsymbol{\tau}} = (\hat{\tau}, \hat{\omega})$ is within the region of convergence of the complex integration (9). Now, we perform a Taylor expansion of $\kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y})$ around $\hat{\boldsymbol{\tau}}$, i.e.,

$$\begin{aligned}\kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y}) \simeq \kappa_{\hat{\tau}\hat{\omega}}(\boldsymbol{x},\boldsymbol{y}) + (\boldsymbol{\tau} - \hat{\boldsymbol{\tau}})^T \boldsymbol{\kappa}'_{\hat{\tau}\hat{\omega}}(\boldsymbol{x},\boldsymbol{y}) + \\ + \frac{1}{2}(\boldsymbol{\tau} - \hat{\boldsymbol{\tau}})^T \boldsymbol{\kappa}''_{\hat{\tau}\hat{\omega}}(\boldsymbol{y})(\boldsymbol{\tau} - \hat{\boldsymbol{\tau}}),\end{aligned} \tag{11}$$

where $\boldsymbol{\kappa}'_{\tau\omega}(\boldsymbol{x},\boldsymbol{y})$ and $\boldsymbol{\kappa}''_{\tau\omega}(\boldsymbol{y})$ are the gradient and the Hessian matrix of $\kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y})$, respectively given by

$$\boldsymbol{\kappa}'_{\tau\omega}(\boldsymbol{x},\boldsymbol{y}) = \begin{bmatrix} \frac{\partial}{\partial \tau} \\ \frac{\partial}{\partial \omega} \end{bmatrix} \kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y}), \tag{12}$$

$$\boldsymbol{\kappa}''_{\tau\omega}(\boldsymbol{y}) = \begin{bmatrix} \frac{\partial^2}{\partial \tau^2} & \frac{\partial^2}{\partial \tau \partial \omega} \\ \frac{\partial^2}{\partial \omega \partial \tau} & \frac{\partial^2}{\partial \tau^2} \end{bmatrix} \kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y}). \tag{13}$$

We note that $\boldsymbol{\kappa}''_{\tau\omega}(\boldsymbol{y})$ does not depend on $\boldsymbol{x}$, as the term $W^n(\boldsymbol{y}|\boldsymbol{x})$ in (10) is linear with τ. Plugging (11) into (9) and making the change of variables $\hat{\tau} + j\tau_i = \tau$ and $\hat{\omega} + j\omega_i = \omega$, we obtain that the probability density function $p(\boldsymbol{z})$ is approximated as

$$\begin{aligned}p(\boldsymbol{z}) \simeq e^{\kappa_{\hat{\tau}\hat{\omega}}(\boldsymbol{x},\boldsymbol{y}) - \hat{\boldsymbol{\tau}}^T \boldsymbol{z}} \cdot \\ \cdot \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty} \mathrm{d}\tau_i \int_{-\infty}^{\infty} \mathrm{d}\omega_i\, e^{-j\boldsymbol{\tau}_i^T \boldsymbol{z}} \varphi(\boldsymbol{\tau}_i),\end{aligned} \tag{14}$$

where $\varphi(\boldsymbol{\tau}_i)$ is the characteristic function of a bidimensional normal distribution with mean $\boldsymbol{\kappa}'_{\hat{\tau}\hat{\omega}}(\boldsymbol{x},\boldsymbol{y})$ and covariance matrix $\boldsymbol{\kappa}''_{\hat{\tau}\hat{\omega}}(\boldsymbol{y})$, i.e.,

$$\varphi(\boldsymbol{\tau}_i) = e^{j\boldsymbol{\tau}_i^T \boldsymbol{\kappa}'_{\hat{\tau}\hat{\omega}}(\boldsymbol{x},\boldsymbol{y}) - \frac{1}{2}\boldsymbol{\tau}_i^T \boldsymbol{\kappa}''_{\hat{\tau}\hat{\omega}}(\boldsymbol{y})\boldsymbol{\tau}_i}. \tag{15}$$

Hence, since $\varphi(\boldsymbol{\tau}_i)$ is integrable in $\mathbb{R}^2$, solving the integration (14) leads to the saddlepoint approximation [14] of $p(\boldsymbol{z})$, i.e.,

$$\begin{aligned}p(\boldsymbol{z}) \simeq e^{\kappa_{\hat{\tau}\hat{\omega}}(\boldsymbol{x},\boldsymbol{y}) - \hat{\boldsymbol{\tau}}^T \boldsymbol{z}} \cdot \\ \cdot \frac{e^{-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\kappa}'_{\hat{\tau}\hat{\omega}}(\boldsymbol{x},\boldsymbol{y}))^T (\boldsymbol{\kappa}''_{\hat{\tau}\hat{\omega}}(\boldsymbol{y}))^{-1} (\boldsymbol{z} - \boldsymbol{\kappa}'_{\hat{\tau}\hat{\omega}}(\boldsymbol{x},\boldsymbol{y}))}}{\sqrt{(2\pi)^2 |\boldsymbol{\kappa}''_{\hat{\tau}\hat{\omega}}(\boldsymbol{y})|}},\end{aligned} \tag{16}$$

Since $\boldsymbol{\kappa}'_{\tau\omega}(\boldsymbol{x},\boldsymbol{y})$ and $\boldsymbol{\kappa}''_{\tau\omega}(\boldsymbol{y})$, grow linearly with n, for sufficiently large n we may neglect the terms in $\boldsymbol{z}$ in the quadratic form in (16), i.e.,

$$p(\boldsymbol{z}) \simeq \frac{e^{\kappa_{\hat{\tau}\hat{\omega}}(\boldsymbol{x},\boldsymbol{y}) - \hat{\boldsymbol{\tau}}^T \boldsymbol{z} - \frac{1}{2}\boldsymbol{\kappa}'_{\hat{\tau}\hat{\omega}}(\boldsymbol{x},\boldsymbol{y})^T (\boldsymbol{\kappa}''_{\hat{\tau}\hat{\omega}}(\boldsymbol{y}))^{-1} \boldsymbol{\kappa}'_{\hat{\tau}\hat{\omega}}(\boldsymbol{x},\boldsymbol{y})}}{\sqrt{(2\pi)^2 |\boldsymbol{\kappa}''_{\hat{\tau}\hat{\omega}}(\boldsymbol{y})|}}. \tag{17}$$

Using the approximation (17) into (8), we obtain

$$\begin{aligned}\mathrm{pep}(\boldsymbol{x},\boldsymbol{y}) \simeq \int_0^\infty \mathrm{d}z \int_{\delta_1}^{\delta_2} \mathrm{d}v \frac{e^{-\hat{\boldsymbol{\tau}}^T \boldsymbol{z}}}{\mu_n} \cdot \\ \cdot \frac{e^{\kappa_{\hat{\tau}\hat{\omega}}(\boldsymbol{x},\boldsymbol{y}) - \frac{1}{2}\boldsymbol{\kappa}'_{\hat{\tau}\hat{\omega}}(\boldsymbol{x},\boldsymbol{y})^T (\boldsymbol{\kappa}''_{\hat{\tau}\hat{\omega}}(\boldsymbol{y}))^{-1} \boldsymbol{\kappa}'_{\hat{\tau}\hat{\omega}}(\boldsymbol{x},\boldsymbol{y})}}{\sqrt{(2\pi)^2 |\boldsymbol{\kappa}''_{\hat{\tau}\hat{\omega}}(\boldsymbol{y})|}}.\end{aligned} \tag{18}$$

Solving the integration w.r.t. $\boldsymbol{z}$, we obtain that for a given channel input $\boldsymbol{x}$ and channel output $\boldsymbol{y}$, the pairwise error probability (4) under the random coding ensemble (1) can be approximated by

$$\mathrm{pep}(\boldsymbol{x},\boldsymbol{y}) \simeq \gamma_n(\boldsymbol{x},\boldsymbol{y}) \cdot e^{\kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y})}, \tag{19}$$

where we redefine $\omega = \hat{\omega}$ and $\tau = \hat{\tau}$ for notation clarity, $\kappa_{\tau\omega}(\boldsymbol{x};\boldsymbol{y})$ is the cumulant generating function (10), and $\gamma_n(\boldsymbol{y})$ is a subexponential related to (12) and (13) as

$$\gamma_n(\boldsymbol{x},\boldsymbol{y}) = \frac{e^{-\omega\delta_1} - e^{-\omega\delta_2}}{\mu_n \tau\omega} \cdot \frac{e^{-\frac{1}{2}\boldsymbol{\kappa}'_{\tau\omega}(\boldsymbol{x},\boldsymbol{y})^T \boldsymbol{\kappa}''_{\tau\omega}(\boldsymbol{y})^{-1} \boldsymbol{\kappa}'_{\tau\omega}(\boldsymbol{x},\boldsymbol{y})}}{\sqrt{(2\pi)^2 |\boldsymbol{\kappa}''_{\tau\omega}(\boldsymbol{y})|}}. \tag{20}$$

We remark that (10) involves the expectation according to the i.i.d. distribution $Q^n(\overline{x}) = \prod_{i=1}^n Q(\overline{x})$. We also note that the optimal auxiliary parameters τ and ω would be chosen as the unique minimizers of $\kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y})$, which would set $\boldsymbol{\kappa}'_{\tau\omega}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{0}$ in the Taylor expansion (11) and in (20). However, this requires one optimization for every $\boldsymbol{x}$ and $\boldsymbol{y}$. Instead, we let τ and ω be fixed for every $\boldsymbol{x}$ and $\boldsymbol{y}$, at the cost of having nonzero $\boldsymbol{\kappa}'_{\tau\omega}(\boldsymbol{x},\boldsymbol{y})$ in $\gamma_n(\boldsymbol{x},\boldsymbol{y})$.

As reported in [2], [5] for the power-constrained AWGN channel, μ_n decays subexponentially in n, hence not affecting the exponent. The saddlepoint approximation (19) states that the pairwise error probability, under cost-constrained i.i.d. random coding ensemble (1), decays exponentially as $e^{\kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y})}$, with a pre-factor $\gamma_n(\boldsymbol{x},\boldsymbol{y})$. We note that $\kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y}) = -i^n_{\tau\omega}(\boldsymbol{x};\boldsymbol{y})$ is the negative tilted information density

$$i^n_{\tau\omega}(\boldsymbol{x};\boldsymbol{y}) = \log \frac{W^n(\boldsymbol{y}|\boldsymbol{x})^\tau}{\mathbb{E}_{Q^n}\left[W^n(\boldsymbol{y}|\overline{\boldsymbol{X}})^\tau e^{\omega c^n(\overline{\boldsymbol{X}})}\right]}. \tag{21}$$

B. Random Coding Union Bound

We start by using the identity $\mathbb{E}[\min\{1, A\}] = \mathbb{P}[A \geq U]$, where U is a uniformly distributed random variable in the $[0,1]$ interval, to write the $\mathrm{rcu}(M)$ as

$$\mathrm{rcu}(M) = \mathbb{E}_{P^n W^n F}\left[\mathbb{1}\left\{(\boldsymbol{X},\boldsymbol{Y},U) \in \mathcal{R}_n\right\}\right] \tag{22}$$

where $\mathcal{R}_n$ is the set

$$\mathcal{R}_n = \{(\boldsymbol{x}, \boldsymbol{y}, u) \,:\, \log(M-1) + \log \mathrm{pep}(\boldsymbol{x}, \boldsymbol{y}) \geq \log u\}\,, \tag{23}$$

and $F(u)$ is the uniform probability distribution. Using the right hand side of (1), we further have that

$$\mathrm{rcu}(M) = \frac{1}{\mu_n} \mathbb{E}_{Q^n W^n F}\left[\mathbb{1}\left\{(\boldsymbol{X}, \boldsymbol{Y}, U) \in \mathcal{R}_n, \boldsymbol{X} \in \mathcal{D}_n\right\}\right]. \tag{24}$$

Similarly to Sec. III-A, we define the random variables $Z = \log(M-1) + \log \mathrm{pep}(\boldsymbol{X}, \boldsymbol{Y}) - \log U$ and $V = c^n(\boldsymbol{X})$, and use the definitions of $\mathcal{R}_n$ and $\mathcal{D}_n$ to write equation (24) for a strongly non-lattice two-dimensional random variable, i.e.,

$$\mathrm{rcu}(M) = \frac{1}{\mu_n} \int_0^{\infty} \mathrm{d}z \int_{\delta_1}^{\delta_2} \mathrm{d}v \; p(\boldsymbol{z}), \tag{25}$$

where $\boldsymbol{z} = (z, v)^T$, and $p(\boldsymbol{z})$ is the joint probability density function of (Z, V), given by

$$p(\boldsymbol{z}) = \left(\frac{1}{2\pi j}\right)^2 \int_{\hat{\rho}-j\infty}^{\hat{\rho}+j\infty} \mathrm{d}\rho \int_{\hat{\lambda}-j\infty}^{\hat{\lambda}+j\infty} \mathrm{d}\lambda \; e^{\chi(\rho,\lambda) - \boldsymbol{\rho}^T \boldsymbol{z}}. \tag{26}$$

Now, $\boldsymbol{\rho} = (\rho, \lambda)^T$, and $\chi(\rho, \lambda) = \log \mathbb{E}_{Q^n W^n F}[e^{\rho Z + \lambda V}]$ is the joint cumulant generating function of Z, V, i.e.,

$$\chi(\rho, \lambda) = \log \mathbb{E}_{Q^n W^n F}\Big[(M-1)^{\rho} \mathrm{pep}(\boldsymbol{X}, \boldsymbol{Y})^{\rho} U^{-\rho} e^{\lambda c^n(\boldsymbol{X})}\Big]. \tag{27}$$

Plugging the saddlepoint approximation (19) into (27), taking $\log(M-1) \simeq nR$, and after some mathematical manipulations, we obtain that

$$\chi(\rho, \lambda) \simeq n\rho R - \log(1-\rho) + \\ + \log \mathbb{E}_{Q^n W^n}\Big[\Big(\gamma_n(\boldsymbol{X}, \boldsymbol{Y}) e^{\kappa_{\tau\omega}(\boldsymbol{X}, \boldsymbol{Y})}\Big)^{\rho} e^{\lambda c^n(\boldsymbol{X})}\Big]. \tag{28}$$

The saddlepoint approximation of the pairwise error probability was found by directly expanding the cumulant generating function (10), as both Z and V were the sum of n i.i.d. random variables. This is not the case of $\chi(\rho, \lambda)$, as there are terms in (28) that are not linear with n. It will prove convenient to write equation (28) as

$$\chi(\rho, \lambda) \simeq n\rho R + \log \mathbb{E}_{Q^n W^n}\left[e^{\rho \kappa_{\tau\omega}(\boldsymbol{X}, \boldsymbol{Y})} e^{\lambda c^n(\boldsymbol{X})}\right] + \\ + \log \mathbb{E}_{Q^n_{\rho\lambda} W^n_{\rho\lambda}}\left[\gamma_n(\boldsymbol{X}, \boldsymbol{Y})^{\rho}\right] - \log(1-\rho), \tag{29}$$

where $Q^n_{\rho\lambda}(\boldsymbol{x}) W^n_{\rho\lambda}(\boldsymbol{y}|\boldsymbol{x})$ is the tilted distribution

$$Q^n_{\rho\lambda}(\boldsymbol{x}) W^n_{\rho\lambda}(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{\nu_n} Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) e^{\rho \kappa_{\tau\omega}(\boldsymbol{x}, \boldsymbol{y})} e^{\lambda c^n(\boldsymbol{x})}, \tag{30}$$

being ν_n a normalization factor. Since $\log(1-\rho)$ and $\gamma_n(\boldsymbol{x}, \boldsymbol{y})$ do not grow exponentially with n, we obtain that

$$\lim_{n\to\infty} \frac{1}{n} \chi(\rho, \lambda) = \rho R - E_0(\rho, \lambda), \tag{31}$$

where

$$E_0(\rho, \lambda) = \lim_{n\to\infty} -\frac{1}{n} \log \mathbb{E}_{Q^n W^n}\left[e^{\rho \kappa_{\tau\omega}(\boldsymbol{X}, \boldsymbol{Y})} e^{\lambda c^n(\boldsymbol{X})}\right]. \tag{32}$$

For a memoryless channel $W^n(\boldsymbol{y}|\boldsymbol{x})$ and i.i.d. input distribution $Q^n(\boldsymbol{x})$, setting $\tau = \frac{1}{1+\rho}$ and $\omega = \lambda$, turns equation (32) into Gallager's E_0 function for constrained inputs [2, Eq. (7.3.43)] given by

$$E_0(\rho, \lambda) = -\log \sum_y \left(\sum_x Q(x) W(y|x)^{\frac{1}{1+\rho}} e^{\lambda c(x)}\right)^{1+\rho}. \tag{33}$$

Defining

$$\theta_n(\rho, \lambda) = \mathbb{E}_{Q^n_{\rho\lambda} W^n_{\rho\lambda}}\Big[\gamma_n(\boldsymbol{X}, \boldsymbol{Y})^{\rho}\Big], \tag{34}$$

the cumulant generating function (27) is of the form

$$\chi(\rho, \lambda) \simeq n\rho R - nE_0(\rho, \lambda) + \log \theta_n(\rho, \lambda) - \log(1-\rho). \tag{35}$$

Using (35) into (26), and then solving (25), we obtain that

$$\mathrm{rcu}(M) = \left(\frac{1}{2\pi j}\right)^2 \int_{\hat{\rho}-j\infty}^{\hat{\rho}+j\infty} \mathrm{d}\rho \int_{\hat{\lambda}-j\infty}^{\hat{\lambda}+j\infty} \mathrm{d}\lambda \cdot \\ \cdot \frac{e^{-\lambda\delta_1} - e^{-\lambda\delta_2}}{\mu_n \lambda} \cdot \frac{e^{n\rho R - nE_0(\rho,\lambda)}}{\rho(1-\rho)} \theta_n(\rho, \lambda). \tag{36}$$

The saddlepoint approximation to the $\mathrm{rcu}(M)$ then involves expanding $\rho R - E_0(\rho, \lambda)$ around $\hat{\boldsymbol{\rho}} = (\hat{\rho}, \hat{\lambda})$, the unique minimizers such that

$$\frac{\partial}{\partial \rho} E_0(\rho, \lambda)\Big|_{\rho=\hat{\rho}} = R, \tag{37}$$

$$\frac{\partial}{\partial \lambda} E_0(\rho, \lambda)\Big|_{\lambda=\hat{\lambda}} = 0. \tag{38}$$

Therefore, around $\hat{\boldsymbol{\rho}}$, we find the expansion

$$\rho R - E_0(\rho, \lambda) \simeq \hat{\rho} R - E_0(\hat{\rho}, \hat{\lambda}) + \frac{1}{2}(\boldsymbol{\rho} - \hat{\boldsymbol{\rho}})^T \boldsymbol{V}_{\hat{\rho}\hat{\lambda}} (\boldsymbol{\rho} - \hat{\boldsymbol{\rho}}), \tag{39}$$

where $\boldsymbol{V}_{\hat{\rho}\hat{\lambda}}$, the cost-constrained dispersion matrix, is the Hessian matrix of $\rho R - E_0(\rho, \lambda)$, given by

$$\boldsymbol{V}_{\hat{\rho}\hat{\lambda}} = -\begin{bmatrix} \frac{\partial^2}{\partial \rho^2} & \frac{\partial^2}{\partial \rho \partial \lambda} \\ \frac{\partial^2}{\partial \lambda \partial \rho} & \frac{\partial^2}{\partial \lambda^2} \end{bmatrix} E_0(\rho, \lambda)\Bigg|_{\rho=\hat{\rho}, \lambda=\hat{\lambda}}. \tag{40}$$

The convergence of equation (36) highly depends on the poles at $\rho = 0$, $\rho = 1$ and $\lambda = 0$. As discussed in [2, Theorem 7.3.2], the error exponent of the constrained input random coding ensemble is given by

$$E(R, Q) = \inf_{0 \leq \rho \leq 1, \lambda \geq 0} \rho R - E_0(\rho, \lambda). \tag{41}$$

When $0 < \hat{\rho} < 1$ and $\hat{\lambda} > 0$, this corresponds to rates R between the critical rate $R^*(Q)$, defined as the rate for which $E(R, Q)$ is achieved at $\hat{\rho} = 1$, and the mutual information $I(Q)$, given by

$$I(Q) = \mathbb{E}_{Q^n W^n}\left[\log \frac{W^n(\boldsymbol{y}|\boldsymbol{x})}{\mathbb{E}_{Q^n}\left[W^n(\boldsymbol{y}|\overline{\boldsymbol{X}})\right]}\right], \tag{42}$$

for which $\hat{\rho} = \hat{\lambda} = 0$. For this range of $\hat{\rho}$ and $\hat{\lambda}$, the complex integration (26) converges for any $\boldsymbol{z}$, so that we can use the Taylor expansion (39) to approximate (36). Conversely, if the rate R is such that the parameter $\hat{\rho}$ satisfying (37) lies outside the $(0,1)$ interval, we need to shift the integration axis of $\hat{\rho}$ at the cost of introducing additional terms due to the Cauchy's residue theorem [13]. For sake of clarity, we consider the more explanatory case of $\rho \in (0,1)$.

To solve (36), it is convenient to use the identity

$$\frac{1}{\rho(1-\rho)} = \frac{1}{\rho} + \frac{1}{1-\rho}, \tag{43}$$

and note that we can equivalently write (25) as

$$\mathrm{rcu}(M) = \frac{1}{\mu_n}\int_{\delta_1}^{\delta_2} \mathrm{d}v \left(\int_{-\infty}^{0} p_1(\boldsymbol{z})\mathrm{d}z + \int_{0}^{\infty} p_2(\boldsymbol{z})\mathrm{d}z \right), \tag{44}$$

where the probability distributions $p_1(\boldsymbol{z})$ and $p_2(\boldsymbol{z})$ are respectively given by

$$p_1(\boldsymbol{z}) = \left(\frac{1}{2\pi j}\right)^2 \int_{\hat{\rho}-j\infty}^{\hat{\rho}+j\infty} \mathrm{d}\rho \int_{\hat{\lambda}-j\infty}^{\hat{\lambda}+j\infty} \mathrm{d}\lambda \; e^{\zeta(\rho,\lambda)+z-\boldsymbol{\rho}^T\boldsymbol{z}}, \tag{45}$$

$$p_2(\boldsymbol{z}) = \left(\frac{1}{2\pi j}\right)^2 \int_{\hat{\rho}-j\infty}^{\hat{\rho}+j\infty} \mathrm{d}\rho \int_{\hat{\lambda}-j\infty}^{\hat{\lambda}+j\infty} \mathrm{d}\lambda \; e^{\zeta(\rho,\lambda)-\boldsymbol{\rho}^T\boldsymbol{z}}. \tag{46}$$

being $\zeta(\rho,\lambda) \simeq n\rho R - nE_0(\rho,\lambda) + \log\theta_n(\rho,\lambda)$. Using the Taylor expansion (39) into (45) and (46), and following the footsteps of the derivation of equations (14), (15) and (16), we obtain the following saddlepoint approximations [14]

$$p_1(\boldsymbol{z}) \simeq e^{\zeta(\hat{\rho},\hat{\lambda})+z-\hat{\boldsymbol{\rho}}^T\boldsymbol{z}} \cdot \frac{1}{\sqrt{(2\pi)^2|n\boldsymbol{V}_{\hat{\rho}\hat{\lambda}}|}} \cdot e^{-\frac{1}{2n}\boldsymbol{z}^T\boldsymbol{V}_{\hat{\rho}\hat{\lambda}}^{-1}\boldsymbol{z}}, \tag{47}$$

$$p_2(\boldsymbol{z}) \simeq e^{\zeta(\hat{\rho},\hat{\lambda})-\hat{\boldsymbol{\rho}}^T\boldsymbol{z}} \cdot \frac{1}{\sqrt{(2\pi)^2|n\boldsymbol{V}_{\hat{\rho}\hat{\lambda}}|}} \cdot e^{-\frac{1}{2n}\boldsymbol{z}^T\boldsymbol{V}_{\hat{\rho}\hat{\lambda}}^{-1}\boldsymbol{z}}. \tag{48}$$

Using the approximations (47) and (48) into (44) and defining the bidimensional integration intervals $\mathcal{I}_1 = (-\infty,0)\times(\delta_1,\delta_2)$ and $\mathcal{I}_2 = (0,\infty)\times(\delta_1,\delta_2)$, we obtain that the saddlepoint approximation of the random coding union bound to the error probability (3) under the cost-constrained random coding ensemble (1) is given by

$$\mathrm{rcu}(M) \simeq \alpha_n \cdot e^{-n(E_0(\hat{\rho},\hat{\lambda})-\hat{\rho}R)}, \tag{49}$$

where the factor α_n is found as

$$\alpha_n = \frac{\theta_n(\hat{\rho},\hat{\lambda})}{\mu_n}\left(\int_{\mathcal{I}_1} \mathrm{d}\boldsymbol{z}\, \frac{e^{z-\hat{\boldsymbol{\rho}}^T\boldsymbol{z}-\frac{1}{2n}\boldsymbol{z}^T\boldsymbol{V}_{\hat{\rho}\hat{\lambda}}^{-1}\boldsymbol{z}}}{\sqrt{(2\pi)^2|n\boldsymbol{V}_{\hat{\rho}\hat{\lambda}}|}} + \right. \left. + \int_{\mathcal{I}_2} \mathrm{d}\boldsymbol{z}\, \frac{e^{-\hat{\boldsymbol{\rho}}^T\boldsymbol{z}-\frac{1}{2n}\boldsymbol{z}^T\boldsymbol{V}_{\hat{\rho}\hat{\lambda}}^{-1}\boldsymbol{z}}}{\sqrt{(2\pi)^2|n\boldsymbol{V}_{\hat{\rho}\hat{\lambda}}|}} \right), \tag{50}$$

and $(\hat{\rho},\hat{\lambda})$ are the saddlepoints obtained from (37)–(38). In (50), $\boldsymbol{z}$ is a bidimensional integration variable, $\boldsymbol{V}_{\rho\lambda}$ is the cost-constrained dispersion matrix given by (40), and the parameter $\theta_n(\hat{\rho},\hat{\lambda})$ is computed as (34), where $\gamma_n(\boldsymbol{x},\boldsymbol{y})$ is given by (20) with $\omega = \hat{\lambda}$ and $\tau = \frac{1}{1+\hat{\rho}}$ and the expectation is under the tilted distribution $Q^n_{\rho\lambda}(\boldsymbol{x})W^n_{\rho\lambda}(\boldsymbol{y}|\boldsymbol{x})$ given in (30).

Computing α_n involves solving two bidimensional integrations, which can be done using standard numerical integration packages. A simpler expression is obtained by further expanding α_n as $n \to \infty$. Neglecting the quadratic form $\frac{1}{2n}\boldsymbol{z}^T\boldsymbol{V}_{\hat{\rho}\hat{\lambda}}^{-1}\boldsymbol{z}$ in (50) to solve the bidimensional integrations, we obtain that α_n can be asymptotically approximated by

$$\alpha_n \simeq \frac{e^{-\hat{\lambda}\delta_1} - e^{-\hat{\lambda}\delta_2}}{\mu_n\hat{\lambda}\hat{\rho}(1-\hat{\rho})} \cdot \frac{\theta_n(\hat{\rho},\hat{\lambda})}{\sqrt{(2\pi)^2|n\boldsymbol{V}_{\hat{\rho}\hat{\lambda}}|}}. \tag{51}$$

Clearly, for rates R satisfying $R^*(Q) < R < I(Q)$, the saddlepoint approximation (49) recovers the correct exponential decay of the error probability, i.e. (41). Furthermore, since $|\boldsymbol{\kappa}''_{\tau\omega}(\boldsymbol{y})|$ and $|n\boldsymbol{V}_{\rho\lambda}|$ both grow as n^2, and μ_n decays as $\frac{1}{\sqrt{n}}$ [5], it follows that α_n in the expansion (51), and hence also the cost-constrained random coding error probability (49), have a polynomial decay of $n^{-\frac{1+\hat{\rho}}{2}}$, exactly the same as the unconstrained case [9, Eq. (134)].

IV. AWGN Channel with Spherical Codes

We numerically evaluate the saddlepoint approximation (49) for the shell input AWGN channel, where codewords satisfy $\sum_{i=1}^n x_i^2 = nP$, being P the power, and the AWGN channel transition probability is given by

$$W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-x_i)^2}{2\sigma^2}}, \tag{52}$$

being σ^2 the noise power. We define the signal-to-noise ratio (SNR) as $\mathrm{snr} = \frac{P}{\sigma^2}$. Shannon gave an expression for the exact RCU bound based on packing spheral cones [11, Eq. (19)], whose accurate computation is challenging, and an asymptotic approximation for large n, given by [11, Eq. (5)].

In order to map our model into the AWGN channel with spherical coding, we consider the function $c^n(\boldsymbol{x}) = \sum_{i=1}^n x_i^2 - nP$, and the set $\mathcal{D}_n = \{\boldsymbol{x} \ : \ nP - \delta \le \sum_{i=1}^n x_i \le nP\}$, corresponding to $\delta_1 = -\delta$ and $\delta_2 = 0$ in (2). We further consider the distribution

$$Q^n(\boldsymbol{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi P}} e^{-\frac{x^2}{2P}}. \tag{53}$$

Under this setting, $\mu_n \simeq \frac{\delta}{\sqrt{4\pi n P^2}}$ [2, Eq. (7.3.29)].

Particularizing our expressions, and taking the limit as $\delta \to 0$ to constrain our code to be spherical, we obtain that the random coding union (RCU) bound can be approximated as (49) where the Gallager function $E_0(\rho,\lambda)$ is

$$E_0(\rho,\lambda) = \lambda P(1+\rho) + \frac{1}{2}\log(1-2\lambda P) + + \frac{\rho}{2}\log\left(1 - 2\lambda P + \frac{\mathrm{snr}}{1+\rho}\right), \tag{54}$$

and α_n is given by

$$\alpha_n = \theta_n(\hat{\rho},\hat{\lambda})\sqrt{2P^2V_{22}^{-1}} \cdot \left(\mathrm{erfcx}\left((1-\hat{\rho})\sqrt{V_{22}^{-1}n|\boldsymbol{V}_{\hat{\rho}\hat{\lambda}}|}\right) + \mathrm{erfcx}\left(\hat{\rho}\sqrt{V_{22}^{-1}n|\boldsymbol{V}_{\hat{\rho}\hat{\lambda}}|}\right)\right). \tag{55}$$

In (55), $\mathrm{erfcx}(t)$ is a modified Gaussian error function given by $\mathrm{erfcx}(t) = \mathrm{sign}(t)\frac{1}{2}\mathrm{erfc}\big(\frac{|t|}{\sqrt{2}}\big)e^{\frac{t^2}{2}}$, V_{22} is given by

$$V_{22} = -\frac{\partial^2}{\partial\lambda^2}E_0(\rho,\lambda)\bigg|_{(\rho,\lambda)=(\hat{\rho},\hat{\lambda})}, \quad (56)$$

and $\theta_n(\hat{\rho},\hat{\lambda})$ can be computed in closed-form from equations (34), (20) and (30), where $\kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y})$ is given by

$$\kappa_{\tau\omega}(\boldsymbol{x},\boldsymbol{y}) = -\frac{1}{2}\frac{(1-2\omega P)\tau}{(1-2\omega P)\sigma^2 + P\tau}\|\boldsymbol{y}\|^2 - \omega nP + \\ +\frac{\tau}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{x}\|^2 - \frac{n}{2}\log\left(1-2\omega P + \tau\,\mathrm{snr}\right). \quad (57)$$

The expression of $\theta_n(\hat{\rho},\hat{\lambda})$, cumbersome and not particularly informative, is not included for the sake of space limitations.

Either expanding (55) as $n \to \infty$, or evaluating (51) as $\delta \to 0$, we obtain that the saddlepoint approximation asymptotically behaves as (49), where the Gallager function $E_0(\rho,\lambda)$ is given by (54), and now α_n is given by

$$\alpha_n \simeq \frac{\theta_n(\hat{\rho},\hat{\lambda})}{\hat{\rho}(1-\hat{\rho})\sqrt{\pi n P^{-2}|\boldsymbol{V}_{\hat{\rho}\hat{\lambda}}|}}. \quad (58)$$

We have derived two approximations of the RCU bound for the shell input AWGN channel, both of the form (49). We denote the approximation by RCU (saddlepoint) for α_n given by (55), and RCU (asymptotic) for α_n is given by (58). Figure 1 shows our two approximations, together with the exact Shannon bound [11, Eq. (19)], and the asymptotic Shannon bound [11, Eq. (5)]. We choose a rate R at 90% of $I(Q) = 1.0286$ bits/channel use, satisfying $R^*(Q) < R < I(Q)$, where

$$I(Q) = \frac{1}{2}\log\big(1+\mathrm{snr}\big), \quad (59)$$

$$R^*(Q) = \frac{1}{2}\log\left(\frac{\mathrm{snr}}{4} + \frac{1}{2}\sqrt{\frac{\mathrm{snr}^2}{4}+1} + \frac{1}{2}\right). \quad (60)$$

For completeness, we also include the Shannon lower bound [11, Eq. (15)]. As it can be seen from the figures, all curves are very close to each other.

A more informative parameter is the fourth-order term of the error probability expansion, denoted as β_n

$$\mathrm{rcu}(M) \simeq \frac{\beta_n}{\sqrt{n^{1+\hat{\rho}}}} \cdot e^{-n(E_0(\hat{\rho},\hat{\lambda})-\hat{\rho}R)}, \quad (61)$$

and shown in Figure 2. The numerical results suggest that the saddlepoint approximation (49) with α_n in (55) is an alternative expression for the asymptotic Shannon bound [11, Eq. (5)] . The main advantage of our saddlepoint approximation is that it is applicable to cost-constrained memoryless channels under some strongly non-lattice conditions.

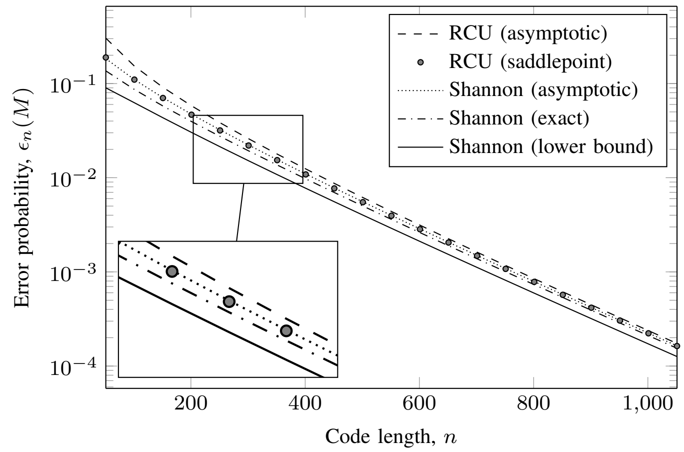

Fig. 1. Random coding error probability bounds versus the code length n at a rate $R = 0.9258$ bits/channel use, and $\mathrm{snr} = 5$ dB.

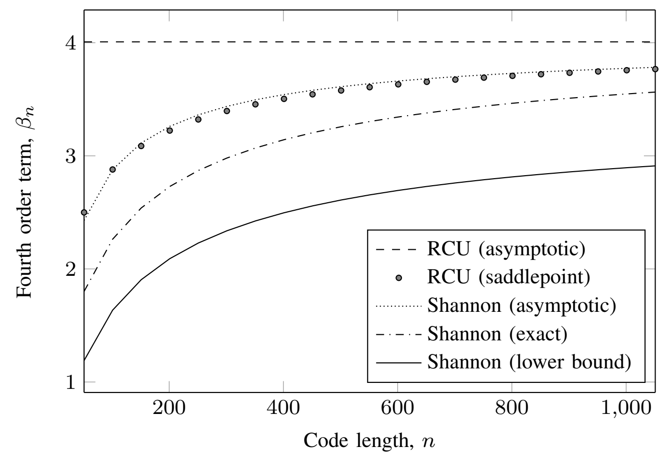

Fig. 2. Fourth order term β_n versus the code length n at a rate $R = 0.9258$ bits/channel use, and $\mathrm{snr} = 5$ dB.

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.

[4] 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.

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

[5] 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.

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

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

[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] 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.

[10] 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.

[11] 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.

[12] 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.

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

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