Efficient Error Probability Simulation of Coded Modulation over Fading Channels

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

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

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

***Abstract*—We use importance sampling to estimate the random-coding union (RCU) bound to the achievable error probability in coded-modulation wireless channels. We provide closed-form expressions of the exponentially-tilted distributions to generate the required samples, and illustrate the technique for coded BPSK modulation over the i.i.d. Rayleigh fading channel.**

I. Introduction

Evaluating the error probability of the transmission of coded data over a continuous-output channel is a common problem in digital communications. Efficient simulation methods of high-performance codes were proposed in, e. g., [1] for low density parity check (LDPC) codes. Together with other powerful codes such as polar codes and turbo codes, LDPC codes assume large code lengths. This assumption is yet not compatible with the ultra-high reliability and low latency requirements for next-generation wireless systems.

Instead of considering a good code, we study the random-coding union (RCU) bound to the achievable error probability [2, Eq. (62)]. Let $\boldsymbol{x}$ denote a transmitted codeword of length n drawn from a constellation $\mathcal{X}$, and let $\boldsymbol{y}$ be the received sequence taking values over $\mathbb{C}^n$. Random-coding arguments show the existence of a code of M codewords, transmitted over a memoryless channel with conditional density $W(y|x)$, whose error probability, the probability of decoding in favor of the wrong codeword, is at most the RCU, given by

$$\mathrm{rcu}_n = \int Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) \min\big\{1, (M-1)\mathrm{pep}_n(\boldsymbol{x},\boldsymbol{y})\big\}\, \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}, \tag{1}$$

where the pairwise error probability $\mathrm{pep}_n(\boldsymbol{x},\boldsymbol{y})$ reads

$$\mathrm{pep}_n(\boldsymbol{x},\boldsymbol{y}) = \int Q^n(\overline{\boldsymbol{x}}) \mathbb{1}\{W^n(\boldsymbol{y}|\overline{\boldsymbol{x}}) \geq W^n(\boldsymbol{y}|\boldsymbol{x})\}\, \mathrm{d}\overline{\boldsymbol{x}}, \tag{2}$$

and $\mathbb{1}\{\cdot\}$ is the indicator function. The expressions (1) and (2) are expectations with respect to the joint probability density

$$Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) Q^n(\overline{\boldsymbol{x}}). \tag{3}$$

The exact computation of the RCU bound is cumbersome even for simple channels and moderate values of n. Instead of resorting to approximations (e.g., [3]–[5]), we explore fast and accurate simulation to estimate (1).

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

II. Importance Sampling

We first note that both expectations (1) and (2) can be cast as follows. Let $f(\boldsymbol{z})$ be a non-negative function of some random variable $\boldsymbol{Z}$ with density $P(\boldsymbol{z})$. The standard Monte Carlo estimate of a quantity $p_n = \mathrm{E}\big[f(\boldsymbol{Z})\big]$ involves drawing N samples $\boldsymbol{z}_i$ from $P(\boldsymbol{z})$ and computing the average

$$\hat{p}_{n,N} = \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{z}_i). \tag{4}$$

The Monte Carlo estimator (4) is unbiased as its expected value satisfies $\mathrm{E}[\hat{p}_{n,N}] = p_n$. Besides, when $f(\boldsymbol{z})$ in (4) is an indicator function, it can be inferred that the number of samples needed to estimate p_n to a given accuracy level grows as $N \propto p_n^{-1}$, [6, Sec. 4.1]. Since the RCU bound decays exponentially with the codeword length n, this implies an exponential growth in the required number of samples.

Alternatively, importance sampling was proposed in [7] to diminish the sampling size in estimating the error probability of a communication scheme. Instead of estimating p_n as (4), this variance-reducing method involves generating i.i.d. samples from another distribution $\bar{P}(\boldsymbol{z})$ [7] to estimate p_n as

$$\hat{p}_{n,N} = \frac{1}{N}\sum_{i=1}^{N} \omega(\boldsymbol{z}_i) f(\boldsymbol{z}_i), \tag{5}$$

where the weights $\omega(\boldsymbol{z})$ that account for the distribution mismatch are given by the ratio $\omega(\boldsymbol{z}) = P(\boldsymbol{z})/\bar{P}(\boldsymbol{z})$.

A good choice for $\bar{P}(\boldsymbol{z})$ is known to be the exponential tilting [6] that exploits the exponential decay of p_n. For any value $s \geq 0$ and a function $g_n(\boldsymbol{z})$, we define the exponentially-tilted distribution

$$\bar{P}_{s,g}(\boldsymbol{z}) = P(\boldsymbol{z}) e^{s g_n(\boldsymbol{z}) - \kappa_n(s)} \tag{6}$$

in terms of the cumulant generating function [8] of $g_n(\boldsymbol{z})$,

$$\kappa_n(s) = \log \mathrm{E}\big[e^{s g_n(\boldsymbol{Z})}\big]. \tag{7}$$

The importance-sampling estimator (5) then becomes

$$\hat{p}_{n,N} = \hat{\alpha}_{n,N}(s) \cdot e^{\kappa_n(s)}, \tag{8}$$

where

$$\hat{\alpha}_{n,N}(s) = \frac{1}{N}\sum_{i=1}^{N} e^{-s g_n(\boldsymbol{z}_i)} f(\boldsymbol{z}_i) \tag{9}$$

and the samples $\boldsymbol{z}_i$ are independently drawn from $\bar{P}_{s,g}(\boldsymbol{z})$.

Roughy speaking, the importance-sampling estimator approximates the pre-exponential factor α_n in the quantity $p_n = \alpha_n(s) \cdot e^{\kappa_n(s)}$ by $\hat{\alpha}_{n,N}$, instead of directly estimating p_n. The importance-sampling estimator (8) is also unbiased [6, Sec. 4.2] with a normalized sample variance

$$\sigma_n^2 = \frac{\mathrm{E}\big[e^{\kappa_n(s)-sg_n(\boldsymbol{Z})}f(\boldsymbol{Z})^2\big]-p_n^2}{p_n^2} \tag{10}$$

that is now reduced by properly choosing the parameters involved in the exponential tilting, namely $s \geq 0$ and $g_n(\boldsymbol{z})$. A good choice of s is the minimizer

$$\hat{s}_n = \arg\min_{s\geq 0} \kappa_n(s), \tag{11}$$

whereas the choice of $g_n(\boldsymbol{z})$ depends on the structure of $f(\boldsymbol{z})$. We next apply the exponentially-tilted importance-sampling method described in this section to estimate (1).

III. Error Probability Estimation

We first note that for a fixed transmitted codeword $\boldsymbol{x}$ and received sequence $\boldsymbol{y}$, a nested estimator of the pairwise error probability (2) is needed. A good choice of $g_n(\overline{\boldsymbol{x}})$ for the importance-sampling estimate of the pairwise error probability in (2) with integration variable $\overline{\boldsymbol{x}}$ is the log-likelihood ratio

$$\ell_n(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}}) = \log\frac{W^n(\boldsymbol{y}|\overline{\boldsymbol{x}})}{W^n(\boldsymbol{y}|\boldsymbol{x})}. \tag{12}$$

As stated later, this choice helps capturing the correct exponential decay of the pairwise error probability in terms of n. The cumulant generating function of $\ell_n(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{X}})$ is given by

$$\kappa_{n,\tau}(\boldsymbol{x},\boldsymbol{y}) = \log\mathrm{E}\big[e^{\tau\cdot\ell_n(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{X}})}\big] \tag{13}$$

and leads to the following tilted distribution $\bar{P}_\tau(\overline{\boldsymbol{x}}|\boldsymbol{y})$ in (6) for the estimation of $\mathrm{pep}_n(\boldsymbol{x},\boldsymbol{y})$

$$\bar{P}_\tau^n(\overline{\boldsymbol{x}}|\boldsymbol{y}) = \frac{1}{\mu_n(\boldsymbol{y})}Q^n(\overline{\boldsymbol{x}})W^n(\boldsymbol{y}|\overline{\boldsymbol{x}})^\tau, \tag{14}$$

where $\mu_n(\boldsymbol{y})$ is a normalizing factor. We remark that while the log-likelihood $\ell_n(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}})$ depends on the transmitted codeword $\boldsymbol{x}$, the conditional distribution (14) for the codeword $\overline{\boldsymbol{x}}$ depends only on the received sequence $\boldsymbol{y}$ through the tilted channel density $W(\boldsymbol{y}|\overline{\boldsymbol{x}})^\tau$.

The importance-sampling estimator of the pairwise error probability generates N_1 independent samples $\overline{\boldsymbol{x}}_j$ from the conditional probability distribution (14), computes the average

$$\hat{\gamma}_{\tau,N_1}(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{N_1}\sum_{j=1}^{N_1} e^{-\tau\cdot\ell_n(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}}_j)}f_{\mathrm{pep}}(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}}_j), \tag{15}$$

where we defined

$$f_{\mathrm{pep}}(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}}) = \mathbb{1}\big\{\ell_n(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}}) \geq 0\big\}, \tag{16}$$

and finally obtains the final estimate

$$\hat{\mathrm{pep}}_{n,N_1}(\boldsymbol{x},\boldsymbol{y}) = \hat{\gamma}_{\tau,N_1}(\boldsymbol{x},\boldsymbol{y})\cdot e^{\kappa_{n,\tau}(\boldsymbol{x},\boldsymbol{y})}. \tag{17}$$

The tilting parameter τ is chosen as $\tau = \hat{\tau}_n(\boldsymbol{x},\boldsymbol{y})$, where

$$\hat{\tau}_n(\boldsymbol{x},\boldsymbol{y}) = \arg\min_{\tau\geq 0}\kappa_{n,\tau}(\boldsymbol{x},\boldsymbol{y}). \tag{18}$$

Note that τ used in the function $\kappa_{n,\tau}(\boldsymbol{x},\boldsymbol{y})$ depends on both $\boldsymbol{x}$ and $\boldsymbol{y}$. Yet, we drop the dependence on $\boldsymbol{x},\boldsymbol{y}$ in $\hat{\tau}_n$ to lighten the notation. Basic results in large-deviation theory imply that for memoryless channels the pairwise error probability (2) behaves exponentially as

$$\lim_{n\to\infty}\frac{\log\mathrm{pep}_n(\boldsymbol{x},\boldsymbol{y})}{\kappa_{n,\hat{\tau}_n}(\boldsymbol{x},\boldsymbol{y})} = 1. \tag{19}$$

We now address the importance-sampling estimate of the random-coding union bound in (1), an expectation with respect to the integration variables $\boldsymbol{x}$ and $\boldsymbol{y}$. In this case, we select the random variable

$$g_n(\boldsymbol{x},\boldsymbol{y}) = \log(M-1) + \kappa_{n,\frac{1}{1+\rho}}(\boldsymbol{x},\boldsymbol{y}) \tag{20}$$

because its cumulant generating function, given by

$$\chi_n(\rho) = \log\mathrm{E}\left[(M-1)^\rho\left(\frac{\mathrm{E}\big[W^n(\boldsymbol{Y}|\overline{\boldsymbol{X}})^{\frac{1}{1+\rho}}|\boldsymbol{Y}\big]}{W^n(\boldsymbol{Y}|\boldsymbol{X})^{\frac{1}{1+\rho}}}\right)^\rho\right], \tag{21}$$

gives the random-coding exponent [9, Sec. 5.6]. As a result, we will restrict the parameter ρ in the $[0,1]$ interval. Using (6), every pair of samples $(\boldsymbol{x}_i,\boldsymbol{y}_i)$ is drawn from

$$\bar{P}_\rho^n(\boldsymbol{x},\boldsymbol{y}) = Q^n(\boldsymbol{x})\bar{W}_\rho^n(\boldsymbol{y}|\boldsymbol{x}), \tag{22}$$

where $\bar{W}_\rho^n(\boldsymbol{y}|\boldsymbol{x})$ is the tilted channel density given by

$$\bar{W}_\rho^n(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{\mu_n}W^n(\boldsymbol{y}|\boldsymbol{x})^{\frac{1}{1+\rho}}\Big(\mathrm{E}[W^n(\boldsymbol{y}|\overline{\boldsymbol{X}})^{\frac{1}{1+\rho}}]\Big)^\rho \tag{23}$$

with normalizing factor μ_n. Inspecting (22), we observe that the transmitted codewords $\boldsymbol{x}_i$ are generated with the original random coding distribution $Q^n(\boldsymbol{x})$, whereas the received sequences $\boldsymbol{y}_i$ are drawn from the modified channel transition probability (23).

The importance-sampling estimator for the RCU bound (1) based on the independently generated pairs of samples $\boldsymbol{x}_i,\boldsymbol{y}_i$ from the probability distribution (22) is given by

$$\hat{\mathrm{rcu}}_{n,N_1,N_2} = \hat{\alpha}_{n,N_1,N_2}(\rho)\cdot e^{\chi_n(\rho)}, \tag{24}$$

where the pre-factor estimate reads

$$\hat{\alpha}_{n,N_1,N_2}(\rho) = \frac{1}{N_2}\sum_{i=1}^{N_2} e^{-\rho\cdot g_n(\boldsymbol{x}_i,\boldsymbol{y}_i)}f_{\mathrm{rcu}}(\boldsymbol{x}_i,\boldsymbol{y}_i) \tag{25}$$

with $f_{\mathrm{rcu}}(\boldsymbol{x}_i,\boldsymbol{y}_i)$ a function that depends on the pairwise error probability estimate (17) as

$$f_{\mathrm{rcu}}(\boldsymbol{x},\boldsymbol{y}) = \min\{1,(M-1)\hat{\mathrm{pep}}_{n,N_1}(\boldsymbol{x},\boldsymbol{y})\}. \tag{26}$$

For choice of

$$\hat{\rho}_n = \arg\min_{0\leq\rho\leq 1}\chi_n(\rho), \tag{27}$$

it follows from basic results in large-deviation theory that

$$\lim_{n\to\infty}\frac{\log\mathrm{rcu}_n}{\chi_n(\hat{\rho}_n)} = 1. \tag{28}$$

In summary, we proposed an importance-sampling estimator for the RCU bound (1) built from two nested estimators. Transmitted codewords $\boldsymbol{x}$ are drawn from the original random-coding distribution and received sequences $\boldsymbol{y}$ are generated from the modified channel transition probability (23) with optimal tilting parameter $\hat{\rho}_n$ related to the random-coding error exponent (28). For a given transmitted codeword and received sequence, the pairwise codewords $\overline{\boldsymbol{x}}$ are generated independently from $\boldsymbol{x}$ but conditioned on $\boldsymbol{y}$ from the conditional distribution (14) with optimal tilting parameter $\hat{\tau}_n$ related to the exponential decay of the pairwise error probability (19).

We remark that (14) and (23) might not be standard probability distributions. Yet, samples can be efficiently generated using, e. g., the rejection method described in [10, Ch. II.3].

We finally briefly discuss the performance analysis of the proposed importance-sampling estimator. We observe that $\mathrm{r\hat{c}u}_{n,N_1,N_2}$ is the sum of N_2 independent terms, each of them a nonlinear function of the inner estimator $\mathrm{p\hat{e}p}_{n,N_1}(\boldsymbol{x}_i, \boldsymbol{y}_i)$ that is also the sum of N_1 independent terms. Using refined central-limits theorems and Taylor expansions in inverse powers of N_1 and N_2, we show in [11] that for memoryless channels and sufficiently large code length n, as both N_1 and N_2 tend to infinity the importance-sampling estimator (24) converges in probability[1] to the exact RCU bound rcu_n according to

$$\mathrm{r\hat{c}u}_{n,N_1,N_2} \xrightarrow[N_1,N_2\to\infty]{\mathrm{p}} \mathrm{rcu}_n\left(1 - \frac{k_{1,n}}{N_1} + \sqrt{\frac{k_{2,n}}{N_2}}\Theta\right), \quad (30)$$

where $k_{1,n}$ and $k_{2,n}$ are positive numbers growing with n as $O(\sqrt{n})$, and Θ is the standard normal random variable.

Since $k_{1,n}$ in (30) is a positive term and Θ is a zero-mean random variable, it implies a negative bias in the estimation of the RCU bound. Yet, the estimator is consistent, as the bias vanishes as N_1 goes to infinity, although the bias might be significant for small values of N_1. The variance term $k_{2,n}$ in (30) grows as the squared root of n, implying a significant reduction in the variance with the importance-sampling estimator, as the number of samples needed to accurately estimate the RCU bound for a given confidence level grows as $N_2 \propto \sqrt{n}$, rather than the typical growth $N_2 \propto \mathrm{rcu}_n^{-1}$ in standard Monte Carlo [6, Sec. 4.1], which would be exponential in the code length n in our setting of a memoryless channel.

IV. Numerical Example

We illustrate the above importance-sampling estimator of the RCU bound for the binary phase-shift keying (BPSK) modulation. We denote the symbol set $\mathcal{X} = \{-\sqrt{P}, +\sqrt{P}\}$, where P is a positive number describing an average power

[1]Two sequences of random variables A_N and B_N indexed by N are said to converge in probability if for all $\varepsilon > 0$, it holds

$$\lim_{N\to\infty} \Pr[|A_N - B_N| > \varepsilon] = 0. \quad (29)$$

We denote the convergence in probability by $A_N \xrightarrow[N\to\infty]{\mathrm{p}} B_N$.

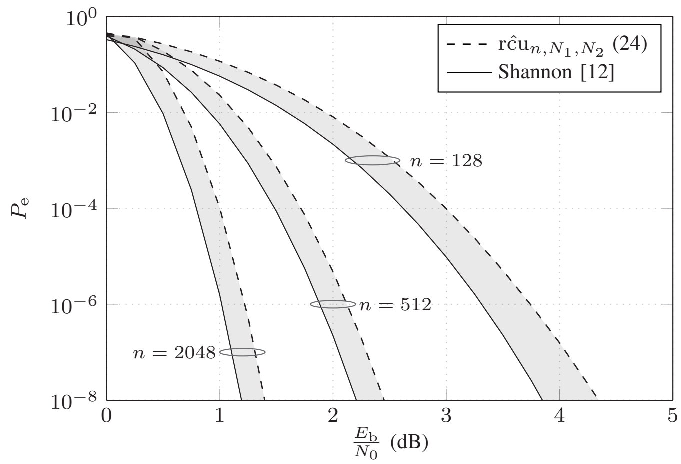

Fig. 1. Error probability versus E_b/N_0 over the AWGN channel, for code rate $R_b = 0.5$, $N_1 = N_2 = 500$ samples, and several code lengths n.

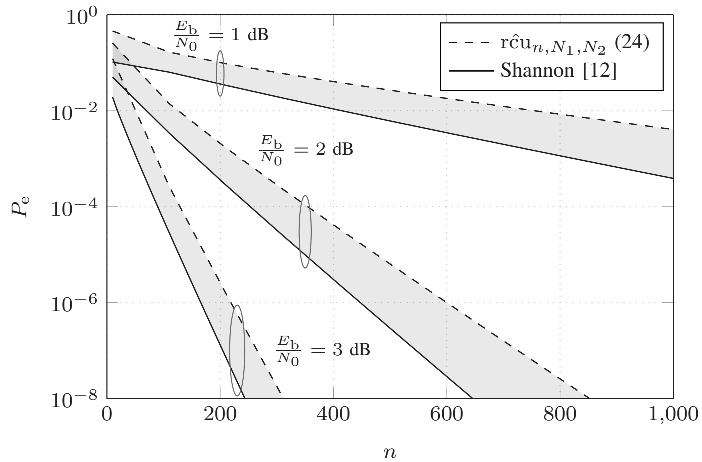

Fig. 2. Error probability versus n over the AWGN channel, for code rate $R_b = 0.5$, $N_1 = N_2 = 500$ samples, and several values of E_b/N_0.

constraint. A codeword $\boldsymbol{x} = (x_1, \ldots, x_n)$ is transmitted over the i.i.d. Rayleigh fading channel described by

$$y_i = h_i x_i + w_i, \quad (31)$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)$ is the received sequence, $\boldsymbol{w} = (w_1, \ldots, w_n)$ is an i.i.d. real-valued zero-mean Gaussian noise with variance σ^2. Since the phase of the fading coefficients is irrelevant, we assume that $\boldsymbol{h} = (h_1, \ldots, h_n)$ is a real-valued i.i.d. Rayleigh distributed with density

$$p^n(\boldsymbol{h}) = \prod_{i=1}^{n} 2h_i e^{-h_i^2} \mathbb{1}\{h_i \geq 0\}. \quad (32)$$

The symmetry of BPSK implies that the input distribution $Q^n(\boldsymbol{x})$ that optimizes both the exponential decay (28) and the channel capacity, denoted as C_b, is the uniform distribution

$$Q^n(\boldsymbol{x}) = \frac{1}{2^n}. \quad (33)$$

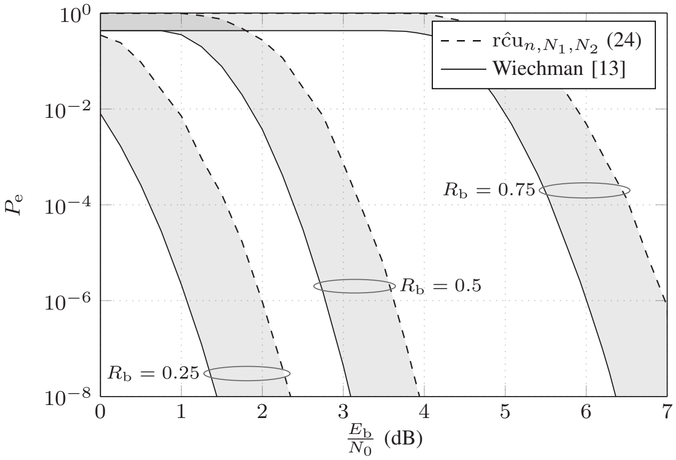

Fig. 3. Error probability versus E_b/N_0 over the i.i.d. Rayleigh channel, for code length $n = 1024$, $N_1 = N_2 = 500$ samples, and several code rates R_b.

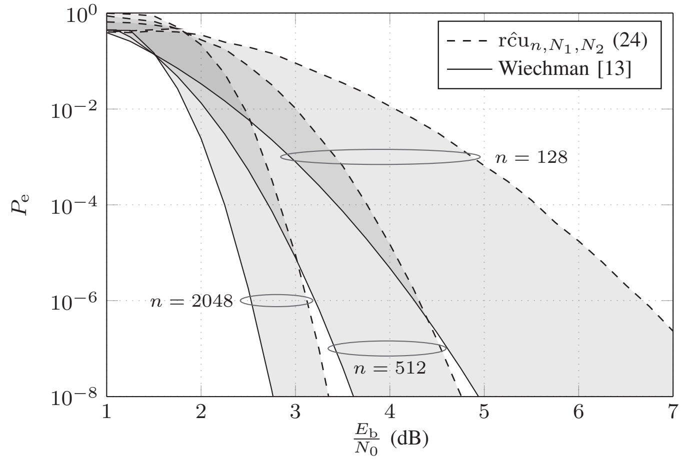

Fig. 4. Error probability versus E_b/N_0 over the i.i.d. Rayleigh channel, for code rate $R_\mathrm{b} = 0.5$, $N_1 = N_2 = 500$ samples, and several code lengths n.

The input distribution $Q^n(\boldsymbol{x})$, together with the channel conditional density given by

$$W^n(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - h_i x_i)^2}{2\sigma^2}}, \tag{34}$$

determine the required parameters for the importance-sampling estimator (24), namely the the cumulant-generating functions (13) and (21), the optimal tilting parameters $\hat{\tau}_n$ and $\hat{\rho}_n$ respectively in (18) and (27), and the tilted distributions (14) and (23). The additive white Gaussian noise (AWGN) channel can be recovered from (34) by setting

$$p^n(\boldsymbol{h}) = \delta_n(\boldsymbol{h} - \boldsymbol{1}) \tag{35}$$

where $\delta_n(\cdot)$ is the n-dimensional Dirac delta, and $\boldsymbol{1}$ is the all-ones length-n vector. As usual, we define the code rate as

$$R_\mathrm{b} = \frac{1}{n}\log_2 M, \tag{36}$$

and the coded average E_b/N_0 ratio as

$$\frac{E_\mathrm{b}}{N_0} = \frac{P}{\sigma^2} \cdot \frac{1}{2R_\mathrm{b}}. \tag{37}$$

We set $N_1 = N_2 = 500$ to estimate the achievable error probability by means of the RCU, and include Shannon's sphere-packing bound [12, Eq. (15)] for the AWGN channel or an improved sphere-packing bound [13, Th. 3.1] for the i.i.d. Rayleigh fading channel. The error probability of good binary codes must lie between the RCU and the sphere-packing bounds, as shown in Figs. 1–4 in gray-shaded regions for several configurations of codeword length n, code rate R_b and coded E_b/N_0 ratio. In the presence of fading, we observe a larger gap between achievability and converse bounds compared to the AWGN case, especially for small values of n. As another example, a performance loss of approximately 2 dB in E_b/N_0 is noticed at $n = 2048$ in Fig. 4 for the fading case when compared to the AWGN case in Fig. 1.

V. Conclusion

In this paper, we proposed an importance-sampling technique to estimate the random-coding union (RCU) bound to the achievable error probability for the transmission of coded data over a continuous-output channel. We derived closed-form expressions for the optimal tilted distributions needed to generate the samples of the two nested estimators involved, and illustrated the transmission of the coded BPSK modulation over the AWGN and i.i.d. Rayleigh fading channels.

References

[1] S. Ahn, K. Yang, and D. Har, "Evaluation of the low error-rate performance of LDPC codes over Rayleigh fading channels using importance sampling," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2166–2177, 2013.

[2] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

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

[4] 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.

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

[6] J. Bucklew, *Introduction to Rare Event Simulation*. New York: Springer-Verlag, 2013.

[7] K. Shanmugam and P. Balaban, "A modified Monte-Carlo simulation technique for the evaluation of error rate in digital communication systems," *IEEE Trans. Commun.*, vol. 28, no. 11, pp. 1916–1924, Nov. 1980.

[8] R. Durrett, *Probability: Theory and Examples*. Belmont, CA: Duxbury, 1996.

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

[10] L. Devroye, *Non-Uniform Random Variate Generation*. New York: Springer-Verlag, 1986.

[11] J. Font-Segura, A. Martinez, and A. Guillén i Fàbregas, "Performance analysis of the RCU importance-sampling estimator," 2019. [Online]. Available: `itc.upf.edu/biblio/1114`

[12] 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.

[13] G. Wiechman and I. Sason, "An improved sphere-packing bound for finite-length codes over symmetric memoryless channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1962–1990, 2008.