Importance Sampling for Coded-Modulation Error Probability Estimation

Josep Font-Segura, *Member, IEEE,* Alfonso Martinez, *Senior Member, IEEE,* and Albert Guillén i Fàbregas, *Senior Member, IEEE,*

***Abstract*—This paper proposes an efficient simulation method based on importance sampling to estimate the random-coding error probability of coded modulation. The technique is valid for complex-valued modulations over Gaussian channels, channels with memory, and naturally extends to fading channels. The simulation method is built on two nested importance samplers to respectively estimate the pairwise error probability and generate the channel input and output. The effect of the respective number of samples on the overall bias and variance of the estimate of the error probability is characterized. For a memoryless channel, the estimator is shown to be consistent and with a small variance, growing with the square root of the code length, rather than the exponential growth of a standard Monte Carlo estimator.**

***Index Terms*—Coded modulation, random coding, error probability, Monte Carlo simulation, importance sampling.**

I. Introduction

IMPORTANCE sampling [2], an improved Monte Carlo simulation in which samples are generated according to tilted distributions, may significantly reduce the sampling size in estimating the error probability of a communication scheme [3]. For instance, the transmission of uncoded symbols over the AWGN channel was studied in [4]–[6] with importance sampling techniques, where sampling distributions involve variance scaling and mean translations. Efficient simulation methods of high-performance codes were proposed in, e. g., [7], [8] for low density parity check (LDPC) codes.

In this paper, we study the error probability in the detection of coded-modulation signals. Currently, most powerful codes such as polar codes, LDPC codes or turbo codes have large code lengths, an assumption incompatible with the low-latency and ultra-high reliability requirements for next-generation wireless systems. Recently, Polyanskiy, Poor and Verdú [9] derived tight bounds to the error probability of random codes valid for short code lengths. With random codes, a common tool used in information theory to show the existence of good codes at rates below the channel capacity, one studies the error probability averaged over all possible randomly-generated codes. Evaluating the random-coding error probability, rather than that of a given code, becomes a useful tool to characterize the performance of coded modulation, primarily in channels where good codes are unexplored or unknown. We thus focus on the random coding union (RCU) bound [9, Eq. (62)] to describe the error probability of good codes of arbitrary length.

The exact computation of the RCU bound is cumbersome even for short code lengths, as it involves high-dimensional integrations; we address numerical simulation instead. Yet, simulation of such small a quantity would require a number of samples exponential in the code length to achieve an acceptable level of precision [10]. To solve this rare-event simulation problem, we find an importance-sampling tilting that explicitly exploits the known exponential decay of the RCU bound with the code length to estimate the pre-exponential factor of the coded error probability, instead of the full probability.

The rest of the paper is organized as follows. In Sec. II, we describe the error probability of random codes and outline the computational challenges to calculate the RCU bound. We present our efficient importance-sampling simulator of the RCU bound in Sec. III, and derive closed-form expressions of the optimal tilted distributions valid for any block coded modulation. Our estimator consists of two nested importance samplers, respectively related to the estimate of a pairwise error probability and to the generation of the channel input and output. In Sec. IV, we carry out an asymptotic performance analysis for memoryless channels to describe the effect of the number of samples on the overall bias and variance of the nested estimator. We consider some examples with coded binary phase-shift keying (BPSK) modulation over the AWGN and the i.i.d. Rayleigh fading channels in Sec. V, and summarize the main contributions of the paper in Sec. VI.

II. Coded-Modulation Error Probability

Let $\mathcal{C}$ be a block code with $M = 2^k$ codewords $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$, where k is the number of information bits. Each codeword $\boldsymbol{x}$ has n symbols drawn from a constellation $\mathcal{X}$. The code rate is given by $R_{\mathrm{b}} = \frac{k}{n}$ bits per channel use. This code is used for transmission over a channel with conditional probability density $W^n(\boldsymbol{y}|\boldsymbol{x})$, where the length-$n$ sequence $\boldsymbol{y}$ represents the equivalent baseband channel output. The error probability, denoted by $P_{\mathrm{e}}(\mathcal{C})$, is the probability of decoding in favor of a codeword $\boldsymbol{x}_j$ other than the transmitted one $\boldsymbol{x}_m$, $j \neq m$.

The computation of $P_{\mathrm{e}}(\mathcal{C})$ is challenging, due to the complex code structure for large values of n, because of the exponentially large number of messages M, or since good codes

J. Font-Segura and A. Martinez are with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona 08018, Spain; A. Martinez is a Serra Húnter Fellow. (e-mail: {josep.font, alfonso.martinez}@ieee.org).

A. Guillén i Fàbregas is with Institució Catalana de Recerca i Estudis Avançats, Barcelona 08010, Spain, the Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona 08018, Spain, and also with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K. (e-mail: guillen@ieee.org).

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.

themselves are unknown for small n. Instead of considering a fixed code $\mathcal{C}$, we study the error probability $P_{\mathrm{e},n}$ averaged over all codes of M codewords generated by independent drawings from some input distribution $Q^n(\boldsymbol{x})$. Shannon's random coding arguments show the existence of at least one good code whose error probability is at most $P_{\mathrm{e},n}$, the expectation of $P_{\mathrm{e}}(\mathcal{C})$ over all possible codes. Evaluating $P_{\mathrm{e},n}$ rather than $P_{\mathrm{e}}(\mathcal{C})$ is not only of theoretical importance but also serves as a performance benchmark for the designers of good codes.

A minor relaxation of the probability $P_{\mathrm{e},n}$ is given by the random coding union (RCU) bound to the random coding error probability [9, Eq. (62)], satisfying $P_{\mathrm{e},n} \leq \mathrm{rcu}_n$, where

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

In (1), the pairwise error probability $\mathrm{pep}_n(\boldsymbol{x},\boldsymbol{y})$ is the probability that the decoder decodes in favor of another independent random codeword $\overline{\boldsymbol{x}}$ for fixed transmitted codeword $\boldsymbol{x}$ and received sequence $\boldsymbol{y}$, i.e.,

$$\mathrm{pep}_n(\boldsymbol{x},\boldsymbol{y}) = \int Q^n(\overline{\boldsymbol{x}}) \mathbb{1}\{\ell_n(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}}) \geq 0\}\, \mathrm{d}\overline{\boldsymbol{x}}, \tag{2}$$

and $\mathbb{1}\{\cdot\}$ is the indicator function taking the value one if the condition is satisfied and zero otherwise, and $\ell_n(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}})$ is the log-likelihood ratio

$$\ell_n(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}}) = \log \frac{W^n(\boldsymbol{y}|\overline{\boldsymbol{x}})}{W^n(\boldsymbol{y}|\boldsymbol{x})}. \tag{3}$$

In short, the RCU bound characterizes the error probability of good codes with rate R_{b} and length n, whose error probability is as good as the right-hand side (r.h.s.) of (1).

The expressions for rcu_n and $\mathrm{pep}_n(\boldsymbol{x},\boldsymbol{y})$ in (1) and (2) respectively are both given by an expectation of a non-negative function $f(\boldsymbol{z})$ of some random variable $\boldsymbol{Z}$ with density $P(\boldsymbol{z})$,

$$p_n = \mathrm{E}\bigl[f(\boldsymbol{Z})\bigr], \tag{4}$$

where from now on we write the expectation operation as $\mathrm{E}[\cdot]$ for the sake of compactness. Evaluating the expectation in (1) involves integrations over joint probability densities

$$Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x}) Q^n(\overline{\boldsymbol{x}}), \tag{5}$$

which is complex even for simple channels and moderate values of n. Instead of resorting to approximations (e.g., [11]–[14]), we explore fast and accurate simulation to estimate (1). While the proposed estimator is valid for generic channel law $W^n(\boldsymbol{y}|\boldsymbol{x})$ and input $Q^n(\boldsymbol{x})$, we restrict our analysis to memoryless channels with product input distributions, for which $W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^n W(y_i|x_i)$ and $Q^n(\boldsymbol{x}) = \prod_{i=1}^n Q(x_i)$.

III. Importance Sampling

The standard Monte Carlo estimate of a quantity p_n as in (4), denoted by $\hat{p}_{n,N}$, involves generating N samples, $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N$, according to $P(\boldsymbol{z})$ and computing the average

$$\hat{p}_{n,N} = \frac{1}{N} \sum_{i=1}^{N} f(\boldsymbol{z}_i). \tag{6}$$

By construction, p_n is the mean of each summand in (6). We also define σ_n^2 to be the variance, normalized to the squared mean, of each summand in (6).

In order to describe the accuracy of an estimator in approaching the exact value of the quantity p_n as N increases, we make use of the notion of convergence in probability. 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

$$\lim_{N\to\infty} \Pr[|A_N - B_N| > \varepsilon] = 0. \tag{7}$$

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

Using the central limit theorem [15, Ch. XV.5], as $N \to \infty$ the estimator $\hat{p}_n$ converges in probability to

$$\hat{p}_{n,N} \xrightarrow[N\to\infty]{\mathrm{p}} p_n \biggl(1 + \frac{\sigma_n}{\sqrt{N}} \Theta_N\biggr), \tag{8}$$

where and Θ_N has a probability density that converges uniformly in θ to the density $p_\Theta(\theta)$ of a standard normal random variable Θ. Equivalently, the relative error converges to a normal random variable with zero mean and variance σ_n^2/N,

$$\frac{\hat{p}_{n,N} - p_n}{p_n} \xrightarrow[N\to\infty]{\mathrm{p}} \frac{\sigma_n}{\sqrt{N}} \Theta. \tag{9}$$

The Monte Carlo estimator $\hat{p}_{n,N}$ is unbiased since its expected value coincides with the quantity to be estimated, namely p_n. Besides, when $f(\boldsymbol{z})$ in (6) is an indicator function, using (9) and the fact that the variance σ_n^2 is normalized to the squared mean, i. e. p_n^2, we infer that the number of samples needed to estimate p_n to a given confidence level grows as $N \propto p_n^{-1}$, [10, Sec. 4.1].

Alternatively, importance sampling is a variance-reducing estimation technique that involves the generation of i.i.d. samples from another distribution $\bar{P}(\boldsymbol{z})$ [2] to estimate p_n as

$$\hat{p}_{n,N} = \frac{1}{N} \sum_{i=1}^{N} \omega(\boldsymbol{z}_i) f(\boldsymbol{z}_i), \tag{10}$$

where the weights $\omega(\boldsymbol{z})$ account for the distribution mismatch and are given by $\omega(\boldsymbol{z}) = P(\boldsymbol{z})/\bar{P}(\boldsymbol{z})$. A good choice for $\bar{P}(\boldsymbol{z})$ is known to be the exponential tilting [10]. For any $s \geq 0$ and a certain function $g_n(\boldsymbol{z})$, define the tilted distribution

$$\bar{P}_{s,g}(\boldsymbol{z}) = P(\boldsymbol{z}) e^{s g_n(\boldsymbol{z}) - \kappa_n(s)}, \tag{11}$$

where $\kappa_n(s)$ is the cumulant generating function [16] of $g_n(\boldsymbol{z})$,

$$\kappa_n(s) = \log \mathrm{E}\bigl[e^{s g_n(\boldsymbol{Z})}\bigr], \tag{12}$$

and the weights are given by

$$\omega_{s,g}(\boldsymbol{z}) = e^{\kappa_n(s) - s g_n(\boldsymbol{z})}. \tag{13}$$

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. Hence, the importance-sampling estimator (10) becomes

$$\hat{p}_{n,N} = \hat{\alpha}_{n,N}(s) \cdot e^{\kappa_n(s)}, \tag{14}$$

where $\hat{\alpha}_{n,N}(s)$ is given by

$$\hat{\alpha}_{n,N}(s) = \frac{1}{N} \sum_{i=1}^{N} e^{-s g_n(\boldsymbol{z}_i)} f(\boldsymbol{z}_i) \tag{15}$$

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

The importance-sampling estimator (14) is unbiased [10, Sec. 4.2] with a normalized sample variance given by

$$\sigma_n^2 = \frac{\mathrm{E}\big[e^{\kappa_n(s)-sg_n(\boldsymbol{Z})}f(\boldsymbol{Z})^2\big]-p_n^2}{p_n^2}. \tag{16}$$

The limit in (8) remains valid with a normalized sample variance that is now reduced by properly choosing the parameters $s \geq 0$ and $g_n(\boldsymbol{z})$. A good choice of s is the minimizer of the cumulant generating function $\kappa_n(s)$, i.e.,

$$\hat{s}_n = \underset{s\geq 0}{\arg\min}\, \kappa_n(s). \tag{17}$$

A. Pairwise Error Probability

For the importance-sampling estimate of the pairwise error probability in (2) with integration variable $\overline{\boldsymbol{x}}$, we select $g_n(\overline{\boldsymbol{x}})$ to be the log-likelihood ratio in (3), $g_n(\overline{\boldsymbol{x}}) = \ell_n(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}})$. As mentioned at the end of this subsection, this choice helps capturing the correct exponential decay of the pairwise error probability in terms of n for memoryless channels. The corresponding cumulant generating function is given by

$$\kappa_{n,\tau}(\boldsymbol{x},\boldsymbol{y}) = \log \mathrm{E}\big[e^{\tau\cdot\ell_n(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{X}})}\big]. \tag{18}$$

For this choice, the tilted distribution $\bar{P}_\tau(\overline{\boldsymbol{x}}|\boldsymbol{y})$ in (11) for the estimation of $\mathrm{pep}_n(\boldsymbol{x},\boldsymbol{y})$ can be explicitly computed as

$$\bar{P}_\tau^n(\overline{\boldsymbol{x}}|\boldsymbol{y}) = \frac{Q^n(\overline{\boldsymbol{x}})W^n(\boldsymbol{y}|\overline{\boldsymbol{x}})^\tau}{\displaystyle\int Q^n(\boldsymbol{x}')W^n(\boldsymbol{y}|\boldsymbol{x}')^\tau \mathrm{d}\boldsymbol{x}'}. \tag{19}$$

While the log-likelihood $\ell_n(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}})$ depends on the channel input $\boldsymbol{x}$, this conditional distribution for the codeword $\overline{\boldsymbol{x}}$ depends only on the channel output $\boldsymbol{y}$.

The importance-sampling estimator of the pairwise error probability draws N_1 independent samples $\overline{\boldsymbol{x}}_j$, for $j = 1,\dots,N_1$, from the probability distribution (19) to compute

$$\hat{\gamma}_{\tau,N_1}(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{N_1}\sum_{j=1}^{N_1} e^{-\tau\cdot\ell_n(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}}_j)} f_{\mathrm{pep}}(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}}_j), \tag{20}$$

where

$$f_{\mathrm{pep}}(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}}) = \mathbb{1}\big\{\ell_n(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}}) \geq 0\big\}, \tag{21}$$

to generate the final estimate

$$\hat{\mathrm{pep}}_{n,N_1}(\boldsymbol{x},\boldsymbol{y}) = \hat{\gamma}_{\tau,N_1}(\boldsymbol{x},\boldsymbol{y}) \cdot e^{\kappa_{n,\tau}(\boldsymbol{x},\boldsymbol{y})}. \tag{22}$$

Based on (17), we select $\tau = \hat{\tau}_n(\boldsymbol{x},\boldsymbol{y})$ given by

$$\hat{\tau}_n(\boldsymbol{x},\boldsymbol{y}) = \underset{\tau\geq 0}{\arg\min}\, \kappa_{n,\tau}(\boldsymbol{x},\boldsymbol{y}). \tag{23}$$

Both the optimal parameter τ used in the function $\kappa_{n,\tau}(\boldsymbol{x},\boldsymbol{y})$ and the estimator $\hat{\gamma}_{\tau,N_1}$ depend on $\boldsymbol{x},\boldsymbol{y}$. Yet, we henceforth drop the dependence on $\boldsymbol{x},\boldsymbol{y}$ in $\hat{\tau}_n$ to lighten the notation. For the optimal choice of $\hat{\tau}_n$, it follows from basic results in large-deviation theory [17, Th. 2.2.3] that for memoryless channels the pairwise error probability (2) behaves as

$$\lim_{n\to\infty} \frac{\log \mathrm{pep}_n(\boldsymbol{x},\boldsymbol{y})}{\kappa_{n,\hat{\tau}_n}(\boldsymbol{x},\boldsymbol{y})} = 1. \tag{24}$$

B. Random-Coding Error Probability

For the importance-sampling estimate of the random-coding union bound in (1), an expectation with respect to the integration variables $\boldsymbol{x}$ and $\boldsymbol{y}$, we choose

$$g_n(\boldsymbol{x},\boldsymbol{y}) = \log(M-1) + \kappa_{n,\frac{1}{1+\rho}}(\boldsymbol{x},\boldsymbol{y}), \tag{25}$$

where $\kappa_{n,\frac{1}{1+\rho}}(\boldsymbol{x},\boldsymbol{y})$ is given in (18). We can compute the cumulant generating function $\chi_n(\rho)$ of $g_n(\boldsymbol{x},\boldsymbol{y})$ as

$$\chi_n(\rho) = \log \mathrm{E}\left[(M-1)^\rho\left(\frac{\mathrm{E}\big[W^n(\boldsymbol{Y}|\overline{\boldsymbol{X}})^{\frac{1}{1+\rho}}|\boldsymbol{Y}\big]}{W^n(\boldsymbol{Y}|\boldsymbol{X})^{\frac{1}{1+\rho}}}\right)^\rho\right]. \tag{26}$$

From (11), the distribution used for generating the pair of samples $(\boldsymbol{x}_i,\boldsymbol{y}_i)$ is given by

$$\bar{P}_\rho^n(\boldsymbol{x},\boldsymbol{y}) = Q^n(\boldsymbol{x})\bar{W}_\rho^n(\boldsymbol{y}|\boldsymbol{x}), \tag{27}$$

where $\bar{W}_\rho^n(\boldsymbol{y}|\boldsymbol{x})$ is a tilted channel transition probability,

$$\bar{W}_\rho^n(\boldsymbol{y}|\boldsymbol{x}) = \frac{W^n(\boldsymbol{y}|\boldsymbol{x})^{\frac{1}{1+\rho}}\Big(\mathrm{E}[W^n(\boldsymbol{y}|\overline{\boldsymbol{X}})^{\frac{1}{1+\rho}}]\Big)^\rho}{\displaystyle\int\Big(\mathrm{E}[W^n(\boldsymbol{y}'|\boldsymbol{X}')^{\frac{1}{1+\rho}}]\Big)^{1+\rho}\mathrm{d}\boldsymbol{y}'}. \tag{28}$$

Equation (27) implies that the channel input sequences $\boldsymbol{x}_i$ are generated with the original random-coding distribution $Q^n(\boldsymbol{x})$ from Sec. II, whereas the channel output sequences $\boldsymbol{y}_i$ are drawn from the modified channel transition probability (28).

Finally, the importance-sampling estimator for the RCU based on the independently generated pairs of samples $\boldsymbol{x}_i,\boldsymbol{y}_i$, for $i = 1,\dots,N_2$, from the probability distribution (27) is

$$\hat{\mathrm{rcu}}_{n,N_1,N_2} = \hat{\alpha}_{n,N_1,N_2}(\rho)\cdot e^{\chi_n(\rho)}, \tag{29}$$

where

$$\hat{\alpha}_{n,N_1,N_2}(\rho) = \frac{1}{N_2}\sum_{i=1}^{N_2} e^{-\rho\cdot g_n(\boldsymbol{x}_i,\boldsymbol{y}_i)} f_{\mathrm{rcu}}(\boldsymbol{x}_i,\boldsymbol{y}_i) \tag{30}$$

and

$$f_{\mathrm{rcu}}(\boldsymbol{x},\boldsymbol{y}) = \min\{1,(M-1)\hat{\mathrm{pep}}_{n,N_1}(\boldsymbol{x},\boldsymbol{y})\}. \tag{31}$$

Based on (17), and taking into account that the cumulant generating function (26) gives the random-coding exponent [18, Sec. 5.6], we select ρ as

$$\hat{\rho}_n = \underset{0\leq\rho\leq 1}{\arg\min}\, \chi_n(\rho). \tag{32}$$

For the choice of $\hat{\rho}_n$, it follows from basic results in large-deviation theory [17, Th. 2.2.3] that for memoryless channels

$$\lim_{n\to\infty}\frac{\log \mathrm{rcu}_n}{\chi_n(\hat{\rho}_n)} = 1. \tag{33}$$

The nested importance-sampling estimator is summarized in pseudo-code in Algorithm 1 on the next page. In summary, the channel $W^n(\boldsymbol{y}|\boldsymbol{x})$, the random-coding input distribution of $Q^n(\boldsymbol{x})$, and the information rate R_b jointly determine the outer tilting parameter $\hat{\rho}_n$, while auxiliary codewords are generated in accordance to the inner tilting parameter $\hat{\tau}_n(\boldsymbol{x},\boldsymbol{y})$ for a given channel input and output pair $\boldsymbol{x},\boldsymbol{y}$. As expected, the algorithm extends the classical Monte Carlo method, which can be recovered by setting $\hat{\rho}_n = \hat{\tau}_n = 0$.

Algorithm 1: Importance-sampling estimate of the RCU bound

Input: $Q^n(\boldsymbol{x})$, $W^n(\boldsymbol{y}|\boldsymbol{x})$, n, R_b, N_2 and N_1
Output: $\hat{\mathrm{rcu}}$
calculate $M = \lfloor 2^{nR_\mathrm{b}} \rfloor$;
calculate $\chi_n(\rho)$ from (26);
select $\rho \leftarrow \arg\min_{0\le\rho\le1} \chi_n(\rho)$;
find $\bar{W}^n_\rho(\boldsymbol{y}|\boldsymbol{x})$ from (28);
$\alpha \leftarrow 0$;
for $i = 1; i \le N_2$ **do**
  generate $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ according to $Q^n(\boldsymbol{x})\bar{W}^n_\rho(\boldsymbol{y}|\boldsymbol{x})$;
  compute $\kappa_{n,\tau}(\boldsymbol{x}_i, \boldsymbol{y}_i)$ from (18);
  select $\tau \leftarrow \arg\min_{\tau\ge0} \kappa_{n,\tau}(\boldsymbol{x}_i, \boldsymbol{y}_i)$;
  find $\bar{P}^n_\tau(\overline{\boldsymbol{x}}|\boldsymbol{y}_i)$ from (19);
  $\gamma \leftarrow 0$;
  for $j = 1; j \le N_1$ **do**
    generate $\overline{\boldsymbol{x}}_j$ according to $\bar{P}^n_\tau(\overline{\boldsymbol{x}}|\boldsymbol{y}_i)$;
    $\gamma \leftarrow \gamma + \frac{1}{N_1} e^{-\tau\cdot\ell_n(\boldsymbol{x}_i,\boldsymbol{y}_i,\overline{\boldsymbol{x}}_j)} \mathbb{1}\{\ell_n(\boldsymbol{x}_i, \boldsymbol{y}_i, \overline{\boldsymbol{x}}_j) \ge 0\}$;
  $\hat{\mathrm{pep}} \leftarrow \gamma \cdot e^{\kappa_{n,\tau}(\boldsymbol{x}_i,\boldsymbol{y}_i)}$;
  $\alpha \leftarrow \alpha + \frac{1}{N_2} e^{-\rho\cdot g_n(\boldsymbol{x}_i,\boldsymbol{y}_i)} \min\{1, (M-1)\hat{\mathrm{pep}}\}$;
$\hat{\mathrm{rcu}} \leftarrow \alpha \cdot e^{\chi_n(\rho)}$;
return $\hat{\mathrm{rcu}}$;

IV. Performance Analysis

In the previous section, we presented in Algorithm 1 an importance-sampling estimator for the RCU bound in (1). Built from two nested estimators, the mapping from the general tilting for importance sampling discussed at the beginning of Sec. III was relatively straightforward and led to the respective tilting parameters $\hat{\tau}_n$ and $\hat{\rho}_n$ given in (23) and (32).

In contrast, the performance analysis is subtler, since the outer estimator $\hat{\mathrm{rcu}}_{n,N_1,N_2}$ is the sum of N_2 independent terms $f_{\mathrm{rcu}}(\boldsymbol{x}_i, \boldsymbol{y}_i)$ in (31), each of them a nonlinear function of the inner estimator $\hat{\mathrm{pep}}_{n,N_1}(\boldsymbol{x}_i, \boldsymbol{y}_i)$. As one estimator is nested inside the other, a more refined analysis than the central-limit theorem is needed to study the consistency, bias, and variance of Algorithm 1. In this section, we derive an asymptotic expansion of our estimator for memoryless channels and large values of N_1 and N_2, as summarized in the following theorem.

Theorem. *For memoryless channels, as both numbers of samples N_1 and N_2 tend to infinity the importance-sampling estimator of Algorithm 1 converges in probability to the exact RCU bound rcu_n according to*

$$\hat{\mathrm{rcu}}_{n,N_1,N_2} \xrightarrow[N_1,N_2\to\infty]{\mathrm{p}} \mathrm{rcu}_n\left(1 - \frac{k_{1,n}}{N_1} + \sqrt{\frac{k_{2,n}}{N_2}}\Theta\right), \quad (34)$$

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.

The positive term $k_{1,n}$ in (34), linked in Sec. IV-A to the variance of the estimate of the pairwise error probability $\hat{\mathrm{pep}}_{n,N_1}(\boldsymbol{x}, \boldsymbol{y})$, induces a negative bias in the estimation of the RCU bound. The estimator is asymptotically consistent, as the bias vanishes as N_1 goes to infinity, although the bias might be significant for small values of N_1 because its value cannot be reduced by increasing the value of N_2. In Sec. V, we numerically validate the expansion in the theorem.

The expression for $k_{2,n}$ in (34) is linked to 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 [10, Sec. 4.1], which would be exponential behaviour in the code length n in our setting of a memoryless channel.

Our result (34) is valid under the assumption that $\mathrm{pep}_n(\boldsymbol{X}, \boldsymbol{Y})$ is a strongly non-lattice random variable with a continuous and differentiable density, and that the terms $f_{\mathrm{pep}}(\boldsymbol{x}, \boldsymbol{y}, \overline{\boldsymbol{X}})$ and $f_{\mathrm{rcu}}(\boldsymbol{X}, \boldsymbol{Y})$ have an integrable characteristic function under the joint densities (5). Both assumptions are plausible for the transmission of complex-valued modulations over continuous-output channels considered in this work.

The two remaining subsections are respectively devoted to deriving the bias $k_{1,n}$ and variance $k_{2,n}$ terms, and to proving the asymptotic expansion (34).

A. Asymptotic Expansion: Bias

In this subsection, we carry out the asymptotic expansion of the RCU bound estimator (29) up to the term with the factor $k_{1,n}$ in (34) in order to characterize the bias of the estimator.

We start by computing the statistical mean $\mathrm{E}[\hat{\mathrm{rcu}}_{n,N_1,N_2}]$ of the estimator of the RCU bound (29). For importance-sampling purposes, pairs of channel input and output samples $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ are generated according to (27). For the analysis, however, it is more convenient to use the relationship between untilted and tilted probability densities (11) to compute the statistical mean $\mathrm{E}[\hat{\mathrm{rcu}}_{n,N_1,N_2}]$ according to the joint probability density

$$Q^n(\boldsymbol{x})W^n(\boldsymbol{y}|\boldsymbol{x}). \quad (35)$$

Defining the random variable $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{Y})$ with density $p_{\boldsymbol{Z}}(\boldsymbol{z})$ in (35), the statistical mean $\mathrm{E}[\hat{\mathrm{rcu}}_{n,N_1,N_2}]$ is given by

$$\mathrm{E}[\hat{\mathrm{rcu}}_{n,N_1,N_2}] = \mathrm{E}\big[\min\{1, (M-1)\hat{\mathrm{pep}}_{n,N_1}(\boldsymbol{Z})\}\big]. \quad (36)$$

This quantity does not depend on N_2 and depends on N_1 through the pairwise error probability estimator $\hat{\mathrm{pep}}_{n,N_1}$. We next expand (36) in a series of inverse powers of N_1.

Making use of the observation that for a random variable D and a uniform random variable U in $[0, 1]$, it holds that

$$\mathrm{E}[\min\{1, D\}] = \Pr[D \ge U], \quad (37)$$

we may thus rewrite (36) after taking logarithms in both sides of the inequality inside the probability as

$$\mathrm{E}[\hat{\mathrm{rcu}}_{n,N_1,N_2}] = \Pr\big[\log(M-1) + \log\hat{\mathrm{pep}}_{n,N_1}(\boldsymbol{Z}) \ge \log U\big]. \quad (38)$$

We continue our analysis by focusing on $\hat{\mathrm{pep}}_{n,N_1}$ for the optimum tilting parameter $\hat{\tau}_n(\boldsymbol{z})$ in (23). Defining

$$\epsilon = \frac{1}{\sqrt{N_1}} \quad (39)$$

and using (8), it follows that the importance-sampling estimator (22) converges as $\epsilon \to 0$ to

$$\hat{\mathrm{pep}}_{n,N_1}(\boldsymbol{z}) \xrightarrow[N_1\to\infty]{\mathrm{p}} \mathrm{pep}_n(\boldsymbol{z})\big(1 + \epsilon\sigma_{\mathrm{pep}}(\boldsymbol{z})\Theta_{N_1}\big), \quad (40)$$

where the sample estimation variance $\sigma^2_{\mathrm{pep}}(\boldsymbol{z})$ is given by

$$\sigma^2_{\mathrm{pep}}(\boldsymbol{z}) = \frac{\mathrm{E}\big[e^{\kappa_{n,\hat{\tau}_n}(\boldsymbol{z}) - \hat{\tau}_n(\boldsymbol{z})\ell_n(\boldsymbol{z},\overline{\boldsymbol{X}})} f_{\mathrm{pep}}(\boldsymbol{z}, \overline{\boldsymbol{X}})\big]}{\mathrm{E}\big[f_{\mathrm{pep}}(\boldsymbol{z}, \overline{\boldsymbol{X}})\big]^2} - 1, \quad (41)$$

and the random variable Θ_{N_1} has a conditional density function $p_{\Theta_{N_1}|\boldsymbol{Z}}(\theta|\boldsymbol{z})$ that converges uniformly to the standard normal density. The convergence is described by the degree-r Edgeworth expansion [15, Ch. XVI.2]

$$p_{\Theta_{N_1}|\boldsymbol{Z}}(\theta|\boldsymbol{z}) = p_\Theta(\theta) \cdot \left(1 + \sum_{k=3}^{r} \epsilon^{k-2} P_k(\theta, \boldsymbol{z})\right) + o(\epsilon^{r-2}). \tag{42}$$

In (42), $P_k(\theta, \boldsymbol{z})$ are degree-k polynomials in θ that depend on the moments of the summation terms in $\hat{\mathrm{pep}}_{n,N_1}(\boldsymbol{z})$ and on the Hermite polynomials [15, Ch. XVI.1] in θ. These moments are assumed to exist, or equivalently, we assume that the summation terms in $\hat{\mathrm{pep}}_{n,N_1}(\boldsymbol{z})$ have an integrable characteristic function under the distribution $Q^n(\overline{\boldsymbol{x}})$.

Taking logarithms in both sides of (40) and doing a Taylor expansion of order 2 in powers of ϵ, we find that $\log \hat{\mathrm{pep}}_{n,N_1}(\boldsymbol{z})$ converges in probability to

$$\log \hat{\mathrm{pep}}_{n,N_1}(\boldsymbol{z}) \xrightarrow[N_1\to\infty]{\mathrm{p}} \log \mathrm{pep}_n(\boldsymbol{z}) + \Xi_n(\epsilon, \boldsymbol{z}, \Theta_{N_1}) \tag{43}$$

since the logarithm is a continuous function, and where

$$\Xi_n(\epsilon, \boldsymbol{z}, \theta) = \epsilon\sigma_{\mathrm{pep}}(\boldsymbol{z})\theta - \frac{1}{2}\epsilon^2\sigma^2_{\mathrm{pep}}(\boldsymbol{z})\theta^2 + O(\epsilon^3). \tag{44}$$

Since the r.h.s. of (38) is a bounded and continuous function of $\hat{\mathrm{pep}}_{n,N_1}(\boldsymbol{z})$, the Mann-Wald continuity theorem [15, p. 431] allows us to substitute (43) into (38). Grouping the summands inside the probability, it proves convenient to define the random variable $A_n(\boldsymbol{Z}, U)$, sometimes written simply as A_n for the sake of compactness, as

$$A_n(\boldsymbol{Z}, U) = \log(M-1) + \log \mathrm{pep}_n(\boldsymbol{Z}) - \log U. \tag{45}$$

With this definition, we find that the statistical mean of the RCU bound estimator converges in probability to

$$\mathrm{E}[\hat{\mathrm{rcu}}_{n,N_1,N_2}] \xrightarrow[N_1\to\infty]{\mathrm{p}} \Pr\bigl[A_n(\boldsymbol{Z}, U) + \Xi_n(\epsilon, \boldsymbol{Z}, \Theta_{N_1}) \geq 0\bigr]. \tag{46}$$

The random variables A_n and Ξ_n have a joint density

$$p_{A_n}(a)p_{\Xi_n|A_n}(\xi|a) = p_{\boldsymbol{Z}}(\boldsymbol{z})p_U(u)p_{\Theta_{N_1}|\boldsymbol{Z}}(\theta|\boldsymbol{z}). \tag{47}$$

Suppose that $p_{A_n}(a)$ is a continuous differentiable density. Under mild assumptions on the joint density of A_n and Ξ_n given in (47), the asymptotic expansion as Ξ_n vanishes to zero [19, Th. 1] implies that the r.h.s. of (46) satisfies

$$\begin{aligned}\Pr[A_n + \Xi_n \geq 0] &= \Pr[A_n \geq 0] + p_{A_n}(0)\mathrm{E}[\Xi_n|A_n = 0] \\ &\quad + \frac{1}{2}p'_{A_n}(0)\mathrm{E}[\Xi_n^2|A_n = 0] + O(\Xi_n^3).\end{aligned} \tag{48}$$

First, notice that using (37) in the first term of the r.h.s. of (48) directly yields the RCU bound in (1), i.e.

$$\Pr\bigl[A_n \geq 0\bigr] = 1 - P_{A_n}(0) = \mathrm{rcu}_n, \tag{49}$$

where $P_{A_n}(a)$ denotes the cumulative distribution of A_n. We conclude that, as $\epsilon \to 0$ or equivalently $N_1 \to \infty$, the RCU bound estimator becomes asymptotically unbiased.

The properties of the Hermite polynomials [15, Ch. XVI.1] imply that

$$\int_{-\infty}^{\infty} p_\Theta(\theta)\theta P_3(\theta, \boldsymbol{z})\,\mathrm{d}\theta = 0. \tag{50}$$

Hence, using (50) and the fact that $p_\Theta(\theta)$ is the density of the standard normal random variable, it follows that under the conditional density expansion (42) for $r = 3$ the remaining expectations in the r.h.s. of (48) are asymptotically given by

$$\mathrm{E}[\Xi_n|A_n = 0] = -\frac{1}{2}\epsilon^2\eta_n + O(\epsilon^3) \tag{51}$$

and

$$\mathrm{E}[\Xi_n^2|A_n = 0] = \epsilon^2\eta_n + O(\epsilon^3), \tag{52}$$

where we defined

$$\eta_n = \mathrm{E}\bigl[\sigma^2_{\mathrm{pep}}(\boldsymbol{Z})|A_n = 0\bigr]. \tag{53}$$

Noting that $O(\Xi_n^3) = O(\epsilon^3)$ we obtain as $\epsilon \to 0$ that

$$\mathrm{E}[\hat{\mathrm{rcu}}_{n,N_1,N_2}] \xrightarrow[N_1\to\infty]{\mathrm{p}} \mathrm{rcu}_n(1 - k_{1,n}\epsilon^2), \tag{54}$$

where the parameter $k_{1,n}$ is given by

$$k_{1,n} = \frac{1}{2}\eta_n\frac{p_{A_n}(0) - p'_{A_n}(0)}{1 - P_{A_n}(0)}. \tag{55}$$

It remains to characterize asymptotically, as $n \to \infty$, the quantities $p'_{A_n}(0)$, $p_{A_n}(0)$, $P_{A_n}(0)$, and η_n. For memoryless channels, the variable A_n behaves asymptotically as the sum of n independent random variables. Standard asymptotic results in the approximation of the density of a strongly non-lattice random variable A_n [20, Ch. 2.2] show that as $n \to \infty$

$$\lim_{n\to\infty} \frac{p_{A_n}(a)}{\hat{p}_{A_n}(a)} = 1, \tag{56}$$

where $\hat{p}_{A_n}(a)$ is the saddlepoint approximation to the density of A_n at a. The saddlepoint $\hat{s}_n$ is the unique solution to

$$\hat{s}_n = \operatorname*{arg\,min}_{0\leq s\leq 1} \varphi_n(s) - sa, \tag{57}$$

where $\varphi_n(s)$ is the cumulant generating function of A_n; the random variable $\log U$ in (45) limits the region of convergence of $\varphi_n(s)$ to the $[0, 1]$ interval. A similar analysis can be carried out to show that the upper tail probability and the derivative of the density respectively satisfy

$$\lim_{n\to\infty} \frac{1 - P_{A_n}(a)}{\frac{1}{\hat{s}_n}\hat{p}_{A_n}(a)} = 1, \tag{58}$$

and

$$\lim_{n\to\infty} \frac{p'_{A_n}(a)}{-\hat{s}_n\,\hat{p}_{A_n}(a)} = 1, \tag{59}$$

for the same saddlepoint $\hat{s}_n$ in (57). Noting the relationship in (49) between the random variable A_n and rcu_n and using the asymptotic equivalence in (33), large-deviation theory results imply that

$$\lim_{n\to\infty} \frac{\varphi_n(s)}{\chi_n(s)} = 1 \tag{60}$$

and therefore

$$\lim_{n\to\infty} \frac{\hat{s}_n}{\hat{\rho}_n} = 1. \tag{61}$$

We recall that $\hat{\rho}_n$ is the optimal tilting parameter used in the importance-sampling estimator of the RCU bound. Combining the limits (56), (58) and (59), it follows that

$$\lim_{n\to\infty} \frac{1}{\hat{\rho}_n(1 - \hat{\rho}_n)}\frac{p_{A_n}(0) - p'_{A_n}(0)}{1 - P_{A_n}(0)} = 1. \tag{62}$$

Since $0 \le \hat{\rho}_n \le 1$ and $\eta_n \ge 0$, by combining (62) and (55) we conclude that $k_{1,n}$ is a non-negative term. For memoryless channels, the optimal tilting parameter satisfies $\hat{\rho}_n = O(1)$ [18, Ch. 5.6]. Therefore, together with (62), we have that

$$\frac{p_{A_n}(0) - p'_{A_n}(0)}{1 - P_{A_n}(0)} = O(1). \tag{63}$$

In the last part of this subsection, we study the asymptotics of η_n as $n \to \infty$. This quantity is a conditional expectation of the sample estimation variance $\sigma^2_{\text{pep}}(\boldsymbol{z})$ given in (41). Noting that $\mathrm{E}\big[f_{\text{pep}}(\boldsymbol{z}, \overline{\boldsymbol{X}})\big] = \text{pep}_n(\boldsymbol{z})$, we start by studying the asymptotics of $\text{pep}_n(\boldsymbol{z})$, whose square appears in the denominator of $\sigma^2_{\text{pep}}(\boldsymbol{z})$. Refined asymptotic results in the approximation of the tail probability of a sum of strongly non-lattice random variables [20, Eq. (2.2.6)] show that

$$\text{pep}_n(\boldsymbol{z}) = \frac{1}{\sqrt{n}} e^{\kappa_{n,\hat{\tau}_n}(\boldsymbol{z}) + O(1)} \tag{64}$$

uniformly in $\boldsymbol{z}$, where $\kappa_{n,\hat{\tau}_n}(\boldsymbol{z})$ is the cumulant generating function (18) evaluated at the optimal tilting parameter (23). The asymptotics of $\mathrm{E}\big[f_{\text{pep}}(\boldsymbol{z}, \overline{\boldsymbol{X}})\big]^2$ are thus obtained by squaring (64), namely

$$\mathrm{E}\big[f_{\text{pep}}(\boldsymbol{z}, \overline{\boldsymbol{X}})\big]^2 = \frac{1}{n} e^{2\kappa_{n,\hat{\tau}_n}(\boldsymbol{z}) + O(1)}. \tag{65}$$

For the remaining terms in $\sigma^2_{\text{pep}}(\boldsymbol{z})$, we note that the numerator is a tilted tail probability that can be written as

$$\Omega_n(\boldsymbol{z}) e^{\kappa_{n,\hat{\tau}_n}(\boldsymbol{z}) + \kappa_{n,-\hat{\tau}_n}(\boldsymbol{z})}, \tag{66}$$

where $\Omega_n(\boldsymbol{z})$ is a quantity given by

$$\Omega_n(\boldsymbol{z}) = \Pr\big[\ell_n(\boldsymbol{z}, \overline{\boldsymbol{X}}) \ge 0\big] \tag{67}$$

and computed under the tilted conditional density

$$\bar{p}_{\overline{\boldsymbol{X}}|\boldsymbol{Z}}(\overline{\boldsymbol{x}}|\boldsymbol{z}) = Q^n(\overline{\boldsymbol{x}}) e^{-\hat{\tau}_n(\boldsymbol{z})\ell_n(\boldsymbol{z},\overline{\boldsymbol{x}}) - \kappa_{n,-\hat{\tau}_n}(\boldsymbol{z})}. \tag{68}$$

Let $\beta_n(t, \boldsymbol{z})$ be the cumulant generating function of $\ell_n(\boldsymbol{z}, \overline{\boldsymbol{X}})$ under the density (68). After some manipulations, we obtain that $\beta_n(t, \boldsymbol{z})$ is related to $\kappa_{n,\tau}(\boldsymbol{z})$ as

$$\beta_n(t, \boldsymbol{z}) = \kappa_{n,t-\hat{\tau}_n}(\boldsymbol{z}) - \kappa_{n,-\hat{\tau}_n}(\boldsymbol{z}). \tag{69}$$

Finding the refined asymptotics for the upper-tail probability $\Omega_n(\boldsymbol{z})$ given by (67) involves computing the unique minimizer

$$\hat{t}_n(\boldsymbol{z}) = \arg\min_{t \ge 0} \beta_n(t, \boldsymbol{z}) \tag{70}$$

in the region of convergence of $\beta_n(t, \boldsymbol{z})$, given by $[0, \infty)$. From (69) and (23), it is given by

$$\hat{t}_n(\boldsymbol{z}) = 2\hat{\tau}_n(\boldsymbol{z}). \tag{71}$$

Therefore, as $n \to \infty$, we obtain using [20, Eq. (2.2.6)] that

$$\Omega_n(\boldsymbol{z}) = \frac{1}{\sqrt{n}} e^{\kappa_{n,\hat{\tau}_n}(\boldsymbol{z}) - \kappa_{n,-\hat{\tau}_n}(\boldsymbol{z}) + O(1)}. \tag{72}$$

Combining (72) with (66), we obtain that the numerator of the pairwise error probability sample variance satisfies

$$\mathrm{E}\big[e^{\kappa_{n,\hat{\tau}_n}(\boldsymbol{z}) - \hat{\tau}_n(\boldsymbol{z})\ell_n(\boldsymbol{z},\overline{\boldsymbol{X}})} f_{\text{pep}}(\boldsymbol{z}, \overline{\boldsymbol{X}})\big] = \frac{1}{\sqrt{n}} e^{2\kappa_{n,\hat{\tau}_n}(\boldsymbol{z}) + O(1)}. \tag{73}$$

Using the asymptotic equalities (73) and (65) into (41), we obtain that $\sigma^2_{\text{pep}}(\boldsymbol{z}) = O(\sqrt{n})$ uniformly in $\boldsymbol{z}$. Since η_n is the conditional expectation of $\sigma^2_{\text{pep}}(\boldsymbol{z})$ given by (53), we find that

$$\eta_n = O(\sqrt{n}). \tag{74}$$

Equations (63) and (74) used in the definition of the term $k_{1,n}$ in (55), together with (54), imply that for sufficiently large code length n and number of samples N_1, the expected value of $\hat{\text{rcu}}_{n,N_1,N_2}$ satisfies

$$\mathrm{E}[\hat{\text{rcu}}_{n,N_1,N_2}] \xrightarrow[N_1 \to \infty]{\text{p}} \text{rcu}_n \bigg(1 - \frac{k_{1,n}}{N_1}\bigg), \tag{75}$$

where $k_{1,n}$ is a parameter that grows with n as $O(\sqrt{n})$. This proves the asymptotic expansion (34) as $N_1 \to \infty$ of the RCU bound estimator (29) up to the term with the factor $k_{1,n}$.

B. Asymptotic Expansion: Variance

In this subsection, we extend the asymptotic expansion of the RCU bound estimator (29) to the term with the factor $k_{2,n}$ in (34) to characterize the variance of the estimator.

To start with, we note that the Lindeberg condition for the central-limit theorem [15, p. 262] is satisfied and that $\hat{\text{rcu}}_{n,N_1,N_2}$ is given by a sum of N_2 independent terms. Therefore, in an analogous manner to (8), we have

$$\hat{\text{rcu}}_{n,N_1,N_2} \xrightarrow[N_2 \to \infty]{\text{p}} \mathrm{E}[\hat{\text{rcu}}_{n,N_1,N_2}] \bigg(1 + \frac{\sigma_{\text{rcu}}}{\sqrt{N_2}}\Theta\bigg), \tag{76}$$

where now from (16) we have

$$\sigma^2_{\text{rcu}} = \frac{\mathrm{E}\big[e^{\chi_n(\hat{\rho}_n) - \hat{\rho}_n g_n(\boldsymbol{Z})} f^2_{\text{rcu}}(\boldsymbol{Z})\big]}{\mathrm{E}\big[f_{\text{rcu}}(\boldsymbol{Z})\big]^2} - 1. \tag{77}$$

As in the previous subsection, we have chosen the optimum tilting parameter $\hat{\rho}_n$ in (32).

We are interested in characterizing the asymptotics of (77) as $n \to \infty$. We first focus on the numerator in the r.h.s. of (77). Using (31) and that $\min[1, D]^2 = \min[1, D^2]$ for $D \ge 0$, the numerator can be rewritten as

$$\Psi_n e^{\chi_n(\hat{\rho}_n) + \chi_n(-\hat{\rho}_n)}, \tag{78}$$

where Ψ_n is a quantity given by

$$\Psi_n = \mathrm{E}\big[\min\big\{1, (M-1)^2 \hat{\text{pep}}_{n,N_1}(\boldsymbol{Z})^2\big\}\big] \tag{79}$$

computed under the tilted density

$$\bar{p}_{\boldsymbol{Z}}(\boldsymbol{z}) = p_{\boldsymbol{Z}}(\boldsymbol{z}) e^{-\hat{\rho}_n g_n(\boldsymbol{z}) - \chi_n(-\hat{\rho}_n)}. \tag{80}$$

Mimicking the steps followed to derive (46), we use the identity $\mathrm{E}[\min\{1, D\}] = \Pr[D \ge U]$ in (37), take logarithms on both sides of the inequality inside the probability, and substitute the expansion (43) in inverse powers of N_1 for $\log \hat{\text{pep}}_{n,N_1}$ to show that Ψ_n converges in probability as

$$\Psi_n \xrightarrow[N_1 \to \infty]{\text{p}} \Pr\big[B_n(\boldsymbol{Z}, U) + \Xi_n(\epsilon, \boldsymbol{Z}, \Theta_{N_1}) \ge 0\big], \tag{81}$$

where ϵ is given in (39), B_n is related to A_n in (45) as

$$B_n(\boldsymbol{Z}, U) = 2A_n(\boldsymbol{Z}, U) + \log U \tag{82}$$

and Ξ_n is given by (44). The tail probability in (81) is computed under the joint density

$$\bar{p}_{\boldsymbol{Z}}(\boldsymbol{z}) p_U(u) p_{\Theta_{N_1}|\boldsymbol{Z}}(\theta|\boldsymbol{z}), \tag{83}$$

where $p_{\Theta_{N_1}|\boldsymbol{Z}}(\theta|\boldsymbol{z})$ has the Edgeworth expansion (42).

Using the asymptotic expansion from [19, Th. 1], as we did in our analysis of (48), and using again the properties of the Hermite polynomials [15, Ch. XVI.1] in an analogous manner to the steps leading to (51) and (52), we find that as $\epsilon \to 0$

$$\Psi_n \xrightarrow[N_1\to\infty]{\mathrm{p}} \Pr\big[B_n \geq 0\big] + O(\epsilon^2). \tag{84}$$

We next turn our attention to the denominator in the r.h.s. of (77), which we identify as $\mathrm{E}[\hat{\mathrm{rcu}}_{n,N_1,N_2}]^2$. Taking squares in the expansion in (75) we find that the squared value of $\mathrm{E}[\hat{\mathrm{rcu}}_{n,N_1,N_2}]$ converges in probability as $N_1 \to \infty$ to rcu_n^2 as

$$\mathrm{E}[\hat{\mathrm{rcu}}_{n,N_1,N_2}]^2 \xrightarrow[N_1\to\infty]{\mathrm{p}} \mathrm{rcu}_n^2 + O(\epsilon^2). \tag{85}$$

Using (84), (85), and (78) in the formula for σ^2_{rcu} in (77), we obtain the asymptotic expansion

$$\sigma^2_{\mathrm{rcu}} \xrightarrow[N_1,N_2\to\infty]{\mathrm{p}} k_{2,n} + O(\epsilon^2), \tag{86}$$

where $k_{2,n}$ is given by

$$k_{2,n} = \frac{\Pr\big[B_n \geq 0\big]}{\mathrm{rcu}_n^2} e^{\chi_n(\hat{\rho}_n)+\chi_n(-\hat{\rho}_n)} - 1. \tag{87}$$

In our next step, we study the asymptotics of $k_{2,n}$ in (87) as $n \to \infty$. We start by studying rcu_n, whose square appears in the equation. Refined asymptotic results in the approximation of the tail probability of a sum of strongly non-lattice random variables for non-singular memoryless channels [11, Th. 1] show that the RCU bound (1) satisfies as $n \to \infty$

$$\mathrm{rcu}_n = \frac{1}{\sqrt{n}} e^{\chi_n(\hat{\rho}_n) - \frac{1}{2}\hat{\rho}_n \log n + O(1)}, \tag{88}$$

where $\chi_n(\rho)$ and $\hat{\rho}_n$ are respectively given by (26) and (32). Roughly speaking, equation (88) can be obtained by plugging the refined asymptotics of the pairwise error probability given by (64) into the definition of A_n in (45), computing the cumulant generating function of A_n using the definitions in (25) and (26), and employing [20, Eq. (2.2.6)] for the tail probability $\mathrm{rcu}_n = \Pr[A_n \geq 0]$. The additional term in the exponent of (88) is the contribution of the $\frac{1}{\sqrt{n}}$ prefactor in (64). The asymptotics of rcu_n^2 are obtained by squaring (88), namely

$$\mathrm{rcu}_n^2 = \frac{1}{n} e^{2\chi_n(\hat{\rho}_n) - \hat{\rho}_n \log n + O(1)}. \tag{89}$$

For the asymptotics of the tail probability $\Pr\big[B_n \geq 0\big]$, we exploit the identity (82) to obtain a similar expansion to (88). To this end, let $\xi_n(\lambda)$ denote the cumulant generating function of B_n, computed under the joint density (83), and let $\mathcal{L}_n$ be the corresponding region of convergence. Also, let $\hat{\lambda}_n$ denote the optimizer of the cumulant generating function,

$$\hat{\lambda}_n = \arg\min_{\lambda\in\mathcal{L}_n} \xi_n(\lambda). \tag{90}$$

The equivalent to (88) for the tail probability $\Pr[B_n \geq 0]$ involves plugging the refined asymptotics of the pairwise error probability given by (64) into the definitions of A_n and B_n in (45) and (82) respectively, and computing the cumulant generating function of B_n under the joint density (83). As a result, we obtain that

$$\xi_n(\lambda) = \chi_n(2\lambda - \hat{\rho}_n) - \chi_n(-\hat{\rho}_n) - \lambda \log n + O(1). \tag{91}$$

Similarly to (89), the $\log n$ term in (91) is the contribution of the $\frac{1}{\sqrt{n}}$ prefactor of (64). Using (32), it follows that

$$\hat{\lambda}_n = \hat{\rho}_n + o(1), \tag{92}$$

where $\hat{\rho}_n$ is the optimal tilting parameter used in the importance-sampling estimator of the RCU bound. Using (91) and (92) in [20, Eq. (2.2.6)], we find that

$$\Pr\big[B_n \geq 0\big] = \frac{1}{\sqrt{n}} e^{\chi_n(\hat{\rho}_n) - \chi_n(-\hat{\rho}_n) - \hat{\rho}_n \log n + O(1)}. \tag{93}$$

Finally, substituting (93) back in (87) the resulting expression together with (89), and simplifying the formula, we find that $k_{2,n}$ grows with n as $O(\sqrt{n})$. Combining this asymptotics with (86) and (75) into (76), we obtain (34).

V. Coded BPSK Modulation

A case of interest is the binary phase-shift keying (BPSK) modulation with symbol set $\mathcal{X} = \{-\sqrt{P}, +\sqrt{P}\}$, where P is a positive number describing an average power constraint. We study the achievable error probability of coded BPSK modulation over both the additive white Gaussian noise (AWGN) and the i.i.d. Rayleigh fading channels. Both cases are examples of continuous-output channels with continuous and differential conditional probability densities for which the required conditions on $f_{\mathrm{pep}}(\boldsymbol{x}, \boldsymbol{y}, \overline{\boldsymbol{X}})$ and $f_{\mathrm{rcu}}(\boldsymbol{X}, \boldsymbol{Y})$ for the theorem in Sec. IV are satisfied.

A. AWGN Channel

A codeword $\boldsymbol{x}$ is sent over the AWGN channel

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{w}, \tag{94}$$

where $\boldsymbol{w}$ is an i.i.d. real-valued zero-mean Gaussian noise with variance σ^2, and we assume perfect synchronization and a coherent receiver. From (94), it follows that

$$W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - x_i)^2}{2\sigma^2}}, \tag{95}$$

where x_i and y_i now denote the ith symbol of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. Thanks to the symmetry of BPSK, the input distribution $Q^n(\boldsymbol{x})$ that optimizes both the exponential decay (33) and the channel capacity C_{b} is the uniform distribution

$$Q^n(\boldsymbol{x}) = \frac{1}{2^n}. \tag{96}$$

For coding rates $R_{\mathrm{b}} < C_{\mathrm{b}}$, where C_{b} is the channel capacity [18, Eq. (2.2.8)], the RCU bound (1) decays exponentially fast since $\chi_n(\hat{\rho}_n) < 0$ for $0 < \hat{\rho}_n \leq 1$ [18, Eq. (5.6.27)]. For fixed P and σ^2, the coded average E_{b}/N_0 ratio is defined as

$$\frac{E_{\mathrm{b}}}{N_0} = \frac{P}{\sigma^2} \cdot \frac{1}{2R_{\mathrm{b}}}. \tag{97}$$

Monte-Carlo evaluation of the RCU bound to the error probability (1) would involve generating triplets $(\boldsymbol{x}, \boldsymbol{y}, \overline{\boldsymbol{x}})$

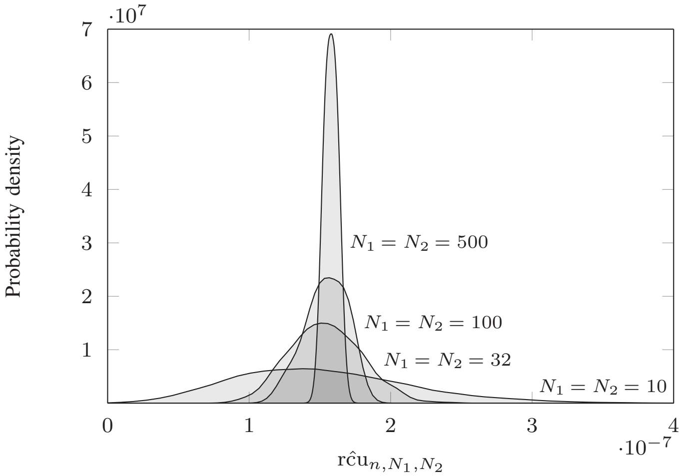

Fig. 1. Empirical probability density of $\hat{\mathrm{rcu}}_{n,N_1,N_2}$ over the AWGN channel, for code length $n = 128$, code rate $R_{\mathrm{b}} = 0.5$, and $E_{\mathrm{b}}/N_0 = 4$ dB.

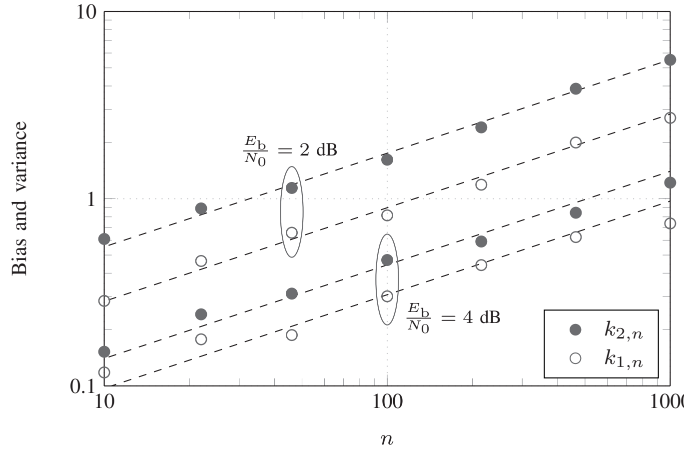

Fig. 2. Empirical bias $k_{1,n}$ and variance $k_{2,n}$ terms versus n over the AWGN channel, for code rate $R_{\mathrm{b}} = 0.5$. Dashed lines represent the $\sqrt{n}$ interpolation.

according to the joint distribution $Q^n(\boldsymbol{x})W^n(\boldsymbol{y}|\boldsymbol{x})Q^n(\overline{\boldsymbol{x}})$, and declaring a decoding error occurs if the codeword $\overline{\boldsymbol{x}}$ has a higher decoding metric than $\boldsymbol{x}$, i.e., if $W^n(\boldsymbol{y}|\overline{\boldsymbol{x}}) \geq W^n(\boldsymbol{y}|\boldsymbol{x})$.

With importance sampling, we need the distributions $\bar{P}^n_\tau(\overline{\boldsymbol{x}}|\boldsymbol{y})$ and $\bar{W}^n_\rho(\boldsymbol{y}|\boldsymbol{x})$. For a given pair of $\boldsymbol{x}$ and $\boldsymbol{y}$, the cumulant generating function $\kappa_\tau(\boldsymbol{x},\boldsymbol{y})$ is given from (18) as

$$\kappa_{n,\tau}(\boldsymbol{x},\boldsymbol{y}) = \sum_{i=1}^n \log\left(e^{-\frac{\tau(y_i-\sqrt{P})^2}{2\sigma^2}} + e^{-\frac{\tau(y_i+\sqrt{P})^2}{2\sigma^2}}\right) + \frac{\tau}{2\sigma^2}\sum_{i=1}^n (y_i - x_i)^2 - n\log 2. \quad (98)$$

Hence, the conditional distribution $\bar{P}^n_\tau(\overline{\boldsymbol{x}}|\boldsymbol{y})$ for the estimation of the pairwise error probability is given from (19) by the following product distribution

$$\bar{P}^n_\tau(\overline{\boldsymbol{x}}|\boldsymbol{y}) = \frac{1}{\mu_{n,\tau}(\boldsymbol{y})}\prod_{i=1}^n e^{-\frac{\tau(y_i-\overline{x}_i)^2}{2\sigma^2}}, \quad (99)$$

where $\mu_{n,\tau}(\boldsymbol{y})$ is a normalizing factor.

Upon substituting (96) and (95) into (26), the error exponent $\chi_n(\rho)$ is given by (100) at the bottom of the page. As a result, the tilted distribution in (27) involves $Q^n(\boldsymbol{x})$, the distribution in (96), and the tilted conditional distribution $\bar{W}^n_\rho(\boldsymbol{y}|\boldsymbol{x})$

$$\bar{W}^n_\rho(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{\mu^n_\rho}\prod_{i=1}^n e^{-\frac{(y_i-x_i)^2}{2(1+\rho)\sigma^2}} \cdot \left(e^{-\frac{(y_i-\sqrt{P})^2}{2(1+\rho)\sigma^2}} + e^{-\frac{(y_i+\sqrt{P})^2}{2(1+\rho)\sigma^2}}\right)^\rho, \quad (101)$$

with normalizing factor μ_ρ.

Importance sampling for coded BPSK in AWGN channels involves generating codewords with equiprobable BPSK symbols, and generating channel outputs according to the equivalent channel transition probability $\bar{W}^n_\rho(\boldsymbol{y}|\boldsymbol{x})$. Even though (101) is not a standard probability distribution, samples can be efficiently generated using the rejection method [21, Ch. II.3]. For a given channel output $\boldsymbol{y}$, the pairwise error probability is estimated by drawing n BPSK symbols from the binary product distribution (99).

We now proceed to consider several numerical examples. We first fix the code length to $n = 128$, the rate $R_{\mathrm{b}} = 0.5$ bits per channel use and the coded $E_{\mathrm{b}}/N_0 = 4$ dB, and show in Fig. 1 a smoothed-kernel density estimation [22] of the importance sampler $\hat{\mathrm{rcu}}_{n,N_1,N_2}$ for several sampling sizes N_1 and N_2. We do not consider the Monte Carlo estimation, since it would require at least $N_2 \approx 10^9$ samples. As we increase both N_1 and N_2, the both the estimation bias and variance are reduced, in accordance with (34).

To validate the growth of the bias and variance terms of (34) as n increases, we obtain empirical estimates of $k_{1,n}$ and $k_{2,n}$. For every value of n, we set $N_2 = 1000$, run Algorithm 1 for several values of N_1, and find the best interpolated value of $k_{1,n}$ in accordance to (34) neglecting the $k_{2,n}$ term. For $k_{1,n}$, we run the algorithm at $N_1 = 1000$ and $N_2 = 1$ several times, compute the sample variance and normalized it according to (34). Fig. 2, plotting $k_{1,n}$ and $k_{2,n}$ against the codeword length n for a rate $R_{\mathrm{b}} = 0.5$ bits per channel use at several coded E_{b}/N_0 ratios, confirms that both terms grow as $\sqrt{n}$.

In the remainder in this subsection, we take $N_1 = N_2 = 500$. As we detailed in Sec. II, the error probability of the optimal code is, at most, the RCU bound. Similarly, using sphere-packing arguments, Shannon established that every code transmitted over the AWGN channel is lower bounded by [1, Eq. (15)]. Therefore, the error probability of good binary codes must lie between the RCU bound and the Shannon lower bound. This region is shaded in gray in the following figures.

In Fig. 3, we depict the coded error probability region with high performance codes [23] for code length $n = 128$, and

$$\chi_n(\rho) = n\rho\log(M-1) - \frac{n}{2}\log(2\pi\sigma^2) - n(1+\rho)\log 2 + n\log\int_{-\infty}^{\infty}\left(e^{-\frac{(y-\sqrt{P})^2}{2(1+\rho)\sigma^2}} + e^{-\frac{(y+\sqrt{P})^2}{2(1+\rho)\sigma^2}}\right)^{1+\rho}\mathrm{d}y \quad (100)$$

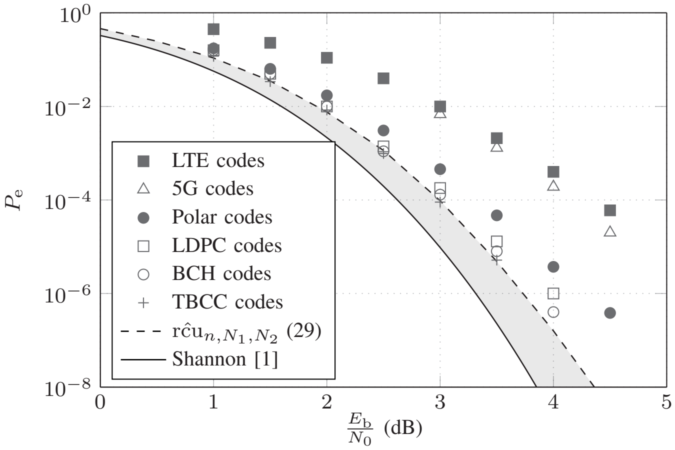

Fig. 3. Error probability versus E_b/N_0 over the AWGN channel, for code length $n = 128$, code rate $R_b = 0.5$, and $N_1 = N_2 = 500$ samples.

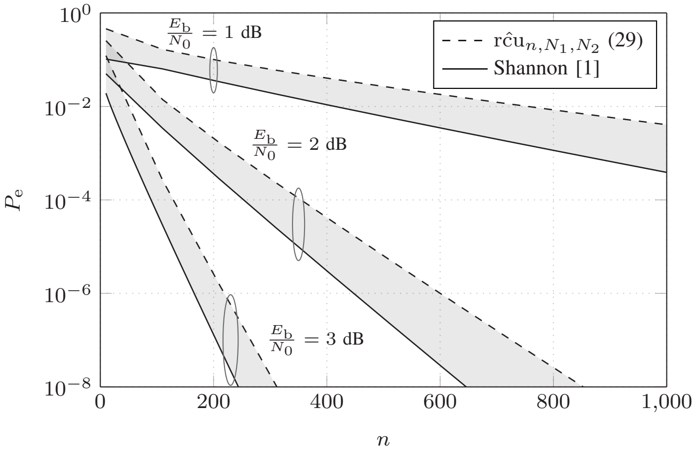

Fig. 4. Error probability versus n over the AWGN channel, for code rate $R_b = 0.5$, $N_1 = N_2 = 500$ samples, and several values of E_b/N_0.

rate $R_b = 0.5$, versus the E_b/N_0 ratio. We include examples of LTE codes (turbo code with 8 states), 5G codes (NR BG 2), polar codes (CRC-7 and $L = 32$), LDPC codes ($\mathbb{F}_{256}$ and OSD $t = 4$), BCH codes (extended OSD $t = 4$) and TBCC codes ($m = 14$). For such a small code length, TBCC codes exhibit good performance, even though there must exist codes with a performance, at most, the RCU bound. Similar conclusions can be obtained from Fig. 4, plotting the coded error probability versus the block length n for varying E_b/N_0.

B. Rayleigh Channel

A second case of interest is the i.i.d. Rayleigh fading channel. For a transmitted codeword $\boldsymbol{x} = (x_1, \dots, x_n) \in \{-\sqrt{P}, +\sqrt{P}\}^n$ the received sequence $\boldsymbol{y} = (y_1, \dots, y_n)$ is

$$y_i = h_i x_i + w_i, \tag{102}$$

where $\boldsymbol{w} = (w_1, \dots, w_n)$ is an i.i.d. real-valued zero-mean Gaussian noise with variance σ^2, and $\boldsymbol{h} = (h_1, \dots, h_n)$ is i.i.d. Rayleigh distributed with density

$$p^n(\boldsymbol{h}) = \prod_{i=1}^n 2h_i e^{-h_i^2} \mathbb{1}\{h_i \geq 0\}. \tag{103}$$

Since $\mathrm{E}[h_i^2] = 1$, the coded average E_b/N_0 ratio remains (97).

Assuming perfect channel state information at the receiver (CSIR), the RCU bound to the error probability, averaged over the i.i.d. Rayleigh fading, is given by the expression

$$\mathrm{rcu}_n = \int p^n(\boldsymbol{h}) Q^n(\boldsymbol{x}) W^n(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h}) \cdot \\ \cdot \min\big\{1, (M-1)\mathrm{pep}_n(\boldsymbol{x},\boldsymbol{y},\boldsymbol{h})\big\} \, \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{h}, \tag{104}$$

where $Q^n(\boldsymbol{x})$ and $p^n(\boldsymbol{h})$ are respectively given by (96) and (103), the channel conditional probability density function $W^n(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h})$ is now given by

$$W^n(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - h_i x_i)^2}{2\sigma^2}}. \tag{105}$$

In (104), the pairwise error probability for a fixed transmitted codeword $\boldsymbol{x}$, received sequence $\boldsymbol{y}$ and fading realization $\boldsymbol{h}$, denoted as $\mathrm{pep}_n(\boldsymbol{x},\boldsymbol{y},\boldsymbol{h})$, reads

$$\mathrm{pep}_n(\boldsymbol{x},\boldsymbol{y},\boldsymbol{h}) = \int Q^n(\overline{\boldsymbol{x}}) \mathbb{1}\{\ell_n(\boldsymbol{x},\boldsymbol{y},\boldsymbol{h},\overline{\boldsymbol{x}}) \geq 0\} \, \mathrm{d}\overline{\boldsymbol{x}} \tag{106}$$

for the log-likelihood ratio function

$$\ell_n(\boldsymbol{x},\boldsymbol{y},\boldsymbol{h},\overline{\boldsymbol{x}}) = \log \frac{W^n(\boldsymbol{y}|\overline{\boldsymbol{x}},\boldsymbol{h})}{W^n(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h})}. \tag{107}$$

The importance-sampling estimator of the RCU bound to the error probability (104) is analogous to the case without fading and involves generating the quadruplet $(\boldsymbol{x},\boldsymbol{y},\overline{\boldsymbol{x}},\boldsymbol{h})$ according to the exponentially-tilted joint distribution

$$p^n(\boldsymbol{h}) Q^n(\boldsymbol{x}) \bar{W}_\rho^n(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h}) \bar{P}_\tau^n(\overline{\boldsymbol{x}}|\boldsymbol{y},\boldsymbol{h}), \tag{108}$$

with the conditional distribution $\bar{P}_\tau^n(\overline{\boldsymbol{x}}|\boldsymbol{y},\boldsymbol{h})$ for the pairwise error probability given by the product distribution

$$\bar{P}_\tau^n(\overline{\boldsymbol{x}}|\boldsymbol{y},\boldsymbol{h}) = \frac{1}{\mu_{n,\tau}(\boldsymbol{y},\boldsymbol{h})} \prod_{i=1}^n e^{-\frac{\tau(y_i - h_i \overline{x}_i)^2}{2\sigma^2}}, \tag{109}$$

where $\mu_{n,\tau}(\boldsymbol{y},\boldsymbol{h})$ is the corresponding normalizing factor, and $\bar{W}_\rho^n(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h})$ is the tilted channel conditional distribution

$$\bar{W}_\rho^n(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h}) = \frac{1}{\mu_\rho^n} \prod_{i=1}^n e^{-\frac{(y_i - h_i x_i)^2}{2(1+\rho)\sigma^2}} \cdot \\ \cdot \left(e^{-\frac{(y_i - h_i\sqrt{P})^2}{2(1+\rho)\sigma^2}} + e^{-\frac{(y_i + h_i\sqrt{P})^2}{2(1+\rho)\sigma^2}} \right)^\rho \tag{110}$$

with μ_ρ the corresponding normalizing factor. Observe that the fading realization $\boldsymbol{h}$ is generated according to the original Rayleigh distribution (103) as, similarly to the case without fading, the exponential tilting of the outer expectation of the RCU (104) only affects the conditional distribution of the received sequence $\boldsymbol{y}$, while leaving the remaining distributions unaltered. The tilting parameters τ and ρ needed in (109) and (110) are optimally selected as the minimizers

$$\hat{\tau}_n(\boldsymbol{x},\boldsymbol{y},\boldsymbol{h}) = \arg\min_{\tau \geq 0} \kappa_{n,\tau}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{h}) \tag{111}$$

and

$$\hat{\rho}_n = \arg\min_{0 \leq \rho \leq 1} \chi_n(\rho), \tag{112}$$

Algorithm 2: Importance-sampling estimate of the RCU bound for BPSK modulation over the i.i.d. Rayleigh fading channel

Input: $Q^n(\boldsymbol{x})$, $W^n(\boldsymbol{y}|\boldsymbol{x})$, $p^n(\boldsymbol{h})$, n, R_b, N_2 and N_1
Output: $\hat{\mathrm{rcu}}$
calculate $M = \lfloor 2^{nR_\mathrm{b}} \rfloor$;
calculate $\chi_n(\rho)$ from (114);
select $\rho \leftarrow \arg\min_{0\le\rho\le1} \chi_n(\rho)$;
find $\bar{W}^n_\rho(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h})$ from (110);
$\alpha \leftarrow 0$;
for $i = 1; i \le N_2$ **do**
  generate $(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{h}_i)$ according to $p^n(\boldsymbol{h})Q^n(\boldsymbol{x})\bar{W}^n_\rho(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{h})$;
  compute $\kappa_{n,\tau}(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{h}_i)$ from (113);
  select $\tau \leftarrow \arg\min_{\tau\ge0} \kappa_{n,\tau}(\boldsymbol{x}_i, \boldsymbol{y}_i)$;
  find $\bar{P}^n_\tau(\overline{\boldsymbol{x}}|\boldsymbol{y},\boldsymbol{h})$ from (109);
  $\gamma \leftarrow 0$;
  for $j = 1; j \le N_1$ **do**
    generate $\overline{\boldsymbol{x}}_j$ according to $\bar{P}^n_\tau(\overline{\boldsymbol{x}}|\boldsymbol{y}_i,\boldsymbol{h}_i)$;
    $\gamma \leftarrow \gamma + \frac{1}{N_1} e^{-\tau\cdot\ell_n(\boldsymbol{x}_i,\boldsymbol{y}_i,\boldsymbol{h}_i,\overline{\boldsymbol{x}}_j)} \mathbb{1}\{\ell_n(\boldsymbol{x}_i,\boldsymbol{y}_i,\boldsymbol{h}_i,\overline{\boldsymbol{x}}_j) \ge 0\}$;
  end
  $\hat{\mathrm{pep}} \leftarrow \gamma \cdot e^{\kappa_{n,\tau}(\boldsymbol{x}_i,\boldsymbol{y}_i,\boldsymbol{h}_i)}$;
  $\alpha \leftarrow \alpha + \frac{1}{N_2} e^{-\rho\cdot g_n(\boldsymbol{x}_i,\boldsymbol{y}_i,\boldsymbol{h}_i)} \min\{1, (M-1)\hat{\mathrm{pep}}\}$;
end
$\hat{\mathrm{rcu}} \leftarrow \alpha \cdot e^{\chi_n(\rho)}$;
return $\hat{\mathrm{rcu}}$;

where the cumulant generating functions $\kappa_{n,\tau}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{h})$ and $\chi_n(\rho)$ are respectively given as

$$\kappa_{n,\tau}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{h}) = \sum_{i=1}^n \log\left(e^{-\frac{\tau(y_i-h_i\sqrt{P})^2}{2\sigma^2}} + e^{-\frac{\tau(y_i+h_i\sqrt{P})^2}{2\sigma^2}}\right) + \frac{\tau}{2\sigma^2}\sum_{i=1}^n (y_i - h_i x_i)^2 - n\log 2 \quad (113)$$

and by (114) at the bottom of the page. The nested importance-sampling estimator for the i.i.d. Rayleigh fading channel is summarized in pseudo-code in Algorithm 2 on the top of the page, where by extension we defined

$$g_n(\boldsymbol{x},\boldsymbol{y},\boldsymbol{h}) = \log(M-1) + \kappa_{n,\frac{1}{1+\rho}}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{h}). \quad (115)$$

Similarly to the AWGN case, we include an error probability lower bound, the improved sphere-packing bound [24, Th. 3.1] valid for discrete-input continuous-output symmetric channels such as the i.i.d. Rayleigh fading channel described in (102).

We take $N_1 = N_2 = 500$ to estimate the RCU bound (104) using the importance-sampling Algorithm 2. In Fig. 5, we depict the coded error probability region against the coded E_b/N_0 for rate $R_\mathrm{b} = 0.5$, several codeword lengths n. For $n = 128$, we obtain that RCU bound with i.i.d. Rayleigh fading incurs a 3 dB loss in E_b/N_0 compared to the AWGN channel in Fig. 3. We also observe that the gap to the error probability lower bound is higher in the presence of i.i.d. Rayleigh fading. By setting $n = 1024$, we finally show

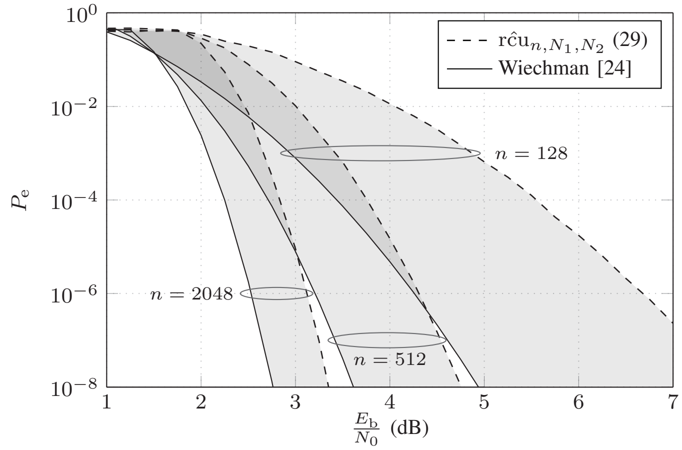

Fig. 5. Error probability versus E_b/N_0 over the i.i.d. Rayleigh channel, for code rate $R_\mathrm{b} = 0.5$, $N_1 = N_2 = 500$ samples, and several code lengths n.

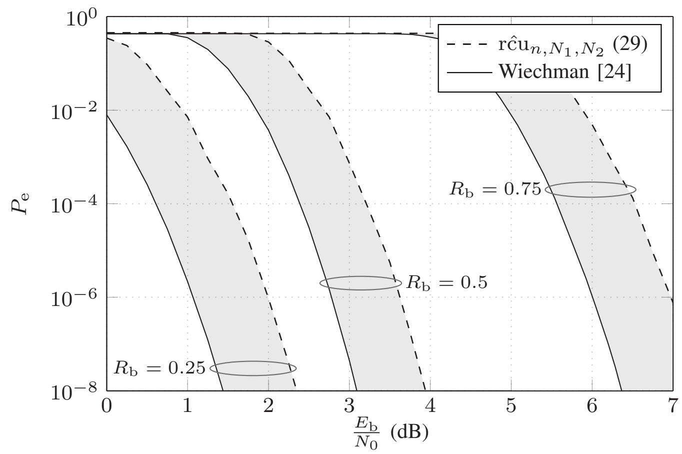

Fig. 6. Error probability versus E_b/N_0 over the i.i.d. Rayleigh channel, for code length $n = 1024$, $N_1 = N_2 = 500$ samples, and several code rates R_b.

in Fig. 2 the coded error probability region against the coded E_b/N_0 for several rates above below channel capacity.

VI. Conclusion

In this paper, we have presented an importance-sampling technique to estimate the achievable error probability, using random coding arguments, for the transmission of coded data over a continuous-output channel. Exploiting the exponential decay of the error probability, we found closed-form expressions for the optimal tilted distributions needed to generate the samples of the two nested estimators involved. We studied the convergence in probability of the estimator and illustrated the transmission of the coded BPSK modulation over the AWGN and i.i.d. Rayleigh fading channels. Our study gives an estimate of the bias and variance of the estimator in terms of the number of samples and the code length, thereby providing guidance on the dimensioning of the nested estimators.

$$\chi_n(\rho) = n\rho\log(M-1) - \frac{n}{2}\log(2\pi\sigma^2) - n(1+\rho)\log 2 + n\log\int_0^\infty\int_{-\infty}^\infty 2he^{-h^2}\left(e^{-\frac{(y-h\sqrt{P})^2}{2(1+\rho)\sigma^2}} + e^{-\frac{(y+h\sqrt{P})^2}{2(1+\rho)\sigma^2}}\right)^{1+\rho} \mathrm{d}y\mathrm{d}h \quad (114)$$

References

[1] 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.

[2] 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.

[3] B. Davis, "An improved importance sampling method for digital communication system simulations," *IEEE Trans. Commun.*, vol. 34, no. 7, pp. 715–719, Jul. 1986.

[4] D. Lu and K. Yao, "Improved importance sampling technique for efficient simulation of digital communication systems," *IEEE J. Sel. Areas Commun.*, vol. 6, no. 1, pp. 67–75, Jan. 1988.

[5] R. J. Wolfe, M. C. Jeruchim, and P. M. Hahn, "On optimum and suboptimum biasing procedures for importance sampling in communication simulation," *IEEE Trans. Commun.*, vol. 38, no. 5, pp. 639–647, May 1990.

[6] J.-C. Chen, D. Lu, J. S. Sadowsky, and K. Yao, "On importance sampling in digital communications—Part I. Fundamentals," *IEEE J. Sel. Areas Commun.*, vol. 11, no. 3, pp. 289–299, Apr. 1993.

[7] E. Cavus, C. L. Haymes, and B. Daneshrad, "Low BER performance estimation of LDPC codes via application of importance sampling to trapping sets," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 1886–1888, 2009.

[8] 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.

[9] 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.

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

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

[12] 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.

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

[14] 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. Science and Sys. (CISS)*, Princeton, NJ, Mar. 2018.

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

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

[17] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer Berlin Heidelberg, 2009.

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

[19] D. R. Cox and N. Reid, "Approximations to noncentral distributions," *Canadian J. of Stat.*, vol. 15, no. 2, pp. 105–114, 1987.

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

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

[22] V. A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," *Theory of Prob. and its Appl.*, vol. 14, pp. 153–158, 1969.

[23] M. C. Coşkun, G. Durisi, T. Jerkovits, G. Liva, W. Ryan, B. Stein, and F. Steiner, "Efficient error-correcting codes in the short blocklength regime," *Phys. Commun.*, vol. 34, pp. 66–79, 2019.

[24] 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.