

Non-parametric kernel-based bit error probability estimation in digital communication systems:

An estimator for soft coded QAM BER computation

Poda Pasteur^{1,2} - Saoudi Samir² - Chonavel Thierry² - Guilloud Frédéric²
- Tapsoba Théodore¹

1 Ecole supérieure d'Informatique, Université Nazi BONI, Bobo-Dioulasso, Burkina Faso.
pasteur.poda@univ-bobo.bf, theo_tapsoba@univ-ouaga.bf.

2 IMT Atlantique, Lab-STICC, UBL, 29238 Brest, France.
{samir.saoudi, thierry.chonavel, frederic.guilloud}@imt-atlantique.fr.



ABSTRACT. The standard Monte Carlo estimations of rare events probabilities suffer from too much computational time. To make estimations faster, kernel-based estimators proved to be more efficient for binary systems whilst appearing to be more suitable in situations where the Probability Density Function (pdf) of the samples is unknown. We propose a kernel-based Bit Error Probability (BEP) estimator for coded M -ary Quadrature Amplitude Modulation (QAM) systems. We defined soft real bits upon which an Epanechnikov kernel-based estimator is designed. Simulation results showed, compared to the standard Monte Carlo simulation technique, accurate, reliable and efficient BEP estimates for 4-QAM and 16-QAM symbols transmissions over the additive white Gaussian noise channel and over a frequency-selective Rayleigh fading channel.

RÉSUMÉ. Les estimations de probabilités d'événements rares par la méthode classique de Monte Carlo souffrent de trop de temps de calculs. Des estimateurs à noyau se sont montrés plus efficaces sur des systèmes binaires en même temps qu'ils paraissent mieux adaptés aux situations où la fonction de densité de probabilité est inconnue. Nous proposons un estimateur de Probabilité d'Erreur Bit (PEB) à noyau pour les systèmes M -aires codés de Modulations d'Amplitude en Quadrature (MAQ). Nous avons défini des bits souples à valeurs réelles à partir desquels un estimateur à noyau d'Epanechnikov est conçu. Les simulations ont montré, par rapport à la méthode Monte Carlo, des estimées de PEB précises, fiables et efficaces pour des transmissions MAQ-4 et MAQ-16 sur canaux à bruit additif blanc Gaussien et à évanouissements de Rayleigh sélectif en fréquence.

KEYWORDS : Bit error probability, Bit error rate, Probability density function, Monte Carlo method, Kernel estimator.

MOTS-CLÉS : Probabilité d'erreur binaire, Taux d'erreur binaire, Fonction de densité de probabilité, Méthode Monte Carlo, Estimateur à noyau.



1. Introduction

In digital communications, the Bit Error Probability (BEP) is usually used for the performance characterisation of a communication system. It is generally determined in the form of an estimate commonly called Bit Error Rate (BER). Several approaches of the BEP estimation have been studied. A few attempts to analytically estimate the BEP were reported in [1] and [2]. However, simulation-based techniques of the BEP estimation have been the most investigated, surely because of the increasing complexity, of the contemporary and emerging digital communication systems, that renders impossible the derivation of a closed-form solution of the BEP estimate. A simulation-based technique of the BEP estimation that does not depend on the complexity of the digital communication system is that technique which proceeds with the errors counting in the received bits sequence and then determines the BEP estimate, i.e. the BER, as the ratio of the number of the observed errors over the total number of the transmitted bits. That technique, called the classical (or standard) Monte Carlo (MC) method, is a universal technique as it does not depend on the digital communication system. For this reason, it is commonly used as the baseline for the other methods. This universal technique which is moreover straightforward unfortunately suffers from its high computational cost. Indeed, it is known as computationally the most costly of the methods [3]. Samples of very large size may be required by this technique to perform BEP estimates with a given accuracy. This is further obvious when small BER values have to be computed. To mathematically exhibit this disadvantage of the classical MC method, one can easily be provided with a lower bound of the sample size N_{mc} in function of the BEP p_e and the normalised error ε of the estimate \hat{p}_e : $N_{mc} > 1/\varepsilon^2 p_e$ (see e.g.: [4]). The normalised error ε is defined as the ratio of the standard deviation of \hat{p}_e over p_e . Therefore, to estimate a BEP $p_e = 10^{-4}$, the classical MC method with standard deviation smaller than $0.1p_e$, should at least be run with a total number of transmitted data equal to 10^6 .

To overcome the problem of the classical MC method efficiency, a class of simulation-based techniques, referred to as variance-reducing techniques, have been developed during the 1970s with the goal to reduce the sample size. These variance-reducing techniques have been discussed in details in [3]. More specifically, they include the importance sampling (also called modified Monte-Carlo) technique [5], the extreme-value theory method [6], the tail extrapolation method [7] and the quasi-analytical estimation technique [3]. The importance sampling method has been the most successful and the most investigated of these variance-reducing techniques. Some recent investigations on this technique dealt with very low BER performance estimation for coded modulations. For instance, Cavus et al. [8] developed an approach that combines the importance sampling technique with trapping sets. Their approach allowed a successful simulation of the performance of Low Density Parity Check (LDPC) codes over an Additive White Gaussian Noise (AWGN) channel at BERs smaller than 10^{-6} .

More recently, new simulation-based techniques of the BEP estimation have been developed [4], [9]. They rely on non-parametric Probability Density Function (pdf) estimations. Soft observations are used to estimate the pdf that is generally assumed unknown. Very few of the non-parametric estimation techniques are based on Gaussian mixture models [9]. Gaussian mixture models rely on the assumption that the pdf of the received observations is a mixture of Gaussians. The parameters of each Gaussian, i.e. the mean, the variance and the *a priori* probabilities, together with the number of Gaussians are estimated by simulation. The expression of the BEP estimate is then derived as a function

of the estimated parameters. On the contrary, the non-parametric kernel-based estimation techniques have been more investigated. They rely on kernel density estimators [10]. The estimation solely relies on the observed samples and there is no need to estimate the number of Gaussian components as for the Gaussian mixture. Several kernel-based BEP estimators built around the Gaussian kernel function have shown to reach good performance on the uncoded binary-input Gaussian channel. This is the case of the kernel-based BEP estimator for soft BERs computation [4], where the estimator performance was analysed for Code Division Multiple Access (CDMA) schemes. Efficient and accurate BEP estimates were reported. This is also the case with the study reported in [12]. A kernel-based estimator for soft BERs computation demonstrated that kernel-based estimators of the BEP can perform well in an unsupervised manner, i.e. without requiring the transmitted data to be known. Always based on the Gaussian kernel function, a maximum likelihood-based smoothing parameter optimisation was studied [13] for the kernel estimators of BEP. Illustrations of the effectiveness of the maximum likelihood-based smoothing parameter selection were made for binary coded transmission schemes involving Turbo and LDPC codes over CDMA systems. For the first time, a work addressing the issue of efficient kernel-based BEP estimation for M -ary transmissions schemes was reported in [14]. This paper extends the work in [14] by thoroughly studying the kernel-based BEP estimation for coded M -ary Quadrature Amplitude Modulations (QAM) transmissions schemes. In the proposed estimator, the samples used for the estimation are soft bits that are sampled from the output of the soft error-correcting code decoder. Beyond the AWGN channels, frequency-selective Rayleigh fading channels are targeted. Under this framework, the issue of kernel-based efficient BEP estimation becomes more challenging with respect to the previous contexts presented above. Firstly, shifting from binary real constellations to M -ary complex constellations involves the estimation of complex pdfs. Secondly, when frequency-selective Rayleigh fading channels are considered, the pdf of the soft observations to be estimated loses its Gaussian nature and finding an appropriate smoothing parameter for the kernel is not straightforward.

The remainder of this paper is organised as follows: in Section 2 a theoretical formulation of the BEP followed by MC-based and kernel-based approaches to estimate it are described; in Section 3 a description of the different features of the proposed kernel-based BEP estimator is provided; in Section 4 the simulation framework and analysis parameters are presented and simulation results are discussed. Section 5 concludes the paper.

2. The bit error rate, an estimate of the bit error probability

2.1. The bit error probability

In order to theoretically formulate the BEP, let us consider a coded digital communication system that operates with Quadrature Amplitude Modulation (QAM) schemes. At the transmitter-end, a signal containing coded M -QAM waveforms is transmitted over a noisy channel. M is the QAM constellation size. The transmitted signal corresponds to a bit sequence $(b_j)_{1 \leq j \leq N}$. At the receiver-end, more precisely at the output of the channel decoder, we assume being provided with independent and identically distributed soft real bits $(X_j)_{1 \leq j \leq N}$ such that the hard decision consists of a bit $\hat{b}_j = 0$ (resp. $\hat{b}_j = 1$) when $X_j < 0$ (resp. $X_j > 0$). Let X denote the univariate real random variable that

describes the soft bits $(X_j)_{1 \leq j \leq N}$ and let $f_X^{(0)}(x)$ (resp. $f_X^{(1)}(x)$) be the conditional pdf of X_j conditional to $b_j = 0$ (resp. $b_j = 1$). The BEP can be stated as:

$$\begin{aligned} p_e &= \Pr[X > 0, b_j = 0] + \Pr[X < 0, b_j = 1] \\ &= \Pr[X > 0 | b_j = 0] \Pr[b_j = 0] + \Pr[X < 0 | b_j = 1] \Pr[b_j = 1] \\ &= \pi_0 \int_0^{+\infty} f_X^{(0)}(x) dx + \pi_1 \int_{-\infty}^0 f_X^{(1)}(x) dx, \end{aligned} \quad (1)$$

where $\pi_0 = \Pr[b_j = 0]$ and $\pi_1 = \Pr[b_j = 1]$ are the *a priori* probabilities of the bits values “0” and “1” respectively. The BER is an estimate of the BEP and can be computed based on simulations. Several simulation-based approaches of the BER computation exist.

2.2. The Monte Carlo methods of BEP estimation

We denote by Monte Carlo methods, the classical MC method and the modified Monte-Carlo simulation technique also called importance sampling. To introduce the classical MC method of BEP estimation, let us consider that the transmitted bits sequence $(b_j)_{1 \leq j \leq N}$ are all binary “0” valued. Consequently, Eq. (1) of the BEP can be written as

$$p_e = \int_0^{+\infty} f_X^{(0)}(x) dx, \quad (2)$$

$$= \int_{-\infty}^{+\infty} f_X^{(0)}(x) I(x > 0) dx, \quad (3)$$

$$= \mathbb{E}[I(X > 0) | b_j = 0, 1 \leq j \leq N]. \quad (4)$$

where $\mathbb{E}[\cdot]$ is the mathematical expectation operator and I is such as,

$$I(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

A natural and straightforward estimator \hat{p}_e of p_e is therefore given by the method of moments as

$$\hat{p}_e = \frac{1}{N} \sum_{j=1}^N I(X_j^{(0)} > 0), \quad (5)$$

where $X_j^{(0)}$, $1 \leq j \leq N$, are the realisations of the random variable X conditioned to $b_j = 0$, $1 \leq j \leq N$. Using Eq. (5), the BEP estimate, i.e. the BER can be computed as the number of errors over the total number of observations: this way of calculating the BER defines the classical MC method.

As about the modified Monte-Carlo simulation technique, it has been introduced by Shanmugam et al. based on the following principle [5]: if the regions of the receiver input (i.e. the channel output) which contribute to an important event that is of interest are known, then the input distribution is biased (i.e. modified) in such a way that more samples are taken from the important regions. In this technique, the statistical properties of the receiver input are assumed to be known and the BEP to be estimated is given by

$$p_e = \int_{-\infty}^{\infty} h(x) f_X^{(0)}(x) dx, \quad (6)$$

where $h(x)$ is equal to 1 if x is greater than the hard decision threshold and 0 otherwise, thus generalising Eq. (3). A modified pdf $f_X^{(0)*}(x)$ can be used to change the distribution of the input samples, yielding:

$$p_e = \int_{-\infty}^{\infty} h^*(x) f_X^{(0)*}(x) dx, \quad (7)$$

where $h^*(x) = B(x)h(x)$ and $B(x) = f_X^{(0)}(x)/f_X^{(0)*}(x)$. The BER is finally computed as the BEP estimate \hat{p}_e given by

$$\hat{p}_e = \frac{1}{N} \sum_{j=1}^N h^*(x_j), \quad (8)$$

where x_1, \dots, x_N are the output sequence corresponding to the input sequence obtained by the modified distribution.

2.3. The kernel method of BEP estimation

In the kernel-based BEP estimation technique, the marginal conditional pdfs $f_X^{(0)}(x)$ and $f_X^{(1)}(x)$ are estimated as follows:

$$\begin{cases} \hat{f}_X^{(0)}(x) = \frac{1}{n_0} \sum_{j=1}^{n_0} \frac{1}{h_0} K\left(\frac{x-X_j}{h_0}\right) \\ \hat{f}_X^{(1)}(x) = \frac{1}{n_1} \sum_{j=1}^{n_1} \frac{1}{h_1} K\left(\frac{x-X_j}{h_1}\right), \end{cases} \quad (9)$$

where K is any even regular pdf with zero mean and unit variance called the kernel, n_0 (resp. n_1) is the cardinality of the subset of the soft observations among X_1, \dots, X_N which are likely to be decoded into a binary “0” (resp. “1”) bit value and h_0 (resp. h_1) is a parameter called smoothing parameter (or bandwidth) that depends on the soft observations X_1, \dots, X_{n_0} (resp. X_1, \dots, X_{n_1}). Let us notice that the sample size N is such as $n_0 + n_1 = N$. Then, the estimate \hat{p}_e of the BEP p_e of Eq. (1) can be expressed as follows,

$$\hat{p}_e = \hat{\pi}_0 \int_0^{+\infty} \hat{f}_X^{(0)}(x) dx + \hat{\pi}_1 \int_{-\infty}^0 \hat{f}_X^{(1)}(x) dx. \quad (10)$$

where $\hat{\pi}_0$ and $\hat{\pi}_1$ are the estimates of π_0 and π_1 respectively. To compute Eq. (10), the prior determination of the kernel K and the smoothing parameters h_0 and h_1 is necessary.

The choice of the kernel function K is related to the density function under estimation. As an example, for the estimation of a power limited process, the kernel should have a finite variance. Also, whenever the observed samples are distributed over a large scale, distributions with an infinite support (e.g., Gaussian distribution) should be chosen. However, distributions such as Epanechnikov, Uniform or Quartic that have finite support should be selected to model the kernel function when the observed samples are bounded. Once the kernel function K is selected, then comes the selection of the smoothing parameters h_0 and h_1 .

The design of the smoothing parameters is indeed a major issue since it is crucial to the performance of the estimator, especially in terms of accuracy. It has been demonstrated in [15] that if the smoothing parameter $h \rightarrow 0$ (i.e., either $h_0 \rightarrow 0$ or $h_1 \rightarrow 0$) when the

sample size $N \rightarrow \infty$ (which means that the quantities $n_0 \rightarrow \infty$ and $n_1 \rightarrow \infty$), then the estimator is asymptotically unbiased. The methods for optimally selecting the smoothing parameter are all based on the minimisation of the estimation error. A common way of measuring the estimation error is the Mean Integrated Squared Error (MISE), a function of the smoothing parameter h , given by [16],

$$MISE(h) = \int \mathbb{E} \left[\hat{f}_X(x) - f_X(x) \right]^2 dx. \quad (11)$$

Under standard technical assumptions (see e.g. [17]), the MISE is asymptotically (i.e., as the sample size $N \rightarrow \infty$) approximated by the Asymptotic Mean Integrated Squared Error (AMISE),

$$AMISE(h) = n_\zeta^{-1} h^{-1} \int K^2(x) dx + h^4 \int f_X''(x)^2 dx \left(\int x^2 K(x)/2 dx \right)^2, \quad (12)$$

where $f_X''(x)$ is the second derivative of the pdf $f_X(x)$, $(n_\zeta)_{\zeta \in \{0,1\}}$ is for designating either n_0 or n_1 , and K is as mentioned above the kernel or more precisely the kernel function. The quantities n_0 and n_1 , already defined above, have to be also seen as the amounts of the soft observations that are used to estimate the marginal conditional pdfs $f_X^{(0)}(x)$ and $f_X^{(1)}(x)$ respectively. The minimisation of (12) with respect to h gives the AMISE-based optimal smoothing parameters as follows,

$$h_\zeta^* = \left[\frac{\int K^2(x) dx}{\int f_X''(x)^2 dx \left(\int x^2 K(x) dx \right)^2} \right]^{1/5} n_\zeta^{-1/5}. \quad (13)$$

Clearly, the constraint in Eq. (13) is the prior knowledge of the second derivative $f_X''(x)$ of the targeted distribution $f_X(x)$ which is of course unknown and searched for. A multitude of techniques [18] that provide a way to bypass this constraint include plug-in methods [19], cross-validation techniques [20], [21] and variable kernel density estimation methods [22], [23]. Unfortunately, none of these techniques has yet been considered as the best in every situation [18]; hence the difficulty to find a universal optimal smoothing parameter. In practice, a convenient technique is to replace the unknown pdf f_X by a reference distribution with mean and variance matching those of the data. In the literature, the Gaussian distribution is a popular choice for the kernel function so that many designs regarding the choice of the optimal smoothing parameter are available. By selecting a Gaussian kernel function K and using a Gaussian reference distribution (i.e. the unknown pdf f_X is assumed Gaussian), the AMISE-based optimal smoothing parameters have been derived from Eq. (13) as follows [24],

$$h_\zeta^* = (4/3)^{1/5} \hat{\sigma} n_\zeta^{-1/5}, \quad (14)$$

where $\hat{\sigma}$ is the estimated standard deviation obtained through the soft observations X_1, \dots, X_n . The expression of Eq. (14) is indeed an estimate of the AMISE-based optimal smoothing parameter and is called the rule-of-thumb bandwidth.

A practical problem with the rule-of-thumb bandwidth is its sensitivity to outliers [28]. A single outlier may cause a too large estimate of σ and hence implies a too large bandwidth. It is noted in [25] that the standard variance estimator is not appropriate for the non-Gaussian densities. So, a more robust estimator is obtained from the interquartile range which is a measure that indicates the range over which the 50% most centered

samples are spread. Let Q_Z^k denote the k -th quartile of the random variable Z , defined by $\Pr(Z < Q_Z^k) = \frac{k}{4}$ for $k \in \{1, 2, 3\}$. Let also R_Z denote the interquartile range of the random variable Z as $R_Z = Q_Z^3 - Q_Z^1$. Still keeping the assumption that the true pdf f_X is Gaussian, we have $X \sim \mathcal{N}(\mu, \sigma^2)$ and $W = (X - \mu)/\sigma \sim \mathcal{N}(0, 1)$. Hence asymptotically [28],

$$\begin{aligned} R_X &= Q_X^3 - Q_X^1, \\ &= (\mu - \sigma Q_W^3) - (\mu - \sigma Q_W^1), \\ &= \sigma(0.67 - (-0.67)), \\ &= 1.34\sigma, \end{aligned} \tag{15}$$

Thus σ can be estimated by $\hat{\sigma} = R_X/1.34$. By plugging $\hat{\sigma} = R_X/1.34$ into Eq. (14), we obtain the following version of the rule-of-thumb bandwidth,

$$h_\zeta^* = 0.79R_X n_\zeta^{-1/5}. \tag{16}$$

By combining Eq. (14) and Eq. (16), a more robust estimate of σ is $\min(\hat{\sigma}, R_X/1.34)$ [26], [17] and hence the following robust version of the rule-of-thumb optimal bandwidth is:

$$h_\zeta^* = (4/3)^{1/5} \min(\hat{\sigma}, R_X/1.34) n_\zeta^{-1/5}. \tag{17}$$

3. The proposed kernel-based BEP estimator

Let us consider a digital communication system with multi-carrier transmissions of M -ary QAM waveforms over frequency-selective Rayleigh fading channels. Such a communication system model involves inter-symbol and inter-carrier interferences and subsequently complex receiver schemes are required. Nowadays, advanced receivers characterise the contemporary and emerging wireless communication systems. As the nature of the pdf of the soft observations depends not only on the type of the receiver but also on the channel model, it is very difficult to find the exact parametric model that describes the received distribution. In these conditions, a method as the modified Monte-Carlo that assumes the pdf of the received soft bits to be known cannot be applied. However, the kernel approach is well justified. So, the objective of this work is to compute soft coded BERs of M -ary QAM transmissions schemes using the kernel approach of the BEP estimation. The digital communication system under consideration also includes a channel codec (encoder/decoder). The soft coded BER is the BER that is computed from soft bits taken at the output of the channel decoder. The soft bits are the soft outputs of the channel decoder that normally serve for the hard decision making. We assume that the channel decoder operates with soft inputs in the form of LLR values and can deliver soft outputs in the form of LLR values. As M -QAM waveforms of alphabet $\{s_1, \dots, s_M\}$ are transmitted, the channel outputs are M -ary waveforms (soft symbols). To provide the channel decoder with appropriate inputs, a symbol-to-bit soft demapping (see 3.1) has to be done in order to convert the channel outputs which are soft symbols into soft bits in the form of LLR values. In addition, suited soft bits in the form of real values have to be given at the input of the BEP estimator. In 3.2, we define these soft real bits and establish in 3.3 the key equation for the BEP estimate computation.

3.1. The symbol-to-bit soft demapping

At the transmitter-end, before the modulation, bunches of k coded bits are grouped $b = (b_1, \dots, b_k)$, $k = \log_2(M)$, and mapped onto constellation points $s \in \chi = \{s_1, \dots, s_M\}$ for transmission. Note that the transmitted bits are uniformly distributed, and thus so are the transmitted symbols. At the receiver-end, the soft bit for each coded bit is calculated based on the received signal r . At the i -th sample period, the observed symbol is r_i and the symbol-to-bit soft demapping [27] results in the computation of k soft bits in the form of LLR $(L_j)_{1 \leq j \leq k}$. The j -th soft bit L_j is then given by

$$L_j = \log \left(\frac{\Pr[b_j = 1|r_i]}{\Pr[b_j = 0|r_i]} \right), \quad (18)$$

and can be rewritten as

$$L_j = \log \left(\frac{\sum_{s \in \chi_j^{(1)}} \Pr[s|r_i]}{\sum_{s \in \chi_j^{(0)}} \Pr[s|r_i]} \right), \quad (19)$$

where $\chi_j^{(0)}$ (resp. $\chi_j^{(1)}$) denotes the signal subset of χ with the j -th bit equal to “0” (resp. “1”). Using Bayes’ rule and since the symbols s_1, \dots, s_M are uniformly distributed ($\Pr[s = s_i] = 1/M, \forall i \in \{1, \dots, M\}$), we get

$$L_j = \log \left(\frac{\sum_{s \in \chi_j^{(1)}} p[r_i|s]}{\sum_{s \in \chi_j^{(0)}} p[r_i|s]} \right). \quad (20)$$

For a fading channel, $r_i = g_i s + \eta_i$, where $s \in \chi$, η_i a complex-valued realisation of the AWGN $\eta \sim \mathcal{N}(0, \sigma^2)$ and g_i , a complex-valued element of the channel matrix (a vector) g . Thus, $p[r_i|s] \sim \mathcal{N}(g_i s, \sigma^2)$ and the conventional Max-Log-MAP demapper allows L_j , $1 \leq j \leq k$, of Eq. (20) to be derived in function of r_i , g_i and s as follows:

$$L_j \approx -\frac{1}{\sigma^2} \left[\min_{s \in \chi_j^{(1)}} |r_i - g_i s|^2 - \min_{s \in \chi_j^{(0)}} |r_i - g_i s|^2 \right]. \quad (21)$$

Let us remark that in the case of the AWGN channel, $r_i = s + \eta_i$ and therefore g_i is a scalar always equal to 1.

3.2. The proposed kernel-based estimator inputs

The outputs of the channel decoder are soft bits in the form of LLR values L_j . However, we define the inputs of the kernel-based estimator to be soft bits $(X_j)_{1 \leq j \leq N}$ in the form of soft real values $+1$ or -1 . The soft bit X_j is given by:

$$X_j = \Pr[b_j = 1|r_i] - \Pr[b_j = 0|r_i]. \quad (22)$$

From Eq. (18) and $\Pr[b_j = 1|r_i] + \Pr[b_j = 0|r_i] = 1$, we derive the expressions of $\Pr[b_j = 1|r_i]$ and $\Pr[b_j = 0|r_i]$ in function of the soft LLR values L_j as follows:

$$\begin{cases} \Pr[b_j = 1|r_i] = e^{L_j} / (1 + e^{L_j}) \\ \Pr[b_j = 0|r_i] = 1 / (1 + e^{L_j}) \end{cases}. \quad (23)$$

From Eq. (23) and Eq. (22), we derive the expression of the soft bit X_j as a function of the channel decoder output L_j :

$$X_j = \frac{1 - e^{-L_j}}{1 + e^{-L_j}}. \quad (24)$$

The soft bits $(X_j)_{1 \leq j \leq N}$ are used by the proposed kernel-based estimator to derive the coded M -QAM BER values. Otherwise, if needed, soft symbol error probabilities can be estimated using soft M -ary symbols. The definition of the soft M -ary symbols is given in Appendix A.

3.3. The proposed kernel-based estimator equation

As defined in Eq. (24), the soft bits $(X_j)_{1 \leq j \leq N}$ are bounded with values in the interval $[-1, +1]$. So, the selection of the popular Gaussian kernel function cannot be strongly justified against the selection of finite support kernel functions. Among the multiple finite support distributions that are candidates for the kernel function selection, the Epanechnikov distribution is the simplest one in a computational point of view. For these reasons, we select the Epanechnikov distribution as the kernel function, i.e., $K(x) = \frac{3}{4} (1 - x^2) I(|x| \leq 1)$. Then it can be checked that the kernel estimator with bandwidth h will be restricted to interval $[-1 - h, 1 + h]$. Since optimally chosen h remains much smaller than 1 for large samples, we can consider that numerically the support constraint for the distribution of X is satisfied when using the Epanechnikov kernel. Therefore, we need to find the corresponding smoothing parameter h_{Epa}^* that approximates well the AMISE-based optimal smoothing parameter of Eq. (13). According to the related literature on bandwidth selection based on the Gaussian kernel, the robust rule-of-thumb optimal smoothing parameter of Eq. (17) is the best approximation of the AMISE-based optimal smoothing parameter. We then determine h_{Epa}^* based on the concept of canonical bandwidths as follows [28],

$$h_{\text{Epa}}^* = \frac{\delta_{\text{Epa}}}{\delta_{\text{Gau}}} h_{\zeta}^*, \quad (25)$$

where h_{ζ}^* is the robust optimal smoothing of Eq. (17), $\delta_{\text{Gau}} \approx (1/4)^{1/10} = 0.7764$ is the canonical bandwidth of the Gaussian kernel and $\delta_{\text{Epa}} \approx 15^{1/5} = 1.7188$ is the canonical bandwidth of the Epanechnikov kernel. Let us notice that in Eq. (25), we represent through h_{Epa}^* two smoothing parameters insofar as h_{ζ}^* is for either h_0^* or h_1^* . The canonical bandwidths are closely related to the rescaling of a kernel function called the canonical kernel [29]. The principle of the canonical kernel is to uncouple the problems of choosing h and K and the idea for separating these choices is to find the canonical bandwidths so that the AMISE will be asymptotically equal for different kernels (for instance, the Gaussian and the Epanechnikov kernels).

Now, the two key parameters that completely define the proposed kernel-based BEP estimator are known: the kernel function $K(x) = \frac{3}{4} (1 - x^2) I(|x| \leq 1)$ and the smoothing parameter as given in Eq. (25). The expressions of the two marginal conditional pdfs $\hat{f}_X^{(0)}(x)$ and $\hat{f}_X^{(1)}(x)$ can be derived from Eq. (9). Then, Eq. (10) of the BEP estimate can be rewritten as follows,

$$\hat{p}_e = \hat{\pi}_0 \int_0^{+\infty} \frac{1}{n_0 h_0^*} \sum_{j=1}^{n_0} K\left(\frac{x - X_j}{h_0^*}\right) dx + \hat{\pi}_1 \int_{-\infty}^0 \frac{1}{n_1 h_1^*} \sum_{j=1}^{n_1} K\left(\frac{x - X_j}{h_1^*}\right) dx, \quad (26)$$

where h_0^* (resp. h_1^*), computed based on Eq. (25), is the selected optimal smoothing parameter which will govern the accuracy of the estimation of $\hat{f}_X^{(0)}(x)$ (resp. $\hat{f}_X^{(1)}(x)$). By the means of some mathematical transformations that are detailed in Appendix B, Eq. (26) leads to the following convenient equation that will serve for the BER computation,

$$\hat{p}_e = \frac{\hat{\pi}_0 L_0}{n_0} + \frac{\hat{\pi}_1 L_1}{n_1} + \frac{3\hat{\pi}_0}{4n_0} \left\{ \sum_{\substack{|\alpha_j| \leq 1, \\ 1 \leq j \leq n_0}} \left(\frac{2}{3} - \alpha_j + \frac{\alpha_j^3}{3} \right) \right\} + \frac{3\hat{\pi}_1}{4n_1} \left\{ \sum_{\substack{|\beta_j| \leq 1, \\ 1 \leq j \leq n_1}} \left(\frac{2}{3} + \beta_j - \frac{\beta_j^3}{3} \right) \right\}, \quad (27)$$

where $\alpha_j = -X_j/h_0^*$, $\beta_j = -X_j/h_1^*$, L_0 (resp. L_1) is the cardinality of the subset of $(\alpha_j)_{1 \leq j \leq n_0}$ (resp. $(\beta_j)_{1 \leq j \leq n_1}$) which are less than -1 (resp. greater than 1).

The optimal smoothing parameters h_0^* and h_1^* of Eq. (27) are derived based on Eq. (25) and Eq. (17) as follows,

$$\begin{cases} h_0^* = 2.3449 \min(\hat{\sigma}, R_X/1.34) n_0^{-1/5} \\ h_1^* = 2.3449 \min(\hat{\sigma}, R_X/1.34) n_1^{-1/5} \end{cases}. \quad (28)$$

Then, the soft coded BERs are computed by simulations using Eq. (27). The simulations have been run for M -ary QAM transmissions over the AWGN channel and over a frequency-selective Rayleigh fading channel. In the next Section, we deal with the simulation results of the proposed kernel-based BEP estimator.

4. Simulation results

4.1. Channel models and performance analysis tools

The proposed kernel-based BEP estimator has been simulated on a single-carrier M -ary QAM transmission scheme over the AWGN channel and also on a multi-carrier M -ary QAM transmission scheme over a frequency-selective Rayleigh fading channel. Gray-coded 4-QAM and 16-QAM constellations were considered. The frequency-selective Rayleigh fading channel was ten taps long with a sample period of $12.8 \mu s$, a maximum Doppler shift set to $8 Hz$ and average taps gains given in *watts* by the vector $[0.0616 \ 0.4813 \ 0.1511 \ 0.0320 \ 0.1323 \ 0.0205 \ 0.0079 \ 0.0778 \ 0.0166 \ 0.0188]$ [30], [31]. To mitigate inter-symbol and inter-carrier interferences, a cyclic prefix Orthogonal Frequency Division Multiplexing (OFDM) technique was implemented [32]. The length of the Cyclic Prefix was set to 9 and the number of OFDM sub-carriers set to 128. A 128-point FFT (Fast Fourier Transform) was performed. The Channel codec was a 4/7-rate LDPC code with a Gallager-based parity check matrix built to be of rank 15. The number of iterations while decoding the LDPC code was set to 10 in the simulations involving the AWGN channel. It was set to 30 in the simulations regarding the frequency-selective Rayleigh fading channel.

To analyse the performance of the proposed kernel-based estimator for soft coded BER computation, we used different analysis parameters, namely the absolute bias, the Confidence Interval (CI), the sample size saving and the CPU time. We defined the absolute bias as $|\mathbb{E}[\hat{p}_e] - p_e|$. To compute the absolute bias, a certain number of the BEP

estimates \hat{p}_e has to be processed. As the theoretical expression of p_e is not available, a reference value has to be determined too. So, benchmark values have been computed for each value of the information bit energy to noise power spectral density ratio denoted E_b/N_0 . The benchmark values were computed using the classical MC simulations whilst handling a threshold on the number of observed errors. The threshold depends on the channel model. In the case of the AWGN channel, at least 100 errors have been observed. As about the frequency-selective Rayleigh fading channel, a threshold of 1 000 errors was set. The absolute bias allows the analysis of the estimator accuracy. The smaller it is, the more accurate is the estimation. Visually, excellent accuracy performance results in plotted data points of the BEP estimates values which are pointwise consistent with their corresponding benchmark data points.

The accuracy alone is not sufficient to evaluate with thoroughness the proposed BEP estimator performance. For instance, a good accuracy result of an estimate can be the fact of a lucky coincidence while suffering from reliability. The CI is intended to measure how reliable the estimator is. The smaller the CI is, the more reliable the estimate is. Let l be the size of a set of the BEP estimates \hat{p}_e . The $(1 - \alpha)$ CI is the interval $I = [\hat{p}_e - \epsilon, \hat{p}_e + \epsilon]$ such that the probability of having the true value of p_e inside I is equal to $(1 - \alpha)$, that is to say, $\Pr(\hat{p}_e - \epsilon \leq p_e \leq \hat{p}_e + \epsilon) = 1 - \alpha$, i.e.,

$$\Pr\left(\frac{-\epsilon}{\hat{\sigma}_l/\sqrt{l}} \leq T \leq \frac{\epsilon}{\hat{\sigma}_l/\sqrt{l}}\right) = 1 - \alpha,$$

where $T = (p_e - \hat{p}_e)/(\hat{\sigma}_l/\sqrt{l})$ and $\hat{\sigma}_l^2$ is a chi square distribution. Thus ϵ can be estimated thanks to the reciprocal cumulative density function of T denoted $F_T^{-1}(x)$ for $0 \leq x \leq 1$ by:

$$\epsilon = \frac{\hat{\sigma}_l}{\sqrt{l}} F_T^{-1}(\alpha/2).$$

Assuming that \hat{p}_e follows a Gaussian distribution with mean p_e and standard deviation σ , the random variable T follows a Student's t -distribution with $(l - 1)$ degrees of freedom. When s is high, T is well approximated by a Gaussian distribution. However, due to complexity issues, we only simulated a limited number of BEP estimates ($l = 21$), which justifies the use of the Student's t -distribution with $(l - 1)$ degrees of freedom for T . Hereafter, a CI for $\alpha = 0.05$ has been chosen.

The efficiency of the proposed estimator is evaluated thanks to the sample size reduction with respect to the classical MC method. If N_K is the sample size required by the proposed estimator to achieve a given performance (of accuracy and reliability) and N_{mc} , the sample size required by the classical MC to achieve equal (or almost equal) performance, then the sample size saving is defined by the ratio N_{mc}/N_K . Therefore, the greater (than 1) the sample size saving is, more efficient is the proposed estimator. The sample size saving is closely linked to the computational cost engendered by the BEP estimate computation. To evaluate this aspect of the estimator efficiency, the CPU time has been computed by simulation. The CPU time shows to how much is the energy consumption. It tells us about the inherent computational cost of the estimator.

4.2. Numerical results and discussions

In this Subsection, we analysed through simulation results the performance of the proposed kernel-based estimator with respect to the analysis parameters defined above. Since the importance sampling method requires (see Eq. (8)) the pdf of the receiver input to be

known in advance, a fair comparison with our proposed estimator is not possible without resorting to additional techniques. For this reason, we only made comparisons with respect to the universal classical MC method. Simulations of the proposed kernel-based estimator for soft coded BERs computation have been performed both in the AWGN channel and in a frequency-selective Rayleigh fading channel. The Rayleigh fading channels are more characteristic of the contemporary and emerging digital communication systems. They are moreover more challenging than the AWGN channels regarding BEP estimations. For these reasons, we'll roughly evaluate the performance of the proposed estimator as far as the AWGN channel is concerned. However, most of the proposed estimator performance analysis will focus on simulation results involving the frequency-selective Rayleigh fading channel.

Insights of the proposed estimator performance over the AWGN channel

To analyse the accuracy of the proposed estimator when performing over the AWGN channel, Figure 1 illustrates the curves of coded BERs in function of E_b/N_0 . Simulations with sample sizes N_K (see Table 1 and Table 2) in the form of 10^p , $p \in \{3, 4, 5, 6\}$, have been run to cover coded BER values from 10^{-1} down to 10^{-5} . The blue dashed curves in Figure 1 are related to the theoretical uncoded BERs and are given for illustration. The data points of the proposed estimator are borne by the green curves with diamond mark at each data point. The curves drawn with red solid lines are for the benchmarks. As Figure 1 lets see, the green curve with diamond mark at each data point and the red solid one are combined in a unique curve as well for the pair (of curves) related to 4-QAM as for that related to 16-QAM. This demonstrated that the proposed estimator achieved very accurate estimates. To quantify the achieved accuracy, Table 1 (for 4-QAM) and Table 2 (for 16-QAM) report numerical data that show how small are the corresponding absolute biases compared to the associated benchmarks.

As about the reliability of the proposed estimator, we show in Table 1 and Table 2 the intervals I representing the achieved CIs. Small CIs, each containing its corresponding estimate value, have been observed. In the case of 4-QAM where true coded BERs values from 1.1×10^{-1} down to 4.4×10^{-6} are estimated, the smallest of the CIs is $[0.94p_e, 1.06p_e]$ and the largest one is $[0.54p_e, 1.46p_e]$. In the case of 16-QAM, the smallest of the CIs is $[0.97p_e, 1.03p_e]$ and the largest of all is $[0.73p_e, 1.27p_e]$ for true coded BER values from 1.9×10^{-1} down to 1.1×10^{-5} . These numerical data show that the proposed estimator is not only accurate but also reliable when performing over the AWGN channel.

We are now interested in knowing roughly whether or not the proposed estimator yielded sample size savings with respect to the classical MC simulation technique. Using simulations, we have been able to note that sample size savings can be obtained from the proposed estimator. Numerical values of the sample size savings N_{mc}/N_K are given in the last columns of Table 1 and Table 2.

Performance analysis over a frequency-selective Rayleigh fading channel

Based on the plots in Figure 2, the accuracy of the proposed estimator can be analysed in both cases of 4-QAM and 16-QAM M -ary symbols transmissions over the frequency-selective Rayleigh fading channel as specified above in 4.1. The blue dashed curves in the figure are those of the theoretical uncoded BERs. They are drawn for illustration. The benchmarks related to both transmission schemes of 4-QAM and 16-QAM are given by the curves drawn in red solid line. The green curves with diamond mark at each data point bear the soft coded BER values that describe the performance of the proposed

Table 1. 4-QAM numerical results (AWGN channel)

E_b/N_0	Benchmark	Bias	I	N_K	N_{mc}/N_K
00 dB	1.1×10^{-1}	0.03×10^{-1}	$[0.94p_e, 1.06p_e]$	10^3	1.8
01 dB	6.7×10^{-2}	0.22×10^{-2}	$[0.90p_e, 1.10p_e]$	10^3	1.6
02 dB	3.1×10^{-2}	0.22×10^{-2}	$[0.82p_e, 1.18p_e]$	10^3	1.9
03 dB	1.2×10^{-2}	0.11×10^{-2}	$[0.93p_e, 1.07p_e]$	10^4	2.0
04 dB	3.0×10^{-3}	0.18×10^{-3}	$[0.81p_e, 1.19p_e]$	10^4	2.6
05 dB	4.7×10^{-4}	0.30×10^{-4}	$[0.89p_e, 1.11p_e]$	10^5	2.5
06 dB	4.9×10^{-5}	0.38×10^{-5}	$[0.66p_e, 1.34p_e]$	10^5	2.5
07 dB	4.4×10^{-6}	0.09×10^{-6}	$[0.54p_e, 1.46p_e]$	10^6	> 5.0

kernel-based estimator. We can see that the soft coded BER data points are very close to their corresponding benchmarks. So, we conclude that the accuracy of the kernel-based estimator is satisfactory. The observed accuracy is the fact of the absolute biases achieved by the estimator. The associated numerical values are given by the 3rd column of Table 3 and Table 4 regarding the 4-QAM scheme and 16-QAM respectively. We can notice that they are quite negligible compared to their respective benchmarks.

To evaluate the performance of the proposed estimator in terms of reliability, let us focus on the 4th columns of Table 3 and Table 4. These columns contain numerical data of the intervals I that describe the CIs achieved by the proposed estimator. An estimator is considered to be acceptable if the estimated value of p_e lies in an interval which is smaller than $I_{max} = [0, 3p_e]$ with a probability of 0.95 [5]. In this paper, we considered $I_{max} = [0.50p_e, 1.50p_e]$ as the largest acceptable interval of the CI. From BER values in the neighbourhood of 2×10^{-1} down to the lower in the neighbourhood of 3×10^{-4} , smaller than I_{max} values of the CIs have been observed for reasonable sample sizes N_K (see 5th columns of Table 3 and Table 4). The largest of the observed CIs values is

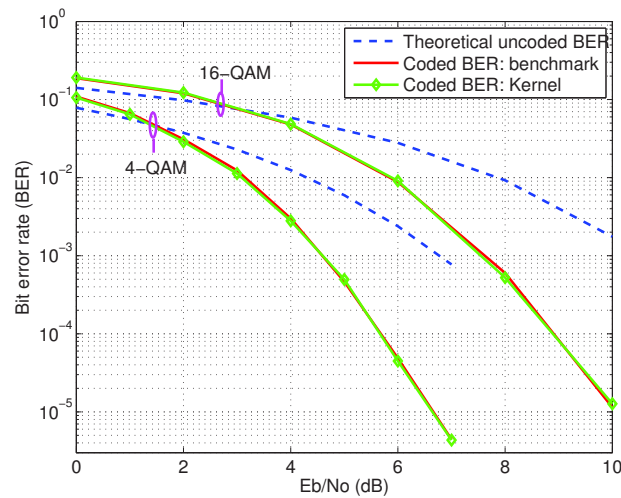

Figure 1. Performance of the proposed estimator over the AWGN channel

Table 2. 16-QAM numerical results (AWGN channel)

E_b/N_0	Benchmark	Bias	I	N_K	N_{mc}/N_K
00 dB	1.9×10^{-1}	0.03×10^{-1}	$[0.97p_e, 1.03p_e]$	10^3	2.1
02 dB	1.2×10^{-1}	0.02×10^{-1}	$[0.95p_e, 1.05p_e]$	10^3	1.8
04 dB	4.8×10^{-2}	0.05×10^{-2}	$[0.96p_e, 1.04p_e]$	10^4	1.6
06 dB	8.8×10^{-3}	0.26×10^{-3}	$[0.91p_e, 1.09p_e]$	10^4	1.8
08 dB	6.0×10^{-4}	0.77×10^{-4}	$[0.89p_e, 1.11p_e]$	10^5	1.6
10 dB	1.1×10^{-5}	0.13×10^{-5}	$[0.73p_e, 1.27p_e]$	10^6	> 3.0

$[0.52p_e, 1.48p_e]$. The smallest one is $[0.89p_e, 1.11p_e]$. In addition, we checked from the numerical data and noted that all the data points of the estimates (see in Figure 2 the green curves with diamond mark at each data point) are associated to soft coded BER mean values that are inside their corresponding intervals I . This checking combined with the observed intervals I that are associated to the achieved CIs allow us to conclude that the proposed estimator is reliable.

As far as the efficiency of the proposed estimator is concerned, we provided comparisons to the classical MC method in terms of the sample size saving given by N_{mc}/N_K . In the last columns of Table 3 and Table 4, sample size savings that characterised the observed accuracy (the absolute biases) and reliability (the CIs) of the proposed estimator are reported. To derive the sample size saving, the sample size N_{mc} is that required by the classical MC estimator to achieve almost equal accuracy and reliability. However, the values of N_{mc} that are preceded by the greater-than symbol (see the last rows in Table 3 and in Table 4) are smaller than the sample sizes which are truly required to meet almost equivalent reliability and accuracy as the proposed estimator. The right values are not

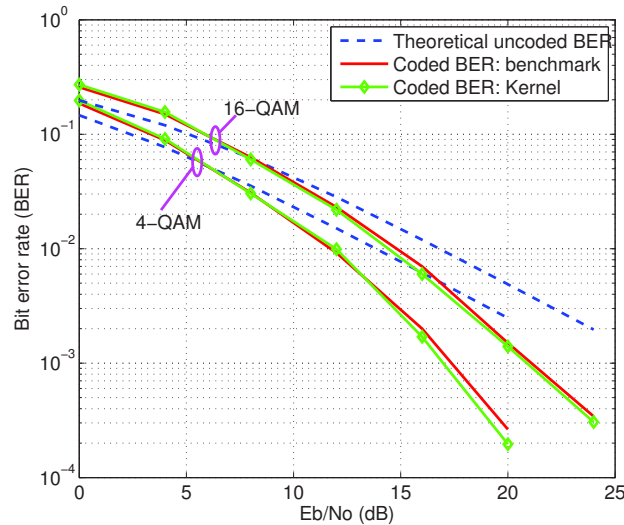


Figure 2. Performance of the proposed estimator over Rayleigh channel

Table 3. 4-QAM numerical results (Rayleigh channel)

E_b/N_0	Benchmark	Bias	I	N_K	N_{mc}/N_K
00 dB	1.86×10^{-1}	0.12×10^{-1}	$[0.88p_e, 1.12p_e]$	2.0×10^3	2.0
04 dB	8.86×10^{-2}	0.24×10^{-2}	$[0.77p_e, 1.23p_e]$	5.0×10^3	1.2
08 dB	3.08×10^{-2}	0.04×10^{-2}	$[0.73p_e, 1.27p_e]$	7.0×10^3	3.7
12 dB	9.20×10^{-3}	0.70×10^{-3}	$[0.70p_e, 1.30p_e]$	2.5×10^4	3.1
16 dB	2.00×10^{-3}	0.34×10^{-3}	$[0.55p_e, 1.45p_e]$	6.0×10^4	3.3
20 dB	2.64×10^{-4}	0.68×10^{-4}	$[0.52p_e, 1.48p_e]$	1.6×10^5	> 8.1

given because of the computational cost. We determined N_{mc} by the means of simulations. As we can see throughout the last columns of Table 3 and Table 4, sample size savings are observed. To illustrate how significant are the observed sample size reductions, let us consider the row of $E_b/N_0 = 12$ dB in Table 4. The proposed kernel-based estimator achieved an efficiency described by a sample size of 50 000 against 127 995 for the classical MC estimator. In the same time, quite equal CIs ($[0.81p_e, 1.19p_e]$ for the proposed estimator versus $[0.80p_e, 1.20p_e]$ for the classical MC estimator) are obtained. While the true BER is equal to 0.0231, the classical MC method performed the estimation with an absolute bias of 0.0011 meanwhile the proposed estimator yielded quite equal absolute bias of 0.0012. So for quite equal accuracy and reliability, the proposed estimator yielded significant sample size reduction. We also observed that for $E_b/N_0 = 20$ dB (in Table 4), both the proposed estimator and the classical MC method performed with equal absolute biases: the true BER is 0.0015 and the estimate is 0.0014. However, the proposed estimator not only performed more efficiently with a sample size saving greater than 5 but also showed to be more reliable: an achieved CI of $[0.67p_e, 1.33p_e]$ against $[0.62p_e, 1.38p_e]$ for the classical MC estimator.

Let us now exhibit the effect of the sample size reductions according to a computational efficiency point of view. We noted that the proposed estimator, by enabling sample size savings, also enabled CPU time savings, i.e., less energy consumption than the classical MC method. We computed the CPU times on a personal computer with an Intel(R) Core(TM) i5-6200U CPU 2.30GHz. We noted that the classical MC method and the proposed estimator yielded almost equal CPU time when the samples used for the estimation are of equal sizes. For instance, at $E_b/N_0 = 16$ dB and for a sample size of 6.0×10^4 , the CPU time equals 13 seconds for the classical MC method against 14 seconds for the

Table 4. 16-QAM numerical results (Rayleigh channel)

E_b/N_0	Benchmark	Bias	I	N_k	N_{mc}/N_K
00 dB	2.58×10^{-1}	0.13×10^{-1}	$[0.89p_e, 1.11p_e]$	1.0×10^3	3.0
04 dB	1.50×10^{-1}	0.06×10^{-1}	$[0.86p_e, 1.14p_e]$	2.0×10^4	1.0
08 dB	6.28×10^{-2}	0.26×10^{-2}	$[0.87p_e, 1.13p_e]$	5.0×10^4	1.0
12 dB	2.31×10^{-2}	0.12×10^{-2}	$[0.81p_e, 1.19p_e]$	5.0×10^4	2.6
16 dB	7.00×10^{-3}	1.00×10^{-3}	$[0.73p_e, 1.27p_e]$	5.0×10^4	2.0
20 dB	1.50×10^{-3}	0.08×10^{-3}	$[0.67p_e, 1.33p_e]$	1.0×10^5	> 5.1
24 dB	3.42×10^{-4}	0.36×10^{-4}	$[0.54p_e, 1.46p_e]$	4.1×10^5	> 6.3

Table 5. *Asymptotic behaviour of the proposed kernel-based estimator*

N_K	Benchmark	Bias	I
AWGN channel:			
1.0×10^3	3.0×10^{-3}	0.55×10^{-3}	$[0.39p_e, 1.61p_e]$
1.0×10^4	3.0×10^{-3}	0.18×10^{-3}	$[0.81p_e, 1.19p_e]$
1.0×10^5	3.0×10^{-3}	0.00×10^{-3}	$[0.94p_e, 1.06p_e]$
Rayleigh channel:			
1.0×10^3	8.9×10^{-2}	0.38×10^{-2}	$[0.76p_e, 1.24p_e]$
1.0×10^4	8.9×10^{-2}	0.24×10^{-2}	$[0.83p_e, 1.17p_e]$
1.0×10^5	8.9×10^{-2}	0.08×10^{-2}	$[0.87p_e, 1.13p_e]$

proposed estimator. However, when the sample size increases it causes the CPU time to increase too. So, the sample size saving brought by the proposed estimator is beneficial in terms of the power consumption. To illustrate this, the performance achieved at $E_b/N_0 = 20$ dB (see Table 3) is at the cost of a CPU time of 1 minute for the proposed estimator while being by far greater than 2 hours for the classical MC method.

Until now, we analysed the performance of the proposed estimator in terms of accuracy, reliability and efficiency. Let us end with an asymptotic analysis that can provide a better understanding of the estimator improvement as the sample size increases. For this purpose, Table 5 reports numerical data that show how the absolute bias and the CI evolve as the sample size increases. It appears that, when the sample size increases, the absolute bias of the estimation decreases and the interval I of the CI becomes smaller.

5. Conclusion

We studied the problem of the universal Monte Carlo (MC) simulation technique efficiency improvement regarding the Bit Error Probability (BEP) estimation of digital communication systems. We designed a kernel-based estimator for efficient and reliable computations of Bit Error Rate (BER) in contemporary and emerging digital communication systems that rely on coded M -ary Quadrature Amplitude Modulation (QAM) transmissions schemes. The proposed kernel-based BEP estimator has been designed to perform with soft real bits that are sampled from soft outputs of the channel decoder. We completely defined the estimator by selecting a kernel function that follows an Epanechnikov distribution with associated smoothing parameters that have been derived based on the canonical kernel concept. Simulation of the proposed estimator was made and results were reported regarding coded 4-QAM and 16-QAM single carrier transmissions over the additive white Gaussian noise channel. Simulation results were also reported for coded 4-QAM and 16-QAM multiple carrier transmissions over a frequency-selective Rayleigh fading channel. Based on the observed simulation results, the performance of the proposed kernel-based BEP estimator has been analysed in terms of accuracy, reliability, computational and sample size efficiency. Better performance than the universal classical MC method has been achieved by the proposed estimator for BEP estimates covering coded BER values from the neighbourhood of 10^{-1} down to the vicinity of 10^{-5} .

Acknowledgements

This work started thanks to the funding of *Projet RESEAU, SCAC-Ambassade de France, Burkina Faso*. Part of it has been performed in the framework of the *FP7 project ICT-317669 METIS*, which is partly funded by the European Union.

6. References

- [1] S. BENEDETTO, E. BIGLIERI, R. DAFFARA, “Modeling and performance evaluation of non linear satellite links-A volterra series approach”, *IEEE Journal on Selected Areas in Communications*, vol. AES-15, 1979, pp. 494-507.
- [2] K. YAO, L. B. MILSTEIN, “The use of moment space bounds for evaluating the performance of a non linear digital communication system”, *IEEE Transactions Communications*, vol. 31, 1983, pp. 677-683.
- [3] M. C. JERUCHIM “Techniques for estimating the bit error rate in the simulation of digital communication systems”, *IEEE Journal on Selected Areas in Communications*, vol. 2, num. 1, 1984, pp. 153-170.
- [4] S. SAOUDI, M. TROUDI, F. GHORBEL, “An Iterative Soft Bit Error Rate Estimation of Any Digital Communication Systems Using a Nonparametric Probability Density Function”, *EURASIP Journal Wireless Commun. and Networking*, vol. doi:10.1155/2009/512192, 2009.
- [5] K. SHANMUGAM, P. BALABAN, “A Modified Monte-Carlo Simulation Technique for the Evaluation of Error Rate in Digital Communication Systems”, *IEEE Transactions on Communications*, vol. 28, num. 11, 1980, pp. 1916-1924.
- [6] S. B. WEINSTEIN “Theory and application of some classical and generalized asymptotic distributions of extreme values”, *IEEE Transactions on Information Theory*, vol. 19, num. 2, 1973, pp. 148-154.
- [7] S. B. WEINSTEIN “Estimation of small probabilities by linearization of the tail of a probability distribution function”, *IEEE Transactions on Communication Technology*, vol. 19, num. 6, 1971, pp. 1149-1155.
- [8] E. CAVUS, C. L. HAYNES, B. DANESHRAJ, “Low BER performance estimation of LDPC codes via application of Importance Sampling to trapping sets”, *IEEE Transactions on Communications*, vol. 57, num. 7, 2009, pp. 1886-1888.
- [9] S. SAOUDI, T. DERHAM, T. AIT-IDIR, P. COUPE “A Fast Soft Bit Error Rate Estimation Method”, *EURASIP Journal Wireless Communications and Networking*, doi:10.1155/2010/372370, 2010.
- [10] M. ROSENBLATT, “Remarks on some non-parametric estimates of a density function”, *The Annals of Mathematical Statistics*, vol. 27, num. 3, 1956, pp. 832-837.
- [11] J. D. LASTER, J. H. REED, W. H. TRANTER “Bit Error Rate Estimation using Probability Density Function Estimators”, *IEEE Transactions on Vehicular Technology*, vol. 51, num. 1, 2003, pp. 260-267.
- [12] S. SAOUDI, T. AIT-IDIR, Y. MOCHIDA, “A Novel Non-Parametric Iterative Soft Bit Error Rate Estimation Technique for Digital Communications Systems”, *In: IEEE International Conference on Communications*, 2011, pp. 1-6.
- [13] J. DONG, T. AIT-IDIR, S. SAOUDI, “Unsupervised bit error rate estimation using Maximum Likelihood Kernel methods”, *In: IEEE Vehicular Technology Conference*, 2012, pp. 1-5.
- [14] P. PODA, S. SAOUDI, T. CHONAVEL, F. GUILLOUD, T. TAPSOBA, “Kernel-based performance evaluation of coded QAM systems”, *In: Proceedings CARI 2016*, 2016, pp. 182-191.

- [15] S. SAOUDI, F. GHORBEL, A. HILLION, "Some statistical properties of the kernel-diffeomorphism estimator", *Applied Stochastic Models and Data Analysis Journal*, vol. 13, 1997, pp. 39-58.
- [16] M. C. JONES, J. S. MARRON, S. J. SHEATER, "A brief survey of bandwidth selection for density estimation", *Journal of the American Statistical Association*, vol. 91, num. 433, 1996, pp. 401-407.
- [17] B. W. SILVERMAN, "Density Estimation for Statistics and Data Analysis", *Chapman and Hall, London*, 1986.
- [18] A. Z. ZAMBOM, R. DIAS, "A review of Kernel density estimation with applications to econometrics", *International Econometric Review*, vol. 5, num. 1, 2013, pp. 20-42.
- [19] D. W. SCOTT, R. A. TAPIA, J. R. THOMPSON, "Kernel density estimation revisited", *Non-linear Analysis, Theory, Methods and Applications*, vol. 1, 1977, pp. 339-372.
- [20] M. RUDEMO, "Empirical choice of histograms and kernel density estimators", *Scandinavian Journal of Statistics*, vol. 9, 1982, 9, pp. 65-18.
- [21] J. S. MARRON, "An asymptotically efficient solution to the bandwidth problem of kernel density estimation", *Annals of Statistics*, vol. 13, 1985, pp. 1011-1023.
- [22] G. R. TERRELL, D. W. SCOTT, "Variable Kernel density estimation", *The Annals of Statistics*, vol. 20, num. 3, 1992, pp. 1236-1265.
- [23] P. HALL, T. C. HU, J. S. MARRON, "Improved variable window Kernel estimates of probability densities", *The Annals of Statistics*, vol. 23, num. 1, 1995, pp. 1-10.
- [24] M. TROUDI, A. M. ALIM, S. SAOUDI, "Fast plug-in method for parzen probability density estimator applied to genetic neutrality study", *International Conference on Computer as a Tool (EUROCON'07)*, 2007, pp. 1034-1039.
- [25] D. FREEDMAN, P. DIACONIS, "On the histogram as a density estimator: L2 Theory", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 57, num. 4, 1981, pp. 453-476.
- [26] A. J. IZENMAN, "Modern Multivariate Statistical Techniques: Regression, Classification and Manifold Learning", *Springer, New York*, 2008.
- [27] Q. WANG, Q. XIE, Z. WANG, S. CHEN, L. HANZO, "A Universal Low-Complexity Symbol-to-Bit Soft Demapper", *IEEE Transactions on Vehicular Technology*, vol. 63, num. 1, 2014, pp. 119-130.
- [28] W. HÄRDLE, M. MÜLLER, S. SPERLICH, A. WERWATZ, "Nonparametric and semiparametric models", *Springer Verlag, Berlin Heidelberg*, 2004.
- [29] J. S. MARRON, D. NOLAN, "Canonical kernels for density estimation", *Statistics & Probability Letters*, vol. 7, 1988, pp. 195-199.
- [30] T. S. RAPPAPORT "Wireless Communications principles and practice" *New Jersey: Prentice Hall PTR* vol. 2, 1996.
- [31] J. G. PROAKIS, M. SALEHI "Digital Communications" *McGraw-Hill, 5th Edition*, 2008.
- [32] J. R. BARRY, E. A. LEE, D. G. MESSERSCHMITT "Digital Communication", *Kluwer Academic Publishers, 3rd edition*, 2004.

Appendix A

Bit-to-symbol soft mapping

In the bit-to-symbol soft mapping, the goal is to map $k = \log_2(M)$ soft bits onto a single complex-valued symbol. The complex-valued symbol is the soft observation that can

be used, e.g., for coded symbol error rate estimate computation. Let \hat{Z}_j be the complex-valued soft observation to be estimated by the means of bit-to-symbol soft mapping. \hat{Z}_j is expressed as follows:

$$\hat{Z}_j = \sum_{m=1}^M s_m \Pr[s_m]. \quad (\text{A.1})$$

As a bit vector (b_1, \dots, b_k) is assigned to each s_m at the transmitter and since s_1, \dots, s_M are independent, we get

$$\Pr[s_m] = \prod_{j=1}^k \Pr[b_{j,m}], \quad (\text{A.2})$$

where $b_{j,m}$ is the j -th bit in the bit vector (b_1, \dots, b_k) assigned to s_m . Finally, with the help of Eq. (24), we conclude with the expression of \hat{Z}_j as follows:

$$\hat{Z}_j = \sum_{m=1}^M s_m \prod_{i=1}^k \frac{e^{b_{i,m} L_i}}{1 + e^{L_i}}. \quad (\text{A.3})$$

Appendix B

Equation for the bit error rate computation

The BER estimate as given in Eq. (26) is

$$\hat{p}_e = \pi_0 \int_0^{+\infty} \frac{1}{n_0} \sum_{j=1}^{n_0} \frac{1}{h_0^*} K\left(\frac{x - X_j}{h_0^*}\right) dx + \pi_1 \int_{-\infty}^0 \frac{1}{n_1} \sum_{j=1}^{n_1} \frac{1}{h_1^*} K\left(\frac{x - X_j}{h_1^*}\right) dx, \quad (\text{B.1})$$

where n_0 (resp. n_1) is the cardinality of the subset of the soft observations among $(X_j)_{1 \leq j \leq N}$ which are likely to be decoded into a binary “0” (resp. “1”) bit value and h_0^* (resp. h_1^*) is the selected optimal smoothing parameter which will govern the accuracy of the estimation of $\hat{f}_X^{(0)}(x)$ (resp. $\hat{f}_X^{(1)}(x)$). More explicitly, as the kernel function $K(x) = \frac{3}{4} (1 - x^2) I(|x| \leq 1)$, we have

$$\begin{aligned} \hat{p}_e &= \frac{\pi_0}{n_0} \int_0^{+\infty} \sum_{j=1}^{n_0} \frac{3}{4h_0^*} \left[1 - \left(\frac{x - X_j}{h_0^*} \right)^2 \right] I\left(\left| \frac{x - X_j}{h_0^*} \right| \leq 1\right) dx \\ &+ \frac{\pi_1}{n_1} \int_{-\infty}^0 \sum_{j=1}^{n_1} \frac{3}{4h_1^*} \left[1 - \left(\frac{x - X_j}{h_1^*} \right)^2 \right] I\left(\left| \frac{x - X_j}{h_1^*} \right| \leq 1\right) dx. \end{aligned} \quad (\text{B.2})$$

Then, using one of the properties of the integral, we get

$$\begin{aligned} \hat{p}_e &= \frac{\pi_0}{n_0} \sum_{j=1}^{n_0} \int_0^{+\infty} \frac{3}{4h_0^*} \left[1 - \left(\frac{x - X_j}{h_0^*} \right)^2 \right] I\left(\left| \frac{x - X_j}{h_0^*} \right| \leq 1\right) dx \\ &+ \frac{\pi_1}{n_1} \sum_{j=1}^{n_1} \int_{-\infty}^0 \frac{3}{4h_1^*} \left[1 - \left(\frac{x - X_j}{h_1^*} \right)^2 \right] I\left(\left| \frac{x - X_j}{h_1^*} \right| \leq 1\right) dx. \end{aligned} \quad (\text{B.3})$$

Now, let $u = (x - X_j)/h_0^*$ and $v = (x - X_j)/h_1^*$ by the change of variables rule. We obtain

$$\begin{aligned} \hat{p}_e &= \frac{3\pi_0}{4n_0} \sum_{j=1}^{n_0} \int_{-X_j/h_0^*}^{+\infty} (1 - u^2) I(|u| \leq 1) du \\ &+ \frac{3\pi_1}{4n_1} \sum_{j=1}^{n_1} \int_{-\infty}^{-X_j/h_1^*} (1 - v^2) I(|v| \leq 1) dv, \end{aligned} \quad (\text{B.4})$$

and then,

$$\hat{p}_e = \frac{3\pi_0}{4n_0} \sum_{j=1}^{n_0} \int_{[\alpha_j, +\infty[\cap [-1, 1]} (1 - u^2) du + \frac{3\pi_1}{4n_1} \sum_{j=1}^{n_1} \int_{]-\infty, \beta_j] \cap [-1, 1]} (1 - v^2) dv, \quad (\text{B.5})$$

where $\alpha_j = -X_j/h_0^*$ and $\beta_j = -X_j/h_1^*$. Depending on the values of α_j (resp. β_j), three cases are possible among which one leads to zero; hence we get,

$$\begin{aligned} \hat{p}_e &= \frac{3\pi_0}{4n_0} \left\{ \sum_{\substack{\alpha_j < -1, \\ 1 \leq j \leq n_0}} \left[t - \frac{t^3}{3} \right]_{-1}^1 + \sum_{\substack{|\alpha_j| \leq 1, \\ 1 \leq j \leq n_0}} \left[t - \frac{t^3}{3} \right]_{\alpha_j}^1 \right\} \\ &+ \frac{3\pi_1}{4n_1} \left\{ \sum_{\substack{\beta_j > 1, \\ 1 \leq j \leq n_1}} \left[t - \frac{t^3}{3} \right]_{-1}^1 + \sum_{\substack{|\beta_j| \leq 1, \\ 1 \leq j \leq n_1}} \left[t - \frac{t^3}{3} \right]_{-1}^{\beta_j} \right\}. \end{aligned} \quad (\text{B.6})$$

Finally, the BER estimate expression is as follows:

$$\begin{aligned} \hat{p}_e &= \frac{\pi_0 L_0}{n_0} + \frac{\pi_1 L_1}{n_1} + \frac{3\pi_0}{4n_0} \left\{ \sum_{\substack{|\alpha_j| \leq 1, \\ 1 \leq j \leq n_0}} \left(\frac{2}{3} - \alpha_j + \frac{\alpha_j^3}{3} \right) \right\} \\ &+ \frac{3\pi_1}{4n_1} \left\{ \sum_{\substack{|\beta_j| \leq 1, \\ 1 \leq j \leq n_1}} \left(\frac{2}{3} + \beta_j - \frac{\beta_j^3}{3} \right) \right\}, \end{aligned} \quad (\text{B.7})$$

where L_0 (resp. L_1) is the cardinality of the subset of $\alpha_1, \dots, \alpha_{n_0}$ (resp. $\beta_1, \dots, \beta_{n_1}$) which are less than -1 (resp. greater than 1).