

Hindawi Publishing Corporation
EURASIP Journal on Wireless Communications and Networking
Volume 2010, Article ID 372370, 13 pages
doi:10.1155/2010/372370

Research Article

A Fast Soft Bit Error Rate Estimation Method

Samir Saoudi,^{1,2,3} Thomas Derham,³ Tarik Ait-Idir,^{4,5} and Patrice Coupe³

¹ *Institut TELECOM/TELECOM Bretagne, UMR CNRS 3192 Lab-STICC, Technopole Brest-Iroise, France*

² *UEB, Université Européenne de Bretagne, France*

³ *Orange Labs Tokyo, Japan*

⁴ *INPT, Madinat Al-Irfane, Rabat, Morocco*

⁵ *College of Computers and Information Systems, Taif University, Saudi Arabia*

Correspondence should be addressed to Samir Saoudi, samir.saoudi@telecom-bretagne.eu

Received 9 March 2010; Revised 15 July 2010; Accepted 30 August 2010

Academic Editor: Tho Le-Ngoc

Copyright © 2010 Samir Saoudi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We have suggested in a previous publication a method to estimate the Bit Error Rate (BER) of a digital communications system instead of using the famous Monte Carlo (MC) simulation. This method was based on the estimation of the probability density function (pdf) of soft observed samples. The kernel method was used for the pdf estimation. In this paper, we suggest to use a Gaussian Mixture (GM) model. The Expectation Maximisation algorithm is used to estimate the parameters of this mixture. The optimal number of Gaussians is computed by using Mutual Information Theory. The analytical expression of the BER is therefore simply given by using the different estimated parameters of the Gaussian Mixture. Simulation results are presented to compare the three mentioned methods: Monte Carlo, Kernel and Gaussian Mixture. We analyze the performance of the proposed BER estimator in the framework of a multiuser code division multiple access system and show that attractive performance is achieved compared with conventional MC or Kernel aided techniques. The results show that the GM method can drastically reduce the needed number of samples to estimate the BER in order to reduce the required simulation run-time, even at very low BER.

1. Introduction

To study the performance of a digital communications system, we need to use, in general, the Monte Carlo (MC) method to estimate the BER. A tutorial exposition of different techniques is provided in [1] with particular reference to four other specific methods: modified Monte Carlo simulation (importance sampling), extreme value theory, tail extrapolation, and quasianalytical method. The modified Monte Carlo is achieved by importance sampling which means that important events, and then errors, are artificially generated by biasing the noise process. At the end of simulation, the error count must be properly unbiased (see also [2]). The extreme value theory (see [3]) assumes that the pdf can be approximated by exponential function. The tail extrapolation method, which is a subset of the previous one, is based on the assumption that only the tail region of the pdf can be described by a generalized exponential class. The quasianalytical method combines noiseless simulation with analytical representation of noise. In [4], High-Order

Statistics (HOS) of the bit Log-Likelihood-Ratio (LLR) are used for evaluating performance of turbo-like codes. In this case, the characteristic function of the bit LLR is estimated by using its first cumulants (or moments). The pdf is therefore computed by using the inverse Fourier transform. The reader can find other recent papers on this topic (see [5–8]).

In [9], we have suggested a soft BER estimation based on the nonparametric computation of the pdf of the received data. We have shown, in this last case, that hard decision is not needed to compute the BER and that the total necessary number of transmitted data is very small compared to the classical MC simulation. This allows a significant reduction in run time for computer simulations and may also be used as the basis for rapid real-time BER estimation in real communication systems.

In this paper, we suggest to use a Gaussian Mixture (GM) model to estimate the pdf of the observed samples. Two conditional pdfs are computed corresponding to the transmitted bits equal to ± 1 . The Expectation Maximisation (EM) algorithm is used to estimate, in an iterative fashion,

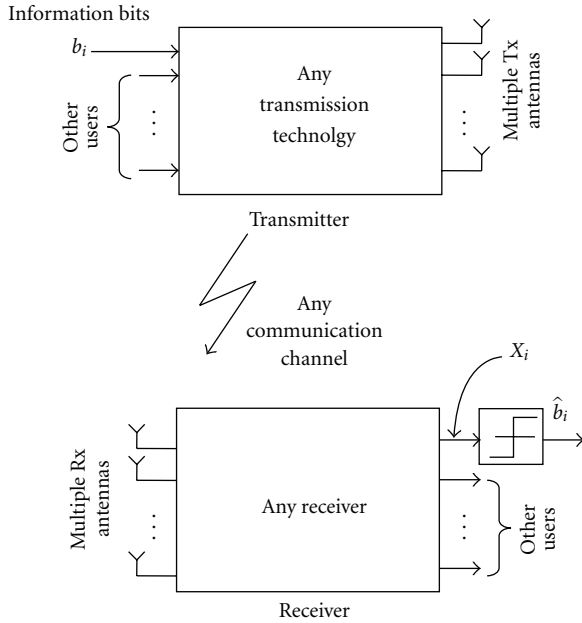


FIGURE 1: General transmission scheme for any transmitter and receiver with soft outputs X_1, \dots, X_N and hard decisions $\hat{b}_1, \dots, \hat{b}_N$.

the different parameters of this mixture, that is, the means, the variances, and the a priori probabilities. The different parameters are estimated by using the Maximum Expectation of joint Likelihood criterion. The analytical expression of the BER is therefore simply given by using the different estimated parameters of the Gaussian Mixture. The choice of the number of Gaussians for each pdf is very important. In [10], a method, based on Mutual Information Theory, was presented to find the optimal number of Gaussians in order to give an accurate estimation of the pdf. Our suggested analytical expression of the BER, based on the Gaussian Mixture model, where parameters are jointly estimated by EM algorithm and Mutual Information theory leads to an efficient fast way to estimate the performance of a digital communications system. Simulation results are carried out to compare the three mentioned methods: Monte Carlo, Kernel, and Gaussian Mixture. We analyze the performance of the proposed BER estimator in the framework of a multiuser code division multiple access (CDMA) system with single user detection and show that attractive performance is achieved compared with conventional Monte Carlo (MC) or Kernel aided techniques. The results show that the GM method can drastically reduce the needed number of samples to estimate the BER in order to reduce the simulation run-time even at very low BER.

The main idea of this paper is the use, in an iterative way, of the Expectation Maximisation (EM) Algorithm, for the pdf estimation with Gaussian Mixture model, jointly with the Mutual Information Theory, for the computation of the optimal number of Gaussians. The analytical expression of the BER is therefore given by using the different estimated parameters of the Gaussian Mixture.

The EM algorithm was introduced for the first time by Dempster et al. [11]. It is an iterative computation method of maximum likelihood estimates of missing data from observable variables. In this paper observable variables are simply given by soft output values at the receiver of a digital communications system. Missing data is given by unknown true component (Gaussian) from which the observation comes. Two conditional pdfs, according to the transmitted bits ± 1 , must be estimated. For each pdf, a Gaussian Mixture model, with a large enough initial number of components, is used. Then, the EM algorithm performs, in an iterative way, the estimation of the parameters for each component, that is, means, variances, and a priori probabilities. The Mutual Information (MI), according to Shannon Theory, is computed. A component with positive MI is assumed to be dependent on others components and could be removed without damaging the pdf estimation. The EM algorithm can be performed with a new decreased value of the number of Gaussians. The algorithm stops when the maximum of the computed MI, over all the components, is nonpositive which means that all the reached components are likely independent and therefore gives an optimal structure of the Gaussian Mixture model. The two conditional pdfs are then estimated, in a parallel fashion, by using the Mutual Information Theory to compute iteratively the optimal number of components and a subiteration for the EM algorithm to estimate the different parameters of each component. An analytical expression of the BER is therefore obtained by using all parameters of the Gaussian Model at the last iteration.

Let us recall that the EM algorithm has mainly been used, in the past twenty last years, in image processing or more precisely in image segmentation for different applications such as image or video compression. The reader can find in [12] an example of an application of SEM (Stochastic version of EM) algorithm in SPOT satellite image segmentation where a Gaussian distribution is assumed for each class. In [13], a hybrid version of SEM is used assuming a generalized Dirichlet distribution.

Nonparametric pdf estimation has also been used in different applications such as speech coding and pattern recognition [14, 15]. The Gaussian Mixture model has also been used in speaker identification [16].

Let us consider a general communications system (see Figure 1) where bit information is transmitted using any kind of transmission schemes such as CDMA, FDMA, TDMA, MC-CDMA techniques, with or without channel coding, space time coding, using single or multitransmit antennas, transmitting over Gaussian, fading or multipath, fixed or time variant, channel. At the receiver, any kind of detection such as MIMO equalization, multiuser detection, turbo techniques detection, or simply Rake receiver, may be implemented. The only assumption we use in this paper is the fact that the receiver is able to perform the soft decision (X_i in Figure 1) which is right before the hard decision, \hat{b}_i , of each transmitted bit. In this paper, only binary phase shift key (BPSK) modulation is used. The case of other kinds of modulation is left for future work.

Let $(b_i)_{1 \leq i \leq N} \in \{+1, -1\}$ be a set of N transmitted bits. The $(b_i)_{1 \leq i \leq N}$ are assumed to be independent and identically distributed with $P[b_i = +1] = \pi_+$ and $P[b_i = -1] = \pi_-$, where $\pi_+ + \pi_- = 1$. Let us note that $(X_i)_{1 \leq i \leq N}$ the corresponding soft output at the receiver such as the hard decision is taken by using its sign: $\hat{b}_i = \text{sgn}(X_i)$. All the received soft output decisions $(X_i)_{1 \leq i \leq N}$ are random variables having the same pdf, $f_X(x)$.

Throughout the paper, the following notation is used. The output decisions $(X_i)_{1 \leq i \leq N}$ are random variables having the same pdf, $f_X(x)$. The cardinality N of set \mathcal{C} is denoted $N = |\mathcal{C}|$. When X is a random variable, $\mathbb{E}[X]$ and $\text{Var}[X]$ denote the mathematical expectation and variance of X , respectively. When f is a second derivative function, $(f)'(x)$ and $(f)''(x)$ denote its first and second derivatives at point x , respectively. $\text{sgn}(\cdot)$ denotes the sign of the argument, and $\ln(\cdot)$ is the natural log function. $P[\cdot]$ is the probability of a given event, and superscript \top denotes the transpose.

The paper is organized as follows: Section 2 briefly shows how the probability density function (pdf) of the soft output signal at the receiver is estimated using the Kernel method in a nonparametric way by estimating the optimal smoothing parameter. The BER is performed based on all soft observations and the smoothing parameter value. In Section 3, we will show how a Gaussian Mixture model can be performed, for each conditional pdf, by using the Expectation Maximization (EM) algorithm. The Mutual Information is used to compute iteratively the optimal number of Gaussians. The BER is, therefore, simply computed by using all parameters (means, variances, and a priori probabilities) for each conditional pdf given by the EM algorithm at the last iteration. Different simulation results are presented in 4. Finally, a brief conclusion is given in Section 5. Proofs of all theoretical results are given in the appendices.

2. Kernel Method for BER Estimation

2.1. Pdf Estimation Based on Kernel Method. A brief description of the kernel method simulation will be given in this section. The reader can find more details in our previous work ([9]). Let us note that $f_X(x)$, the pdf of the output observations, is a mixture of the two conditional pdfs and then can be written as

$$f_X(x) = \pi_+ f_X^{b_+}(x) + \pi_- f_X^{b_-}(x), \quad (1)$$

Where $f_X^{b_+}(\cdot)$ (resp., $f_X^{b_-}(\cdot)$) is the conditional pdf of X such that $b_i = +1$ (resp., $b_i = -1$), $\pi_+ = P[b_i = +1]$, and $\pi_- = P[b_i = -1]$, with $\pi_+ + \pi_- = 1$. We assume that we know the exact partitions of the observations $\{(X_i)_{1 \leq i \leq N}\}$ into two classes (or partitions) \mathcal{C}_+ and \mathcal{C}_- which, respectively, contains the observed received soft bit X_i such as the corresponding transmitted bit information $b_i = +1$ (resp., $b_i = -1$). Let N_+ (resp., N_-) be the cardinality of \mathcal{C}_+ (resp., \mathcal{C}_-). The kernel method ([17–19]) is used to

estimate the different pdfs. In this case, the estimation of the conditional pdf, $f_X^{b_+}$, can be given by the following formula:

$$\hat{f}_{X, N_+}^{b_+}(x) = \frac{1}{N_+ h_{N_+}} \sum_{X_i \in \mathcal{C}_+} K\left(\frac{x - X_i}{h_{N_+}}\right), \quad (2)$$

Where h_{N_+} is the smoothing parameter which depends on the length of the observed samples, N_+ . $K(\cdot)$ is any pdf (called the kernel) assumed to be an even and regular (i.e, square integrated) function with unit variance and zero mean. For simplicity reasons, we will not give the corresponding equations for the conditional pdf, $f_X^{b_-}$. The reader can easily find them by replacing “+” by “-”.

The choice of the smoothing parameter h_{N_+} is very important ([14, 15]). Let us note that for conditional Gaussian distribution $f_X^{b_+} \sim \mathcal{N}(m, \sigma^2)$, and for Gaussian kernel, the optimal smoothing parameter is given by

$$h_{N_+}^* = \left(\frac{4}{3N_+}\right)^{1/5} \sigma. \quad (3)$$

2.2. BER Estimation Based on Kernel Method. The BER is given by:

$$p_e = \pi_+ \int_{-\infty}^0 f_X^{b_+}(x) dx + \pi_- \int_0^{+\infty} f_X^{b_-}(x) dx, \quad (4)$$

To estimate the BER of our system, we must evaluate the expression of (4). We can show that for the chosen Gaussian kernel, a soft BER estimation can be given by the following expression (see proof in [20]):

$$\hat{p}_{e, N} = \frac{\pi_+}{N_+} \sum_{X_i \in \mathcal{C}_+} Q\left(\frac{X_i}{h_{N_+}^*}\right) + \frac{\pi_-}{N_-} \sum_{X_i \in \mathcal{C}_-} Q\left(-\frac{X_i}{h_{N_-}^*}\right), \quad (5)$$

where $Q(\cdot)$ denotes the complementary unit cumulative Gaussian distribution, for example, $Q(x) = \int_x^{+\infty} (1/\sqrt{2\pi}) \exp(-t^2/2) dt$.

We have given some theoretical studies, in [20], regarding the convergence of this BER estimator. We have shown that this estimator is asymptotically unbiased. Different details can be found in [20].

3. Gaussian Mixture for BER Estimation

In this section, instead of using the Kernel method (given by (2)), a Gaussian Mixture (GM) model will be used. The mixture model is used in general for its mathematical flexibilities. For example, a mixture of two Gaussian distributions with different means and different variances results in a density with two modes, which is not a standard parametric distribution model. Mixture distributions can model extreme events better than the basic Gaussian ones. More details about Mixture distributions can be found in [21].

The following sub section will show how to estimate the two conditional pdfs using a Gaussian Mixture model. The Expectation Maximization (EM) algorithm will be performed to compute the mean, the variance and the a priori

probability of each component of this mixture. Therefore, Section 3.2 will show how the BER can be simply computed by using these different estimated parameters. For simplicity reason, equations are developed only for one conditional pdf, $f_X^{b+}(\cdot)$. The reader can easily find all the corresponding equations for the estimation of $f_X^{b-}(\cdot)$.

3.1. Pdf Estimation Based on Gaussian Mixture Method. In this section, we will assume that the conditional pdf f_X^{b+} is a mixture of K_+ Gaussians as follows: (see [21])

$$\hat{f}_{X,N_+}^{b+}(x) = \sum_{k=1}^{K_+} \alpha_k^+ f_k^+(x; \mu_k^+, \sigma_k^{+2}), \quad (6)$$

where α_k^+ is the a priori probability of the k th component for the Gaussian mixture and $f_k^+(x; \mu_k^+, \sigma_k^{+2})$ is a Gaussian pdf with μ_k^+ as a mean and σ_k^{+2} as a variance. We have the constraint that $\sum_{k=1}^{K_+} \alpha_k^+ = 1$.

Let $X = (X)_{1 \leq i \leq N_+}$ be the soft observed samples corresponding to the transmitted bits equal to +1. As the pdf of the observed samples is a mixture of K_+ Gaussians, this means (see [21]) that each X_i is produced by one component k of this mixture ($1 \leq k \leq K_+$). We have to find the value of this component: this is the missing data that we will try to compute. Let $Z = (Z_i)_{1 \leq i \leq K_+}$ be the missing data which is a sequence of variables that determines the component from which the observations originate. $Z_i = k$ means that X_i is generated by the k th component of the Gaussian mixture, that is, $X_i \sim \mathcal{N}(\mu_k^+, \sigma_k^{+2})$. The a priori probability α_k^+ represents the probability that $Z_i = k$, that is, $P[Z_i = k] = \alpha_k^+$.

In order to estimate the conditional pdf, from (6), we have to estimate the unknown parameters represented by $\theta_+ = (\alpha_k^+, \mu_k^+, \sigma_k^{+2})_{1 \leq k \leq K_+}$. The criterion we will use is the maximization of the conditional expectation of the joint likelihood of both observed samples, X , and missing data, Z (see [11] for details about this criterion). The likelihood function is given by:

$$L(\theta_+, X, Z) = \prod_{i=1}^{N_+} \left(\sum_{k=1}^{K_+} \mathbb{1}_k(Z_i) \alpha_k^+ f_k^+(X_i; \mu_k^+, \sigma_k^{+2}) \right), \quad (7)$$

where $\mathbb{1}_k(\cdot)$ is the indicator function given by:

$$\mathbb{1}_k(Z_i) = \begin{cases} 1, & \text{if } Z_i = k \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

In this section, we will use the Expectation Maximization algorithm to estimate, in an iterative way, the unknown parameter θ_+ . For each new iteration t and for a given estimate of the parameter $\theta_+^{(t-1)}$, computed at a previous iteration, two steps are performed. In the first one, that is, Estimation step, we will compute the different a posteriori probabilities (APP): $\rho_{k,i}^{+(t)} = P[Z_i = k | X_i = x_i, \theta_+^{(t-1)}]$. In the second one, that is, Maximization step, we will compute the new parameter $\theta_+^{(t)}$ by maximizing the conditional expectation of the Joint Log likelihood of observed samples, X , and missing data, Z .

3.1.1. Estimation Step. In this step, at iteration t , we estimate the unobserved component in the Gaussian mixture for each observed sample $(X_i)_{1 \leq i \leq N_+}$ using the parameter value $\theta_+^{(t-1)}$, computed at the last Maximization step at the previous iteration ($t-1$). Then, using simple Bayes' rule, we have:

$$\begin{aligned} \rho_{k,i}^{+(t)} &= P[Z_i = k | X_i = x_i, \theta_+^{(t-1)}], \\ &= \frac{P[X_i = x_i | Z_i = k, \theta_+^{(t-1)}] P[Z_i = k | \theta_+^{(t-1)}]}{P[X_i = x_i | \theta_+^{(t-1)}]}, \end{aligned} \quad (9)$$

Therefore, for $i = 1, \dots, N_+$, and for $k = 1, \dots, K_+$, we have:

$$\rho_{k,i}^{+(t)} = \frac{\alpha_k^{+(t-1)} f_k^+(X_i; \mu_k^{+(t-1)}, \sigma_k^{+(t-1)2})}{\sum_{k=1}^{K_+} \alpha_k^{+(t-1)} f_k^+(X_i; \mu_k^{+(t-1)}, \sigma_k^{+(t-1)2})} \quad (10)$$

3.1.2. Maximization Step. Now, at the current iteration t , we will maximize the conditional expectation of the log-likelihood of the joint event, assuming independent observation X_i . Then, we have:

$$\theta_+^{(t)} = \arg \max \mathbb{E}[\ln(L(\theta_+^{(t-1)}, X, Z)) | X] \quad (11)$$

Where $L(\theta_+^{(t-1)}, X, Z)$ is the joint likelihood event given by (7).

We can show that for $k = 1, \dots, K_+$, the new parameters are given by: (see Appendices A–D for proofs)

$$\alpha_k^{+(t)} = \frac{1}{N_+} \sum_{i=1}^{N_+} \rho_{k,i}^{+(t)}, \quad (12)$$

$$\mu_k^{+(t)} = \frac{\sum_{i=1}^{N_+} \rho_{k,i}^{+(t)} X_i}{\sum_{i=1}^{N_+} \rho_{k,i}^{+(t)}}, \quad (13)$$

$$\sigma_k^{+(t)2} = \frac{\sum_{i=1}^{N_+} \rho_{k,i}^{+(t)} (X_i - \mu_k^{+(t)})^2}{\sum_{i=1}^{N_+} \rho_{k,i}^{+(t)}}. \quad (14)$$

3.2. BER Estimation Based on Gaussian Mixture Method.

In this section, we derive the expression of BER estimate assuming Gaussian Mixture based pdf estimator. Let $\theta_+^{(T)} = (\alpha_k^{+(T)}, \mu_k^{+(T)}, \sigma_k^{+(T)2})_{1 \leq k \leq K_+}$ and $\theta_-^{(T)} = (\alpha_k^{-(T)}, \mu_k^{-(T)}, \sigma_k^{-(T)2})_{1 \leq k \leq K_-}$ the reached values at the last iteration of the EM algorithm described in the previous Section 3.1. The parameter $\theta_+^{(T)}$ (resp., $\theta_-^{(T)}$) allows the estimation of the conditional pdf $f_X^{b+}(\cdot)$ (resp., $f_X^{b-}(\cdot)$). Let us underline that we need to perform the EM algorithm two times and in independent way. At each time, a different data base is used: \mathcal{C}_+ (resp., \mathcal{C}_-), of soft observations corresponding to the transmitted bits equal +1 (resp., -1) to estimate $\theta_+^{(T)}$ (resp., $\theta_-^{(T)}$).

At the last iteration T of EM algorithm, reliable estimates of $\theta_+^{(T)}$ and $\theta_-^{(T)}$ are reached and the BER is computed using the obtained estimates. Let us recall the expression of the BER

given by (4). Replacing the two conditional pdfs by their Gaussian Mixture based estimates (6) using the parameters $\theta_+^{(T)}$ and $\theta_-^{(T)}$, the BER estimates is simply computed as

$$\hat{p}_{e,N} = \pi_+ \sum_{k=1}^{K_+} \alpha_k^{+(T)} Q\left(\frac{\mu_k^{+(T)}}{\sigma_k^{+(T)}}\right) + \pi_- \sum_{k=1}^{K_-} \alpha_k^{-(T)} Q\left(-\frac{\mu_k^{-(T)}}{\sigma_k^{-(T)}}\right), \quad (15)$$

Where $Q(\cdot)$ denotes the classical complementary unit cumulative Gaussian distribution. Details regarding the derivation of (15) are provided in Appendix D.

3.3. Optimal Choice of the Number of Components. The choice of K_- and K_+ is very important. It is clear that if this number of components is too low, the corresponding pdf will be too smooth and then the BER less reliable. On the other hand, if this number is too high, this means that the same class of observed samples comes from different components and then these components should be correlated which is not useful for simulation since all the observed data are assumed to be independent. Consequently, the optimal number of components has to be the largest one such that all the components are independent. Similar method has been suggested by [16] for speaker identification applications by increasing the number of classes in the k -means algorithm. More mathematical details can be found in [10]. Here, we suggest to initialize the algorithm with a high enough value, to perform the EM algorithm to estimate the different components and test their independence after the last iteration. If it is not the case, we have to decrease iteratively the number of components until the independence is reached.

To test the independence of two components k_1 and k_2 , mutual information theory, as proposed by Shannon [22], can be used. For speaker identification application, in [16], the mutual relationship of the two components has been defined as

$$MI_+(k_1, k_2) = p_+(k_1, k_2) \log_2 \frac{p_+(k_1, k_2)}{p_+(k_1)p_+(k_2)}, \quad (16)$$

Where, $p_+(k_1) = \alpha_k^+$ is the probability of the mixture k_1 (see (6)), and $p_+(k_1, k_2)$ is the joint probability of these two components as follow,

$$p_+(k_1) = \frac{1}{N_+} \sum_{i=1}^{N_+} \rho_{k_1, i}^+, \quad (17)$$

$$p_+(k_1, k_2) = \frac{1}{N_+} \sum_{i=1}^{N_+} \rho_{k_1, i}^+ \rho_{k_2, i}^+, \quad (18)$$

The sign of the expression of (16) allows us to know whether the two components are statistically independent: if $\text{sgn}(MI_+(k_1, k_2)) = 0$ then the two components are independent ($p_+(k_1, k_2) = p_+(k_1)p_+(k_2)$), if $\text{sgn}(MI_+(k_1, k_2)) > 0$ then the two components are statistically dependent and then one of these components can be removed without damaging the estimation of the pdf. If $\text{sgn}(MI_+(k_1, k_2)) < 0$, the two

components are much less correlated. So, the following quantity,

$$I_+ = \max(MI_+(k_1, k_2))_{1 \leq k_1, k_2 \leq K_+} \quad (19)$$

allows us to know if we have to reduce the number of components or not. To find the optimal number of components K_- and K_+ , we have to just choose a large enough initial number and at the end of the EM algorithm, the sign of I_+ (or I_-) allows us to know if we have to reduce K_+ (or K_-). If $\text{sgn}(I_+) > 0$, we decrease the number of component by one, otherwise we stop the algorithm. The computation of the optimal values K_+ and K_- , which could of course be different, can be performed in a parallel fashion. For a new decreased value, $(K_+ - 1)$, initial GM parameters of EM algorithm could be given by the output parameters at the last iteration of the previous EM algorithm, where we remove the k th component given by

$$k = \arg \max_{1 \leq k_1 \leq K_+} \sum_{k_2=1}^{K_+} MI_+(k_1, k_2). \quad (20)$$

The quantity $\sum_{k_2=1}^{K_+} MI_+(k_1, k_2)$ presents the mutual information for the component k_1 and denotes whether this component has a significant and independent contribution to the pdf estimation. The biggest positive value has a less and dependent contribution to the GM estimation and should therefore be removed. The proposed Gaussian Mixture based BER estimation using EM algorithm and Mutual Information theory can now be summarized in Algorithm 1. Figure 2 gives the flow chart of the suggested algorithm.

4. Performance Evaluation

To evaluate the performance of the three methods, we consider the framework of a synchronous CDMA system with two users using binary phase-shift keying (BPSK) and operating over an additive white Gaussian noise (AWGN) channel. We restrict ourselves to the conventional single user CDMA detector. Performance assessment in the case of advanced signaling/receivers is not reported in this paper due to space limitation and is left for future contributions.

With respect to the considered framework, the received $L_{SF} \times 1$ chip-level signal vector at discrete time instant i can be expressed as

$$\mathbf{r}_i = A_1 b_i^{(1)} \mathbf{s}_1 + A_2 b_i^{(2)} \mathbf{s}_2 + \mathbf{n}_i, \quad (21)$$

where L_{SF} denotes the spreading factor, and $\mathbf{s}_k \in \{\pm 1/\sqrt{L_{SF}}\}^{L_{SF}}$ is the spreading code corresponding to user k . A_k is the amplitude of user $k = 1, 2$, $b_i^{(k)}$ is the information bit value $\in \{\pm 1\}$ of user k at time instant i , and $\mathbf{n}_i \in \mathbb{R}^{L_{SF}}$ is the temporally and spatially white Gaussian noise, that is, $\mathbf{n}_i \sim \mathcal{N}(0, \sigma^2 I_{L_{SF}})$. The *a priori* probabilities of information bits are supposed to be identical and uniform for both users, that is, $\pi_+ = \pi_- = 1/2$.

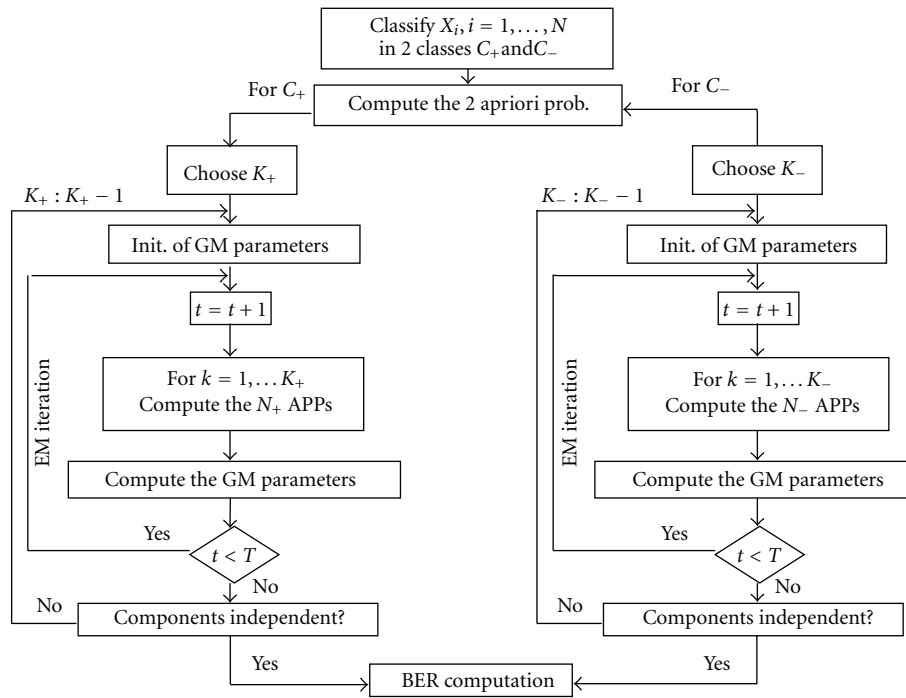


FIGURE 2: Flow Chart for the proposed Fast iterative BER estimation based on EM-Gaussian Mixture and Mutual Information measures.

1. Initialization:

- 1.1. Classify soft outputs X_1, \dots, X_N according to their transmitted bits, that is, $C_+ = \{X_i | b_i = +1\}$, $C_- = \{X_i | b_i = -1\}$. Let $N_+ = |C_+|$, $N_- = |C_-|$
- 1.2. Deduce the two priori probabilities: $\pi_+ = N_+/N$, $\pi_- = N_-/N$
- 1.3. Choose a large enough number of the Gaussian mixture: K_+ and K_- .

2. EM algorithm for C_+ :

- 2.1. Initialization of $\theta_+^{(0)}$
- 2.2. At each iteration t , ($t = 1, \dots, T$)
 - 2.2.1 Estimation step: Estimate APPs, $\rho_{k,i}^{+(t)}$ using (10).
 - 2.2.2 Maximization step: Computation of $\theta_+^{(t)}$ by using (12), (13) and (14).

3. Verification of optimality of K_+ :

- 3.1. Compute I_+ (19)
- 3.2. If $I_+ \leq 0$, save $\theta_+^{(T)}$ go to step 4. Otherwise, $K_+ : K_+ - 1$, go to 2.

4. EM algorithm for C_- :

- 4.1. Initialization of $\theta_-^{(0)}$
- 4.2. At each iteration t , ($t = 1, \dots, T$)
 - 4.2.1 Estimation step: Estimate APPs, $\rho_{k,i}^{-(t)}$ using (10).
 - 4.2.2 Maximization step: Computation of $\theta_-^{(t)}$ by using (12), (13) and (14).

5. Verification of optimality of K_- :

- 5.1. Compute I_- (19)
- 5.2. If $I_- \leq 0$, save $\theta_-^{(T)}$ go to step 6. Otherwise, $K_- : K_- - 1$, go to 4.

6. BER Computation:

- 6.1. Compute the BER estimate from (15) using $\theta_+^{(T)}$ and $\theta_-^{(T)}$.

ALGORITHM 1: Summary of the Gaussian Mixture based BER estimation using EM algorithm and Mutual Information theory. $\theta_+^{(T)}$ and $\theta_-^{(T)}$ are computed in a parallel fashion.

The decision statistic that serves for detecting user 1 at time instant i is $X_i^{(1)} = s_1^\top \mathbf{r}_i$ [23] and is given as,

$$X_i^{(1)} = A_1 b_i^{(1)} + A_2 b_i^{(2)} \rho + \tilde{n}_i^{(1)}, \quad (22)$$

Where ρ is the normalized cross-correlation between the two spreading codes s_1 and s_2 , and $\tilde{n}_i^{(1)}$ is the Gaussian noise at the output of the single user detector, that is, $\tilde{n}_i^{(1)} \sim \mathcal{N}(0, \sigma^2)$. The decision about information bit $b_i^{(1)}$ corresponds to the sign of decision statistic $X_i^{(1)}$, that is, $\hat{b}_i^{(1)} = \text{sgn}(X_i^{(1)})$. Note that the soft output $X_i^{(1)}$ in (22) contains a mixture of a Gaussian noise. Using (22), we can easily show that the BEP for user 1 is

$$p_{e1} = \frac{1}{2} Q\left(\frac{A_1 - A_2 \rho}{\sigma}\right) + \frac{1}{2} Q\left(\frac{A_1 + A_2 \rho}{\sigma}\right). \quad (23)$$

In the following, we use the two spreading codes

$$s_1 = \frac{1}{\sqrt{7}} [+ + + + - - -]^\top, \quad (24)$$

$$s_2 = \frac{1}{\sqrt{7}} [- - + + - - -]^\top, \quad (25)$$

Where the cross-correlation is $\rho = 3/7$. We consider the case where the two users have equal powers $A_1 = A_2 = 1$. The SNR at the output of the MF of each user is therefore $\text{SNR} = 1/2\sigma^2$.

4.1. Output Pdf Estimation Comparison. First of all, in this simulation, we would like to compare the three different pdf estimation: Histogram method which leads to MC method, Kernel method and Gaussian Mixture method. In order to make a fair comparison, the three methods are used in optimal conditions. In particular, the length of the bins of the Histogram is chosen equal to the smoothing parameter computed for the kernel method so as the convergence of the histogram in the MSE and IMSE criterion can be guaranteed.

For this first simulation, we have chosen a pdf as a mixture of 3 different pdfs according to Gaussian, Rayleigh and Beta first kind laws with fixed different parameters. So the true chosen pdf is

$$f(x) = a_1 f_1(x) + a_2 f_2(x) + a_3 f_3(x), \quad (26)$$

Where, $f_1(x) = e^{-(x+1)^2/2}/\sqrt{2\pi}$, $f_2(x) = x e^{-x^2/2} \mathbb{1}_{\mathbb{R}^+}$ and $f_3(x) = 6x(1-x) \mathbb{1}_{[0,1]}(x)$. $a_1 = 0.50$, $a_2 = 0.40$ and $a_3 = 0.10$. Figure 3 plots the true pdf with the estimated pdf for the three cases. $N = 2,000$ samples has been generated for this first simulation. GM method has found that 4 components are sufficient to estimate the pdf as $\hat{f}_{GM}(x) = 0.27 \mathcal{N}(x; -1.64, 0.51) + 0.20 \mathcal{N}(x; -0.38, 0.39) + 0.24 \mathcal{N}(x; 0.67, 0.11) + 0.29 \mathcal{N}(x; 1.53, 0.50)$.

The Integrated Square Error (ISE), which is defined as $\text{ISE} = \int (f(x) - \hat{f}(x))^2 dx$, has been computed for this simulation. We have carried out one hundred different trials and computed the Mean ISE (MISE) and the variance of the ISE for the three methods. Table 1 summarizes

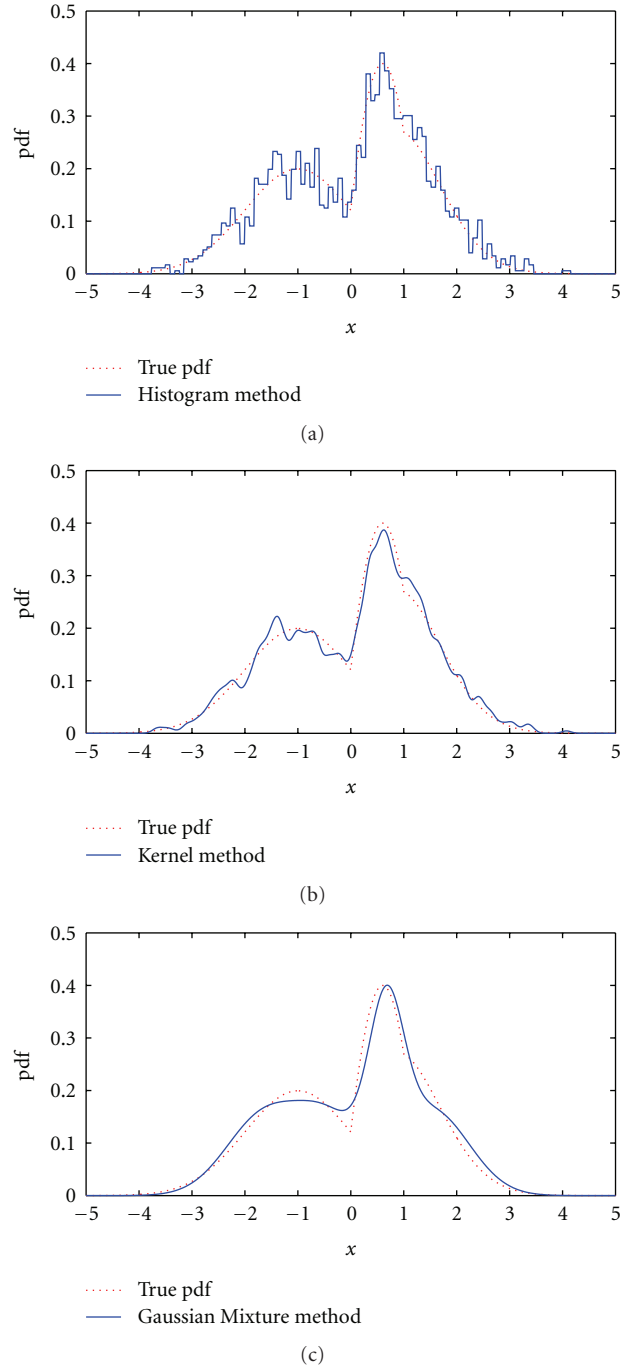


FIGURE 3: Comparison between the real and the estimated pdfs. Three different methods are used. From top to bottom: Histogram, Kernel, and Gaussian methods. $N = 2,000$ Samples are used. The same samples are used for the three methods.

these results and shows that Kernel method gives the best estimation of the pdf in the sense of the minimum of MISE. Gaussian Mixture method seems to be the worst one in the MISE criterion. So, for general applications, such as pattern recognition or speech coding, the Kernel method seems to be the best one to choose. For our application, that is, BER estimation, we do not need an accurate estimate of the whole

TABLE 1: Mean and Standard deviation of Integrated Square Error pdf estimation for the three methods by using 1,000 different trials. $N = 2,000$ samples are used for each trials.

ISE	Histogram	Kernel	Gaussian Mixture
Mean	7.44×10^{-2}	7.29×10^{-2}	8.09×10^{-2}
Std	5.40×10^{-3}	5.40×10^{-3}	5.00×10^{-3}

TABLE 2: Mean and Standard deviation of error estimation of $A = \int_{+3}^{\infty} f(x)dx$ using the three methods with 1,000 different trials. $N = 2,000$ samples are used for each trial.

$A = \int_{+3}^{\infty} f(x) - \hat{f}(x)dx$	Histogram	Kernel	Gaussian Mixture
Mean	5.10×10^{-3}	8.10×10^{-3}	3.60×10^{-3}
Std	2.23×10^{-2}	2.30×10^{-2}	2.38×10^{-2}

pdf. We only need an accurate estimate of the tail of the pdf. That is why, in another simulation, we are interested in estimating the area of the tail delimited between, for example, +3 and $+\infty$. In this area, the tail is a mixture of Gaussian and Rayleigh laws. We can easily show that in this case $A = \int_{+3}^{\infty} f(x)dx = a_1Q(4) + a_3e^{-4.5} = 11 \times 10^{-4}$. Let us now estimate the quantity A using the three different pdf estimates with one hundred different trials. Table 2 shows the mean and the Standard deviation for the three methods. GM method clearly gives the best estimate of the area computation. This result will be confirmed in the following sub section for BER estimation in the framework of CDMA systems described in the begining of Section 4.

4.2. Performance Comparison for the Three Methods. In order to compare the three methods (MC, Kernel and GM method), $N = 2,000$ soft outputs were generated for each SNR to estimate the BER. 1,000 different trials were simulated, for each method, to compute the Mean, Minimum and Maximum of BER values. The number of EM iterations is fixed to 20. All these results are given in Figure 4.

In addition, the Mean, the Standard deviation with the theoretical value of BER for different values of SNR are given in Table 3. This table shows that GM method provides the best performance in the sense of the minimum of Mean Squared Error of the BER estimation, even if a small bias is observed for medium values of SNR. This bias will completely disappear for increased number of EM iterations as it will be seen in Section 4.3. In the same time, MC technique fails to do so and stops between 8 and 14 dB because of the very limited number of transmitted information bits and lack of errors. The Kernel method leads to a smaller bias but has the greatest standard deviation and then less reliable than the GM method.

On the other hand, for GM method and for a fixed SNR value (10 dB), we have computed the Mean BER and the Standard deviation for different number of soft outputs (between 1,000 and 100,000) with 100 different trials.

The number of EM iterations is fixed to 20. All these results are given in Table 4. The precision of the estimation,

TABLE 3: Mean and Standard deviation of BER estimation with the three methods (MC, Kernel, GM) at different SNR. $N = 2,000$ samples are used for each simulation. 1,000 different trials are performed to compute the Mean and the Standard deviation.

SNR (dB)	Theoretical BER	MC	Kernel	GM
		Mean Std. Dev.	Mean Std. Dev.	Mean Std. Dev.
0	1.2×10^{-1}	1.1×10^{-1}	1.2×10^{-1}	1.2×10^{-1}
		7.2×10^{-3}	6.8×10^{-3}	6.2×10^{-3}
4	5.0×10^{-2}	5.0×10^{-2}	5.1×10^{-2}	5.1×10^{-2}
		4.8×10^{-3}	4.6×10^{-3}	4.3×10^{-3}
8	1.1×10^{-2}	1.1×10^{-2}	1.1×10^{-2}	9.6×10^{-3}
		2.3×10^{-3}	2.1×10^{-3}	1.6×10^{-3}
12	3.2×10^{-4}	3.2×10^{-4}	4.1×10^{-4}	5.0×10^{-4}
		4.0×10^{-4}	3.6×10^{-4}	2.4×10^{-4}
16	8.5×10^{-8}	–	4.4×10^{-7}	6.6×10^{-7}
		–	8.5×10^{-6}	4.6×10^{-6}
20	1.6×10^{-16}	–	2.5×10^{-16}	8.7×10^{-16}
		–	7.8×10^{-15}	1.9×10^{-15}

TABLE 4: Mean, Standard deviation and precision of BER estimation GM method, for SNR = 10 dB, at different number of samples are used for each simulation. 100 different trials are performed to compute the Mean, the Standard deviation and the precision.

Nb of samples	Mean of BER	Std. Dev.	Precision
1,000	2.7×10^{-3}	5.71×10^{-4}	0.220
2,000	2.7×10^{-3}	6.00×10^{-4}	0.220
5,000	2.7×10^{-3}	5.95×10^{-4}	0.220
10,000	2.8×10^{-3}	6.45×10^{-4}	0.232
15,000	2.8×10^{-3}	5.97×10^{-4}	0.225
20,000	2.7×10^{-3}	6.19×10^{-4}	0.227
25,000	2.7×10^{-3}	6.24×10^{-4}	0.228
30,000	2.7×10^{-3}	6.19×10^{-4}	0.230
35,000	2.7×10^{-3}	6.59×10^{-4}	0.242
40,000	2.7×10^{-3}	7.08×10^{-4}	0.259
100,000	2.7×10^{-3}	6.13×10^{-4}	0.226

defined as the standard deviation to the mean of BER ratio, does not decrease with the number of samples. This is linked to the observed bias mentioned before.

Figure 5 shows the performance of the receiver, for GM method, with different number of EM iterations and using $N = 2,000$ samples. This figure clearly shows that performance of GM method increases with the number of iterations.

4.3. Performance of GM Method in the High SNR Region. We would like to test our suggested algorithm in severe conditions for high values of SNR while using, in the same time, few samples such $N = 1,000$ output observations. Figure 6 shows the performance of the receiver (one random simulation), using 50 EM-iterations. We can see that we

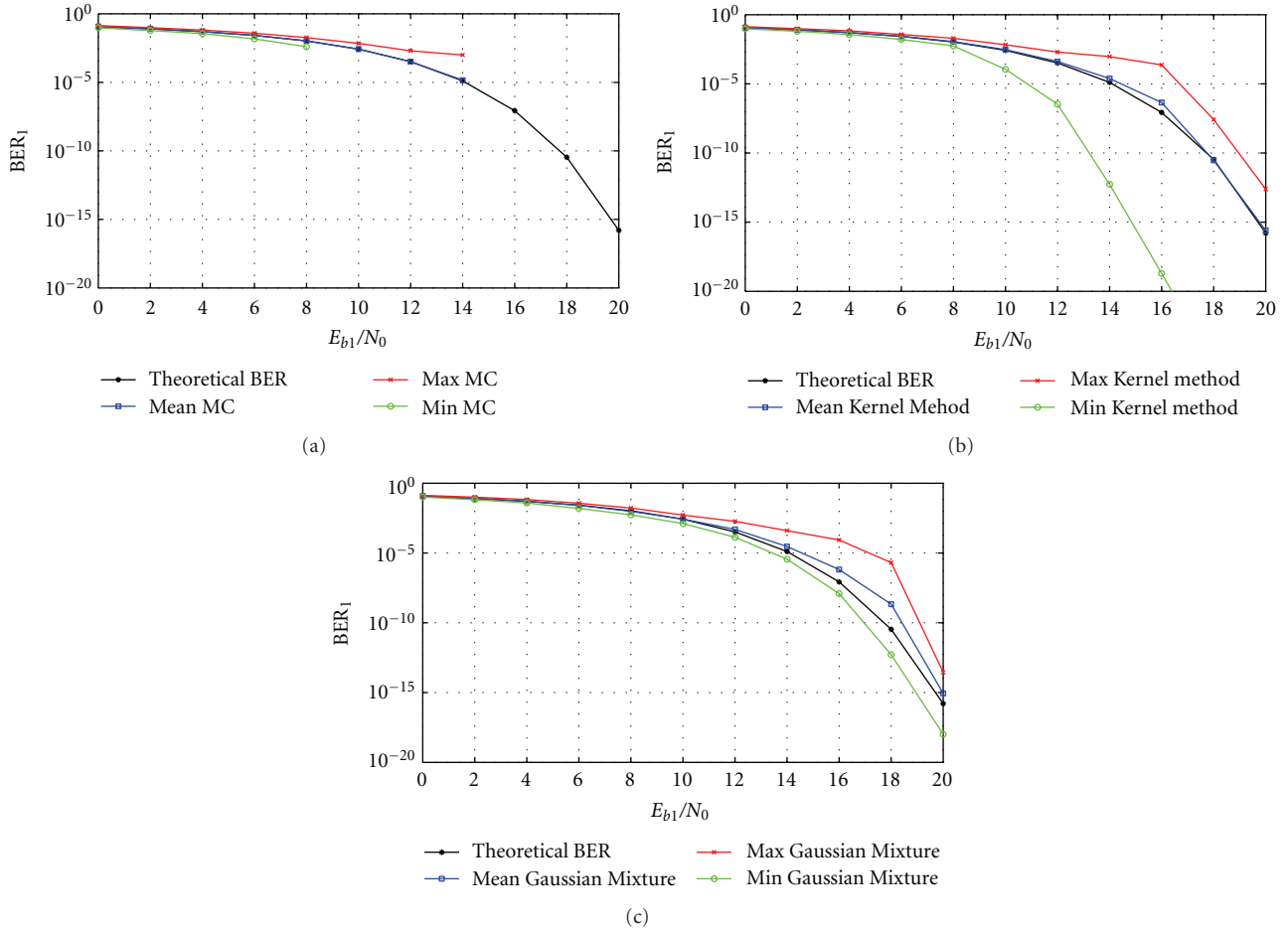


FIGURE 4: BER Comparison: Three different methods are used. From top to bottom: MC, Kernel, and Gaussian mixture method. $N = 2,000$ Samples are used for each simulation. 1,000 different trials are performed to compute the Mean, Min, and Max.

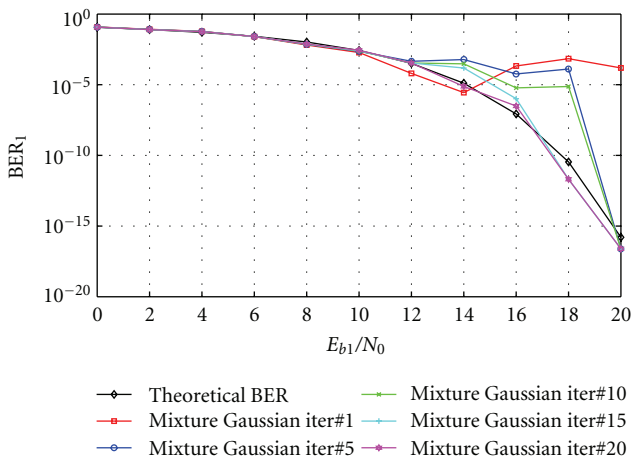


FIGURE 5: BER estimated by Gaussian mixture method with different iteration. $N = 2,000$ Samples are used.

computer precision. To obtain this figure, the simulation run time is 31 seconds on a conventional PC. For each SNR point, the simulation takes less than 2 seconds. This time is computed by using the CPU-time command of Matlab software. The run time does not depend on the value of SNR which is a huge advantage of our suggested method. For 20 EM iterations, the run time takes less than one second for each SNR value.

It is, of course unimaginable and impossible today, to plot this kind of figures using Monte Carlo method and waiting for 100 errors to have an accurate estimate. Indeed, using our computer with Matlab Software, we need 43 milliseconds to generate 1,000 output observations. Let us now assume that in the world, there are 10 billion inhabitants, that each one has got 10 PC and each PC is 10,000 times more powerful than the one we used for our simulation. Let us also assume that we can use all these PCs in a parallel and optimized structure. A simple calculation shows that we need to wait for more than one year to estimate a BER at only 10^{-25} , hoping that there will be no failure or cut electricity. The CPU time for MC method at different values of SNR is given in Table 5.

obtain an unbelievable result, a BER estimate down to 10^{-200} has been measured. This result is only limited by the

TABLE 5: CPU time for Monte Carlo simulation at different BER, assuming using 10^{15} computers with equivalent power to the one we used for Gaussian Mixture simulation.

BER	CPU time					
	Years	Months	Days	Hours	mn	s
10^{-20}					7	10
10^{-21}				1	11	40
10^{-22}				11	56	40
10^{-23}			4	23	26	40
10^{-24}		1	19	18	26	40
10^{-25}	1	4	17	16	26	40
10^{-26}	13	9	26	20	26	40

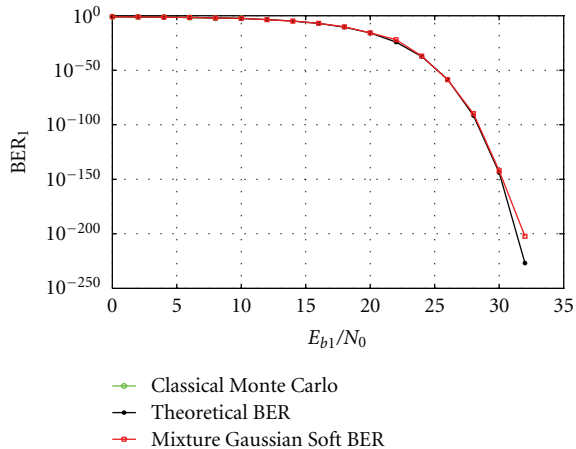


FIGURE 6: BER estimated by Gaussian mixture method for very high SNR values and only using $N = 1,000$ samples with 50 iterations. See Figure 7 for a zoom of this figure at high BER to see Monte Carlo simulation.

4.4. Comparison with Importance Sampling Technique. Importance sampling technique, also known as Modified Monte Carlo method, was introduced by Shanmugam and Balaban (see [2]) to estimate error probabilities in digital communications systems. The importance sampling technique is used to modify the probability density function of the noise process in a way to make simulation possible. An estimate of the pdf of soft decision is constructed in a histogram form. The idea is to modify this pdf, that is, the statistical properties of the soft decision sequence, in such a way that higher rate of errors occur in the simulation process. Therefore, the error count has to be modified appropriately to obtain an unbiased estimate of the true error probability. The main drawback of the importance sampling technique is the difficulty, for complex systems, of determining which regions of the pdf to bias and how to bias these regions.

Compared to classical Monte Carlo method, the importance sampling technique, reduces the sample size requirement by a factor ranging from 10 up to 1,000. Let us assume that this factor is equal to 1,000 in order to compare importance sampling technique with our suggested GM method. In this case, the CPU time given in Table 5 for Monte Carlo simulation has to be divided by a factor of

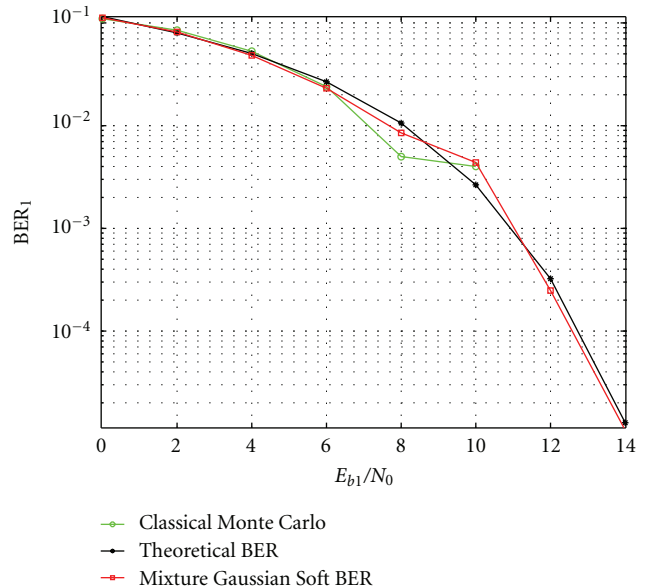


FIGURE 7: A zoom of Figure 6 at high BER to see Monte Carlo simulation.

1,000 to obtain the CPU time for the importance sampling technique simulation. Our suggested GM method has still a huge advantage as the run time does not depend on the value of SNR. In fact, for each SNR point, the simulation takes less than 2 seconds (see Section 4.2). Another advantage of our suggested GM method is the possibility to estimate the performance of a system at a very very low BER (down to 10^{-200} for a CDMA system case in this paper, see Figure 6). The only limitation is given by the precision of the used computer.

5. Conclusions

In this paper, we considered the problem of BER estimation for a digital communications system using any transmission technology or channel coding. BPSK modulation is used. The receiver is assumed to be able to compute soft decision. We proposed a BER estimation algorithm where only soft observations that serve for computing hard decisions about information bits are used. First of all, we provided a formulation of the problem where we showed that BER estimation is equivalent to the estimation of conditional pdfs of soft observations corresponding to the transmitted bits equal to ± 1 . We then proposed a BER computation technique using Gaussian Mixture-based pdf estimation. The Expectation Maximisation (EM) algorithm was used to estimate, in an iterative way, the different parameters of this mixture, that is, the means, the variances and the a priori probabilities. The analytical expression of the BER was therefore simply given by using the different estimated parameters of the Gaussian Mixture. The optimal number of Gaussians was computed using Mutual Information Theory. Finally, we evaluated the performance of the proposed BER estimation technique in the framework of CDMA systems.

Performance comparison with MC techniques and Kernel method was simulated. Interestingly, we showed that while classical MC method fails to perform BER estimation in the region of high SNR, the proposed GM estimator provides reliable estimates and better, in the sense of minimum Mean Squared Error, than Kernel method using only few soft observations. A measure of BER down to 10^{-200} has been reached in less than 2 seconds using only 1,000 soft outputs samples.

Appendices

A. Proof of (12)

Proof. Using (7), the conditional Expectation of the log likelihood function can be written as:

$$\begin{aligned} Q(\theta_+) &= \mathbb{E}[\ln L(\theta_+, X, Z) | X, \theta_+(t)] \\ &= \mathbb{E} \left[\sum_{i=1}^{N_+} \sum_{k=1}^{K_+} \mathbb{1}_k(Z_i) \left[\ln \alpha_k^+ - \frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln \sigma_k^{+2} \right. \right. \\ &\quad \left. \left. - \frac{(X_i - \mu_k^+)^2}{2\sigma_k^{+2}} \right] | X, \theta_+(t) \right], \\ &= \sum_{i=1}^{N_+} \sum_{k=1}^{K_+} \rho_{k,i}^{+(t)} \left[\ln \alpha_k^+ - \frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln \sigma_k^{+2} \right. \\ &\quad \left. - \frac{(X_i - \mu_k^+)^2}{2\sigma_k^{+2}} \right], \end{aligned} \quad (\text{A.1})$$

We must maximize $Q(\theta_+)$ taking into account the constraint $\sum_{k=1}^{K_+} \alpha_k^+ = 1$. If we add a Langrange Multiplier, we get:

$$\begin{aligned} L(\theta_+) &= \sum_{i=1}^{N_+} \sum_{k=1}^{K_+} \rho_{k,i}^{+(t)} \left[\ln \alpha_k^+ - \frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln \sigma_k^{+2} \right. \\ &\quad \left. - \frac{(X_i - \mu_k^+)^2}{2\sigma_k^{+2}} \right] - \lambda \left(\sum_{k=1}^{K_+} \alpha_k^+ - 1 \right). \end{aligned} \quad (\text{A.2})$$

Setting the derivative $\partial L(\theta_+)/\partial \alpha_k^+$ to zero, we find, for $k = 1, \dots, K_+$

$$\alpha_k^{+(t)} = \frac{1}{\lambda} \sum_{i=1}^{N_+} \rho_{k,i}^{+(t)} \quad (\text{A.3})$$

By invoking the fact that $\sum_{k=1}^{K_+} \alpha_k^+ = 1$ and that $\sum_{k=1}^{K_+} \rho_{k,i}^{+(t)} = 1$, it follows from (A.3) that $\lambda = N_+$. Equation (12) is then obtained. \square

B. Proof of (13)

Proof. Let us use the conditional Expectation of the log likelihood function given in Appendix A, (A.1). Setting the derivative $\partial Q(\theta_+)/\partial \mu_k^+$ to zero, we find, for $k = 1, \dots, K_+$

$$\begin{aligned} \frac{\partial Q(\theta_+)}{\partial \mu_k^+} &= \sum_{i=1}^{N_+} \rho_{k,i}^{+(t)} \left[+2 \frac{(X_i - \mu_k^+)}{2\sigma_k^{+2}} \right] \\ &= \sum_{i=1}^{N_+} \rho_{k,i}^{+(t)} X_i - \mu_k^+ \sum_{i=1}^{N_+} \rho_{k,i}^{+(t)} \\ &= 0 \end{aligned} \quad (\text{B.1})$$

Then, for $k = 1, \dots, K_+$, we have,

$$\mu_k^{+(t)} = \frac{\sum_{i=1}^{N_+} \rho_{k,i}^{+(t)} X_i}{\sum_{i=1}^{N_+} \rho_{k,i}^{+(t)}}. \quad (\text{B.2})$$

\square

C. Proof of (14)

Proof. Let us use the conditional Expectation of the log likelihood function given in Appendix A,(A.1). Setting the derivative $\partial Q(\theta_+)/\partial \sigma_k^{+2}$ to zero, we find, for $k = 1, \dots, K_+$

$$\begin{aligned} \frac{\partial Q(\theta_+)}{\partial \sigma_k^{+2}} &= \sum_{i=1}^{N_+} \rho_{k,i}^{+(t)} \left[-\frac{1}{2\sigma_k^{+2}} + \frac{(X_i - \mu_k^+)^2}{2\sigma_k^{+4}} \right] \\ &= -\frac{1}{2\sigma_k^{+2}} \sum_{i=1}^{N_+} \rho_{k,i}^{+(t)} + \frac{1}{2\sigma_k^{+4}} \sum_{i=1}^{N_+} \rho_{k,i}^{+(t)} (X_i - \mu_k^+)^2 \\ &= 0 \end{aligned} \quad (\text{C.1})$$

Then, for $k = 1, \dots, K_+$, we have,

$$\sigma_k^{+(t)2} = \frac{\sum_{i=1}^{N_+} \rho_{k,i}^{+(t)} (X_i - \mu_k^{+(t)})^2}{\sum_{i=1}^{N_+} \rho_{k,i}^{+(t)}}. \quad (\text{C.2})$$

\square

D. Proof of (15)

Proof. By using the expression of the Bit Error Probability(4) and the two conditional pdfs estimates, $\hat{f}_{X_+,N_+}^{b_+}$ and $\hat{f}_{X_-,N_-}^{b_-}$, we can express the BER estimate as,

$$\hat{p}_{e,N} = \pi_+ \underbrace{\int_{-\infty}^0 \hat{f}_{X_+,N_+}^{b_+}(x) dx}_{D_+} + \pi_- \underbrace{\int_0^{+\infty} \hat{f}_{X_-,N_-}^{b_-}(x) dx}_{D_-}. \quad (\text{D.1})$$

Given the fact that the two conditional pdfs are estimated using the Gaussian Mixture based pdf estimate, by EM algorithm, according to (6), and using the following change

of variables, $t = (x - \mu_k^+)/\sigma_k^+$ (resp., $t = (x - \mu_k^-)/\sigma_k^-$), for D_+ (resp., D_-), we get,

$$\begin{aligned}
 D_+ &= \int_{-\infty}^0 \hat{f}_{X,N_+}^{b_+}(x) dx \\
 &= \int_{-\infty}^0 \sum_{k=1}^{K_+} \alpha_k^+ f_k^+(x; \mu_k^+, \sigma_k^{+2}) dx \\
 &= \sum_{k=1}^{K_+} \alpha_k^+ \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} e^{-(x-\mu_k^+)^2/2\sigma_k^{+2}} dx \\
 &= \sum_{k=1}^{K_+} \alpha_k^+ \int_{-\infty}^{-\mu_k^+/\sigma_k^+} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dx \\
 &= \sum_{k=1}^{K_+} \alpha_k^+ \int_{\mu_k^+/\sigma_k^+}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dx \\
 &= \sum_{k=1}^{K_+} \alpha_k^+ Q\left(\frac{\mu_k^+}{\sigma_k^+}\right),
 \end{aligned} \tag{D.2}$$

$$\begin{aligned}
 D_- &= \int_0^{+\infty} \hat{f}_{X,N_-}^{b_-}(x) dx \\
 &= \int_0^{+\infty} \sum_{k=1}^{K_-} \alpha_k^- f_k^-(x; \mu_k^-, \sigma_k^{-2}) dx \\
 &= \sum_{k=1}^{K_-} \alpha_k^- \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-\mu_k^-)^2/2\sigma_k^{-2}} dx \\
 &= \sum_{k=1}^{K_-} \alpha_k^- \int_{-\mu_k^-/\sigma_k^-}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dx \\
 &= \sum_{k=1}^{K_-} \alpha_k^- Q\left(-\frac{\mu_k^-}{\sigma_k^-}\right),
 \end{aligned} \tag{D.3}$$

The BER estimate given by (15) is simply obtained by combining (D.1), (D.2) and (D.3). \square

References

- [1] 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, no. 1, pp. 153–170, 1984.
- [2] K. S. Shanmugam and P. Balaban, "A modified Monte Carlo simulation technique for the evaluation of error rate in digital communication systems," *IEEE transactions on Communications Systems*, vol. 28, no. 11, pp. 1916–1924, 1980.
- [3] E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, NY, USA, 1958.
- [4] A. Abedi, M. E. Thompson, and A. K. Khandani, "Application of cumulant method in performance evaluation of turbo-like codes," in *Proceedings of the IEEE International Conference on Communications (ICC '07)*, pp. 986–989, June 2007.
- [5] E. Cavus, C. L. Haymes, and B. Daneshrad, "Low BER performance estimation of LDPC codes via application of importance sampling to trapping sets," *IEEE Transactions on Communications*, vol. 57, no. 7, pp. 1886–1888, 2009.
- [6] T. Sakai and K. Shibata, "Fast BER estimation of LDPC codes," in *Proceedings of the International ITG conference on Source and Channel Coding*, pp. 1–6, Siegen, Germany, January 2010.
- [7] A. Mahadevan and J. M. Morris, "SNR-invariant importance sampling for hard-decision decoding performance of linear block codes," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 100–111, 2007.
- [8] L. Dong, W. Wenbo, and L. Yonghua, "Monte Carlo simulation with error classification for multipath Rayleigh fading channel," in *Proceedings of the 16th International Conference on Telecommunications (ICT '09)*, pp. 223–227, 2009.
- [9] S. Saoudi, M. Troudi, and F. Ghorbel, "An iterative soft bit error rate estimation of any digital communication systems using a nonparametric probability density function," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 512192, 9 pages, 2009.
- [10] Z. R. Yang and M. Zwoilinski, "Mutual information theory for adaptive mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 4, pp. 396–403, 2001.
- [11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B*, vol. 39, no. 1, pp. 1–38, 1977.
- [12] P. Masson and W. Pieczynski, "SEM algorithm and unsupervised statistical segmentation of satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 31, no. 3, pp. 618–633, 1993.
- [13] N. Bouguila and D. Ziou, "A hybrid SEM algorithm for high-dimensional unsupervised learning using a finite generalized Dirichlet mixture," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2657–2668, 2006.
- [14] S. Saoudi, F. Ghorbel, and A. Hillion, "Non-parametric probability density function estimation on a bounded support: applications to shape classification and speech coding," *Applied Stochastic Models and Data Analysis*, vol. 10, no. 3, pp. 215–231, 1994.
- [15] S. Saoudi, F. Ghorbel, and A. Hillion, "Some statistical properties of the kernel-diffeomorphism estimator," *Applied Stochastic Models and Data Analysis*, vol. 13, no. 1, pp. 39–58, 1997.
- [16] L. Younjeong, L. Ki Yong, and L. Joohun Lee, "The estimating optimal number of Gaussian mixtures based on incremental k-means for speaker identification," *International Journal of Information Technology*, vol. 12, no. 7, pp. 13–21, 2006.
- [17] M. Rosenblatt, "Remarks on some non-parametric estimates of a density function," *The Annals of Mathematical Statistics*, vol. 27, no. 3, pp. 832–837, 1956.
- [18] E. Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.
- [19] R. Kronmal and M. Tarter, "The estimation of probability densities and cumulatives by Fourier series methods," *Journal of the American Statistical Association*, vol. 63, pp. 925–952, 1968.
- [20] S. Saoudi, T. Ait-Idir, and Y. Mochida, "An unsupervised and nonparametric iterative soft bit error rate estimation," submitted to *IEEE Transactions on Communications ICC2011*, 5–9 June, Kyoto, Japan.
- [21] D. Titterington, A. Smith, and U. Makov, *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons, New York, NY, USA, 1985.

- [22] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.
- [23] S. Verdu, *Multuser Detection*, Cambridge University Press, Cambridge, UK, 1998.