Unsupervised and Non Parametric Iterative Soft Bit Error Rate Estimation for Any Communications System

Samir Saoudi, Member, IEEE, Tarik Ait-Idir, Member, IEEE, and Yukou Mochida, Fellow, IEEE

Abstract

This paper addresses the problem of unsupervised soft bit error rate (BER) estimation for any communications system, where no prior knowledge either about transmitted information bits, or the transceiver scheme is available. We show that the problem of BER estimation is equivalent to estimating the conditional probability density functions (pdf)s of soft channel/receiver outputs. Assuming that the receiver has no analytical model of soft observations, we propose a non parametric Kernel-based pdf estimation technique, and show that the resulting BER estimator is asymptotically unbiased and point-wise consistent. We then introduce an iterative Stochastic Expectation Maximization (EM) algorithm for the estimation of both a priori and a posteriori probabilities of transmitted information bits, and the classification of soft observations according to transmitted bit values. These inputs serve in the iterative Kernel-based estimation procedure of conditional pdfs. We analyze the performance of the proposed unsupervised and non parametric BER estimator in the framework of a multiuser code division multiple access (CDMA) system with single user detection, and show that attractive performance are achieved compared with conventional Monte Carlo (MC)-aided techniques.

Index Terms

Unsupervised soft BER estimation, MC simulation, Stochastic EM algorithm, Non parametric pdf estimation, Kernel method.

S. Saoudi is with the Signal and Communications department, Institut Telecom / Telecom Bretegne/LabSticc, Brest, France. He is also with Université Européenne de Bretagne.(email: samir.saoudi@telecom-bretagne.eu).

T. Ait-Idir is with the Communications Systems Department, INPT, Madinat Al-Irfane, Rabat, Morocco. He is also with Signal and Communications department, Institut Telecom / Telecom Bretegne/LabSticc, Brest, France.

Y. Mochida is with Orange-Labs, Tokyo, Japan.
I. Introduction

Monte Carlo (MC) techniques are generally used to evaluate the bit error rate (BER) or block error rate (BLER) of digital communication systems. In [1], a tutorial exposition of different MC-aided techniques has been provided, with particular reference to four methods: i) Modified Monte Carlo simulation (i.e., importance sampling); ii) extreme value theory; iii) tail extrapolation; and iv) quasi-analytical method. The modified Monte Carlo technique is achieved by importance sampling which means that important events, i.e., errors, are artificially generated by biasing the noise process. At the end of the simulation, the error count must be properly unbiased [2]. The extreme value theory [3], assumes that the probability density function (pdf) can be approximated by exponential functions. The tail extrapolation method, which is a subclass of the extreme value theory technique, is based on the assumption that the tail region of the pdf can be described by a generalized exponential class. The quasi-analytical method combines noiseless simulation with analytical representation of the noise. Except for the MC method, all these techniques assume perfect knowledge of the type and/or the form of noise statistics. In [4], high order statistics (HOS) of the bit log-likelihood-ratio (LLR) are used to evaluate the error rate performance of turbo-like codes. In this case, the characteristic function of the bit LLR is estimated using its first cumulants, i.e., moments, and the pdf is deduced with the aid of the inverse Fourier transform (IFT). In [5], a BER estimation technique where the transmitter sends a fixed information bit value has been proposed. At the receiver side, the BER is computed by estimating the pdf of received soft channel/receiver outputs. This technique is called soft BER estimation since the BER is estimated using soft channel observations/outputs without requiring hard decisions about information bits 1. Unfortunately, all these methods assume that the estimator perfectly knows transmitted data.

In many practical communication systems, the BER is required to be on-line estimated in order to perform system-level functions such as scheduling, resource allocation, power control, or link adaptation where the transmission scheme is adapted to the channel conditions (See for instance [6]). Under this framework, the BER estimation problem becomes very challenging because of the following main reasons:

- In MC-based techniques, the unknown transmitted information bit values are required for computing the BER estimate, while in practical communications systems the BER estimation

---

1 In soft BER estimation, soft observations are used instead of hard decisions. This provides reliable BER estimates and reduces the number of required samples/observations compared with hard decision-based BER estimation techniques [5].
should be performed in an unsupervised fashion because the estimator has no information about transmitted data.

- Most of practical communication channels are quasi-static block fading, i.e., constant over a certain block duration, and randomly changes from block to block. Therefore, in order to provide the transmitter with a reliable BER information feedback for a given block, i.e., channel state information (CSI), only the soft observations corresponding to the actual block have to be used for estimating the instantaneous BER. In the case of MC-based techniques, this results in unreliable or even wrong BER estimates because the number of observations/bits is not sufficient.

- The knowledge of the transmitter scheme, channel and interference model, and receiver technique greatly impacts the reliability of the BER estimate. In other words, it is generally quite hard to derive analytical expressions of the bit error probability (BEP) when the system suffers from interference or in the case of non-linear receivers (e.g. iterative interference cancellation receivers).

This paper was mainly motivated by the above considerations. We assume that the estimator has no knowledge either about transmitted data, or the transmitter/receiver scheme and the communication model. We focus on the problem of BER estimation in a completely unsupervised fashion. To the best of the authors knowledge, this issue has not been addressed before in the literature. We first provide a problem formulation where we show that BER estimation is equivalent to estimating the pdfs of conditional soft observations corresponding to transmitted information bits. Then, we introduce a non parametric Gaussian Kernel-based pdf estimation technique where no analytical model of soft observations is assumed. We study the asymptotic behavior of the resulting BER estimator and show that it is unbiased and point-wise consistent. As pdf estimation requires the knowledge of both a priori probabilities of information bits and the classification of soft observations according to the $+1$ and $-1$ values of transmitted bits, we introduce an iterative Stochastic Expectation Maximization (EM) algorithm for iteratively computing these parameters. We analyze the performance of the proposed unsupervised technique in the case of a multiuser code division multiple access (CDMA) system with conventional single-user detection. Performance comparison shows that the proposed estimator clearly outperforms supervised MC-aided estimation. Interestingly, reliable BER estimates are achieved even in the high signal to noise ratio (SNR) and using only a few number of soft observations.

Throughout the paper, the following notation is used. The cardinality $N$ of set $C$ is denoted $N = |C|$.

2The aim of unsupervised estimation is to estimate the BER based only on soft observations that serve for computing hard decisions about information bits without requiring any information about transmitted data and transceiver scheme/model.
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 derivative 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 remainder of the paper is organized as follows. In Section II, we provide some preliminaries about the proposed BER estimator and detail soft output pdf estimation; In Section III, we introduce the Gaussian Kernel-based BER estimator and study its asymptotic behavior; Section IV details the Stochastic EM iterative algorithm; In Section V, we carry out performance evaluation; The paper is concluded in Section VI. The proofs of the results are provided in the Appendixes A–E.

II. PRELIMINARIES AND OUTPUT PDF ESTIMATION

A. System Model

We consider a general communication system where a sequence of $N$ independent and identically distributed (i.i.d) information bits $(b_i)_{1 \leq i \leq N} \in \{+1, -1\}$ is transmitted using any transmission scheme. The transmitter can use any type of channel coding, modulation, multiple access method (e.g., CDMA, orthogonal frequency division multiplexing (OFDM), etc...), and/or multiple input multiple output (MIMO) technique such as spatial multiplexing, or space-time coding. The communication channel can either be time-variant or invariant with either flat or frequency selective fading. At the receiver side, any reception technique is performed ranging from the very simple matched filter (MF) receiver to advanced reception schemes such as multiuser detection (MUD) [7], interference cancellation-aided turbo detection/equalization [8], and space-time turbo equalization (See for instance [9]–[12] and references therein).

Let $X$ denote the random variable (RV) corresponding to channel/receiver soft outputs, i.e., observations, and $(X_i)_{1 \leq i \leq N}$ be the sequence of the realizations of $X$ where each $X_i$ corresponds to the decision statistic that serves for the computation of hard decision $\hat{b}_i$ about transmitted information bit $b_i$ as $\hat{b}_i = \text{sgn}(X_i)$. Note that $X_i$ may contain any type of interference such as co-channel interference (CCI), intersymbol interference (ISI), and multiple antenna interference (MAI), etc... . We assume no knowledge either about the transmission scheme, or about the channel model and the receiver technique. Our main purpose is to estimate in an unsupervised fashion the BER at the output of the receiver. Let $\pi_+$ and $\pi_-$ denote the probability that the transmitted bit $b_i$ is equal to $+1$ and $-1$, respectively, i.e.,
\[
\begin{aligned}
\pi_+ & \triangleq P[b_i = +1], \\
\pi_- & \triangleq P[b_i = -1],
\end{aligned}
\]  
(1)

where \( \pi_+ + \pi_- = 1 \). Note that \( \pi_+ \) and \( \pi_- \) are not known at the receiver. The soft outputs \((X_i)_{1\leq i\leq N}\) are random variables having the same pdf \( f_X(x) \). The BEP is then given by

\[
p_e = \pi_+ \int_{-\infty}^{0} f_{X}^{b_+}(x) \, dx + \pi_- \int_{0}^{\infty} f_{X}^{b_-}(x) \, dx,
\]

(2)

where \( f_{X}^{b_+}(\cdot) \) (respectively, \( f_{X}^{b_-}(\cdot) \)) is the conditional pdf of \( X \) such that \( b_i = +1 \) (respectively, \( b_i = -1 \)).

\( f_X(x) \) is a mixture of the two conditional pdfs \( f_{X}^{b_+}(x) \) and \( f_{X}^{b_-}(x) \) and can therefore be written as

\[
f_X(x) = \pi_+ f_{X}^{b_+}(x) + \pi_- f_{X}^{b_-}(x).
\]

A general transmitter–receiver scheme with soft outputs \((X_i)_{1\leq i\leq N}\) and hard decisions \((\hat{b}_i)_{1\leq i\leq N}\) is depicted in Fig. 1.

B. Brief Overview of the Proposed BER Estimator

As it can be seen from the generic expression of the BEP \( p_e \) in (2), the BER can be estimated using the two conditional pdfs \( f_{X}^{b_+}(\cdot) \), \( f_{X}^{b_-}(\cdot) \), and the \textit{a priori} probabilities \( \pi_+ \) and \( \pi_- \). Note that all these parameters are unknown, and are required to be estimated before computing the BER. The estimation of conditional pdfs \( f_{X}^{b_+}(\cdot) \) and \( f_{X}^{b_-}(\cdot) \) can be performed based on observations \( X_1, \ldots, X_N \), where only those corresponding to transmitted information bit value +1 (respectively, −1) are used to estimate \( f_{X}^{b_+}(\cdot) \) (respectively, \( f_{X}^{b_-}(\cdot) \)). Unfortunately, the problem of classifying soft outputs \( X_1, \ldots, X_N \) according to transmitted information bit values is itself a detection problem. The unsupervised BER estimator we propose in this paper allows us to classify soft outputs \( X_1, \ldots, X_N \) and estimate both the conditional pdfs \( f_{X}^{b_+}(\cdot) \) and \( f_{X}^{b_-}(\cdot) \) and \textit{a priori} probabilities \( \pi_+ \) and \( \pi_- \) in an iterative fashion using \textit{a posteriori} probability (APP) values \( P[b_i = +1 \mid X_i] \) and \( P[b_i = -1 \mid X_i] \) \( \forall \, i = 1, \ldots, N \).

Let \( T \) denote the number of iterations (index \( t = 1, \ldots, T \)). At each iteration \( t \), all APP values are computed using \textit{a priori} probabilities and pdf estimates obtained at the previous iteration \( t - 1 \). The resulting APPs allow us to update the estimates of \textit{a priori} probabilities and classify soft outputs \( X_1, \ldots, X_N \) into two classes according to transmitted information bit values. The two conditional pdfs are then re-estimated. This process is repeated \( T \) times, and the BER is estimated with the aid
of parameter estimates obtained at the last iteration $T$. Note that, at the first iteration, since no prior information is available, the classification of soft outputs $X_1, \ldots, X_N$ can be performed with respect to the sign of each observation $X_i$, while the initial values of a priori probabilities $\pi_+$ and $\pi_-$ can be derived using this initial classification. The diagram of the proposed iterative BER estimation algorithm is depicted in Fig. 2.

C. Output PDF Estimation

Both conditional pdfs $f_{X_+}^{b+}(x)$ and $f_{X_-}^{b-}(x)$ depend on the communication channel model and receiver scheme. Therefore, it is extremely difficult or even infeasible to find out the exact parametric model of these distributions. In this subsection, we introduce a Kernel-based method [13]–[15] to estimate both pdfs $f_{X_+}^{b+}(x)$ and $f_{X_-}^{b-}(x)$. The proposed technique is non-parametric, and only requires soft outputs for each class.

Let us assume that we can classify the set $C = \{X_1, \ldots, X_N\}$ into two classes (i.e., partitions) $C_+$ and $C_-$. Let $C_+$ (respectively, $C_-$) contains the observed received soft output $X_i$ such that the corresponding transmitted bit is $b_i = +1$ (respectively, $b_i = -1$). In Section IV, we will introduce a Stochastic EM-based algorithm that allows us to classify the sets $C_+$ and $C_-$. Let $N_+$ (respectively, $N_-\,)$ denote the cardinality of $C_+$ (respectively, $C_-$), i.e., $N_+ = |C_+|$, and $N_- = |C_-|$. Using the Kernel technique, it follows that each conditional distribution can be estimated using the elements of sets $C_+$ and $C_-$ as,

$$\hat{f}_{X,N_+}^{b+}(x) = \frac{1}{N_+ h_{N_+}} \sum_{X_i \in C_+} K\left(\frac{x - X_i}{h_{N_+}}\right),$$

$$\hat{f}_{X,N_-}^{b-}(x) = \frac{1}{N_- h_{N_-}} \sum_{X_i \in C_-} K\left(\frac{x - X_i}{h_{N_-}}\right),$$

where $h_{N_+}$ (respectively, $h_{N_-}$) is a smoothing parameter which depends on the number of observed samples, i.e., $N_+$ (respectively, $N_-\,)$, Function $K(\cdot)$ is any pdf (called the Kernel) assumed to be an even and regular (i.e., square integrated) function with zero mean and unit variance. In the following, we provide equations only in the case of class $C_+$ as those corresponding to $C_-$ can easily be obtained by replacing “+” by “−”.

The choice of the optimal smoothing parameter is very critical since it impacts the accuracy of the pdf estimate. In [16], [17], it has been shown that if $h_{N_+}$ tends towards 0 when $N_+$ tends towards $+\infty$, the estimator $\hat{f}_{X,N_+}^{b+}(x)$ is asymptotically unbiased. It has also been shown that if $h_{N_+} \to 0$ and
$N_+ h_{N_+} \to +\infty$ when $N_+ \to +\infty$, then the MSE of the Kernel estimator tends towards zero. The optimal integrated mean squared error (IMSE)-based smoothing parameter $h_{N_+}^*$ is given as [5],

$$h_{N_+}^* = N_+^{-\frac{1}{5}} \left( J \left( f_X^{b_+} \right) \right)^{-\frac{2}{5}} \left( M(K) \right)^{\frac{1}{5}},$$

(6)

where $M(K) = \int_{-\infty}^{+\infty} K^2(x) \, dx$, and $J(f_X^{b_+}) = \int_{-\infty}^{+\infty} \{(f_X^{b_+})'x\}^2 \, dx$ is the integrated square second derivative of the conditional pdf $f_X^{b_+}$.

As it can be seen from (6), the computation of $h_{N_+}^*$ unfortunately depends on the unknown pdf $f_X^{b_+}$. We therefore suggest to use the Gaussian Kernel whose parameter $M(K)$ is given as [5]

$$M(K) = \frac{1}{2\sqrt{\pi}},$$

(7)

while $J(f_X^{b_+})$ can be expressed in the case of a Gaussian conditional distribution, i.e., $f_X^{b_+} \sim \mathcal{N}(m_+, \sigma_+^2)$, as

$$J(f_X^{b_+}) = \frac{3}{8\sqrt{\pi}\sigma_+^5},$$

(8)

where $m_+$ and $\sigma_+^2$ are the mean and variance of subset $C_+$, respectively. Details regarding the derivation of (8) can be found in Appendix I. It follows from (6) that the optimal smoothing parameter in the case of a Gaussian Kernel, and a Gaussian conditional pdf is given as,

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

(9)

### III. BER Estimation

#### A. Gaussian Kernel-Based BER Estimation

In this subsection, we derive the expression of the BER estimate assuming Gaussian Kernel-based pdf estimator. Let $\theta \triangleq (\pi_+, N_+, h_{N_+}, \pi_-, N_-, h_{N_-})$. The value of $\theta$ is unknown and is iteratively estimated by iteratively classifying soft outputs $X_1, \cdots, X_N$ into two classes. At the last iteration $T$, a reliable estimate of $\theta$ is reached and the BER is computed using the obtained estimate. Let us recall the expression of the BEP (2). Replacing the two conditional pdfs by their Gaussian Kernel-based estimates (4) and (5), and given the value of $\theta$, the BER estimate is simply computed as

$$\hat{p}_{e, N} = \frac{\pi_+}{N_+} \sum_{X_i \in C_+} Q \left( \frac{X_i}{h_{N_+}} \right) + \frac{\pi_-}{N_-} \sum_{X_i \in C_-} Q \left( -\frac{X_i}{h_{N_-}} \right),$$

(10)

where $Q(\cdot)$ denotes the complementary unit cumulative Gaussian distribution, i.e., $Q(x) = \int_{x}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp \left( -t^2/2 \right) dt$. Details regarding the derivation of (10) are provided in Appendix II.
B. Theoretical Analysis

In this subsection, we study the MSE-based convergence of the proposed iterative BER estimator. The first theorem shows that the proposed estimator is asymptotically unbiased.

**Theorem 1:** Assume that the two conditional pdfs, \( f_{b+}^{b+} \) and \( f_{b-}^{b-} \), are second derivative functions, that \( h_{N+} \to 0 \) and \( h_{N-} \to 0 \) as \( N \to +\infty \). Then the soft BER estimator \( \hat{p}_{e,N} \) given by (10) is asymptotically unbiased, i.e.,

\[
\lim_{N \to +\infty} E[\hat{p}_{e,N}] = p_e.
\]

**Proof:** See Appendix III for the proof.

The following theorem shows that the variance of the proposed estimator also tends towards zero.

**Theorem 2:** Assume that the two conditional pdfs, \( f_{b+}^{b+} \) and \( f_{b-}^{b-} \), are second derivative functions, that \( h_{N+} \to 0 \) and \( h_{N-} \to 0 \) as \( N \to +\infty \). Then the variance of the soft BER estimator \( \hat{p}_{e,N} \) tends towards zero as \( N \) tends towards \(+\infty\), i.e.,

\[
\lim_{N \to +\infty} E[(\hat{p}_{e,N} - E[\hat{p}_{e,N}])^2] = 0.
\]

**Proof:** See Appendix IV for the proof.

Using Theorems 1 and 2, it is easy to show that the proposed estimator is point-wise consistent, i.e., the MSE tends towards zero as the number of samples \( N \) tends towards \(+\infty\). This result can be formulated in the following corollary:

**Corollary 1:** Assume that the two conditional pdfs, \( f_{b+}^{b+} \) and \( f_{b-}^{b-} \), are second derivative functions, that \( h_{N+} \to 0 \) and \( h_{N-} \to 0 \) as \( N \to +\infty \). Then the MSE of the soft BER estimator \( \hat{p}_{e,N} \) tends towards zero as \( N \) tends towards \(+\infty\), i.e.,

\[
\lim_{N \to +\infty} E[(\hat{p}_{e,N} - p_e)^2] = 0.
\]

**Proof:** The proof of this corollary is straightforward. First of all, we write \( E[(\hat{p}_{e,N} - p_e)^2] = E[(\hat{p}_{e,N} - E[\hat{p}_{e,N}])^2] + (E[\hat{p}_{e,N}] - p_e)^2 \). Then, by noting that \( \hat{p}_{e,N} \) is asymptotically unbiased, i.e., \( E[\hat{p}_{e,N}] - p_e \to 0 \) (see Theorem 1), and that the variance of \( \hat{p}_{e,N} \) tends towards 0 as \( N \to +\infty \) (see Theorem 2), we get \( \lim_{N \to +\infty} E[(\hat{p}_{e,N} - p_e)^2] = 0 \).

IV. STOCHASTIC EM-BASED PARAMETER ESTIMATION

In this section, we introduce a Stochastic EM-based algorithm to iteratively classify soft outputs \( X_1, \cdots, X_N \) into the two classes \( C_+ \) and \( C_- \), and estimate \( \theta \). The EM algorithm was first introduced by Dempster et. al. in [18]. It iteratively computes, with the aid of the estimation and maximization
steps, maximum likelihood (ML) estimates when the observations can be viewed as incomplete data
that contain missing information (i.e., the unknown data we have to decide about). Stochastic EM
presents a generalization of EM techniques by introducing a random rule in the EM algorithm. Some
applications of Stochastic EM can be found in [19], [20]. In our case, as we want to estimate the
BER in an unsupervised fashion, the incomplete data correspond to observations \( X_1, \cdots, X_N \), while
the missing data are transmitted information bits \( b_1, \cdots, b_N \).

A. Estimation Step

In the estimation step of iteration \( t \), we estimate the APPs

\[
\begin{align*}
\rho_{+}^{(t)} & \triangleq P[b_i = +1 \mid X_i, \theta^{(t-1)}], \\
\rho_{-}^{(t)} & \triangleq P[b_i = -1 \mid X_i, \theta^{(t-1)}],
\end{align*}
\]

of unobserved information bits \( b_i \), for \( i = 1, \cdots, N \), conditioned on observations \( X_1, \cdots, X_N \) and
the estimate \( \theta^{(t-1)} \) of \( \theta \) obtained at the maximization step of previous iteration \( t - 1 \). Using simple
mathematical manipulations, we show that the likelihood probabilities of bit \( b_i \) at iteration \( t \) can be
computed as,

\[
\begin{align*}
\tilde{P}_{+}^{(t)} & = \frac{\pi_+^{(t-1)} f_{X,N_i}^{+}(X_i)}{\pi_+^{(t-1)} f_{X,N_i}^{+}(X_i) + \pi_-^{(t-1)} f_{X,N_i}^{-}(X_i)}, \\
\tilde{P}_{-}^{(t)} & = \frac{\pi_-^{(t-1)} f_{X,N_i}^{-}(X_i)}{\pi_+^{(t-1)} f_{X,N_i}^{+}(X_i) + \pi_-^{(t-1)} f_{X,N_i}^{-}(X_i)},
\end{align*}
\]

Note that for the sake of notation simplicity, the iteration index is dropped from pdf estimates
\( f_{X,N_i}^{+}(X_i) \) and \( f_{X,N_i}^{-}(X_i) \).

B. Maximization Step

At iteration \( t \), the maximization step allows us to compute the estimate \( \theta^{(t)} \) based on conditional
probabilities obtained at the estimation step of the same iteration \( t \). The estimate \( \theta^{(t)} \) is obtained by
maximizing the conditional expectation \( Q(\theta^{(t)}) \) of the log-likelihood of the joint event at iteration \( t \).
Assuming independent soft observations \( X_1, \cdots, X_N \), we can express \( Q(\theta^{(t)}) \) as

\[
Q(\theta^{(t)}) = \mathbb{E} \left[ \ln \left( \prod_{i=1}^{N} P[X_i, b_i] \mid X_1, \cdots, X_N \right) \right],
\]

\[
= \sum_{i=1}^{N} \left\{ \rho_{+}^{(t)} \ln \left\{ \pi_+^{(t)} f_{X,N_i}^{+}(X_i) \right\} + \rho_{-}^{(t)} \ln \left\{ \pi_-^{(t)} f_{X,N_i}^{-}(X_i) \right\} \right\}. \tag{13}
\]
By invoking the fact that $\pi_+^{(t)} + \pi_-^{(t)} = 1$, the maximization of $Q(\theta)$ leads to the new estimates of a priori probabilities at iteration $t$,

$$
\begin{align*}
\pi_+^{(t)} &= \frac{1}{N} \sum_{i=1}^{N} \rho_{i+}^{(t)}, \\
\pi_-^{(t)} &= \frac{1}{N} \sum_{i=1}^{N} \rho_{i-}^{(t)}.
\end{align*}
$$

(14)

The proof of (14) is provided in Appendix V.

C. Classification Step

The remaining parameters $N_+^{(t)}$, $h_{N+}^{(t)}$, $N_-^{(t)}$, $h_{N-}^{(t)}$ and the two conditional pdf estimates $\hat{f}_{X, N+}^{(t)}(\cdot)$ and $\hat{f}_{X, N-}^{(t)}(\cdot)$ depend on the outcome of the classification procedure of subsets $C_+$ and $C_-$ at iteration $t$. Therefore, this procedure should be carefully performed since it greatly impacts the reliability of both pdf estimates and a posteriori probabilities in subsequent iterations, and consequently the accuracy of the BER estimate. Given estimates $\theta^{(t-1)}$, $\hat{f}_{X, N+}^{(t-1)}(\cdot)$, and $\hat{f}_{X, N-}^{(t-1)}(\cdot)$ available at iteration $t$, we can classify soft outputs $X_1, \cdots, X_N$ according to joint probabilities $P[X_i, b_i = +1]$, and $P[X_i, b_i = -1]$. Using Bayes rule, we can obtain subsets $C_+^{(t)}$ and $C_-^{(t)}$ at iteration $t$ as $C_+^{(t)} = \{X_i : \pi_+^{(t-1)} \hat{f}_{X, N+}^{(t-1)}(X_i) \geq \pi_-^{(t-1)} \hat{f}_{X, N-}^{(t-1)}(X_i)\}$, and $C_-^{(t)} = C_+^{(t)}$.

Unfortunately, this classification procedure will prevent some erroneous soft outputs (i.e., those positive soft outputs corresponding to transmitted information bit value $-1$ and vise versa) from being exchanged in the course of iterations between subsets $C_+$ and $C_-$. In the following, we introduce a Stochastic EM-based algorithm that randomly performs the classification of soft outputs $X_1, \cdots, X_N$ using APP values. This method relaxes the condition on the exchange of soft outputs between the two subsets $C_+$ and $C_-$. The Stochastic EM technique uses a random Bayesian rule. At iteration $t$, it combines likelihood probabilities (15) and realizations $U_1^{(t)}, \cdots, U_N^{(t)}$ of a uniform random variable $U$ defined over the interval $[0, 1]$. The classification of soft outputs is performed as follows,

$$
\begin{align*}
C_+^{(t)} &= \big\{ X_i : \rho_{i+}^{(t)} \geq U_i^{(t)} \big\}, \\
C_-^{(t)} &= C_+^{(t)}.
\end{align*}
$$

(15)

Parameters $N_+^{(t)}$ and $N_-^{(t)}$ are simply obtained as $N_+^{(t)} = |C_+^{(t)}|$, and $N_-^{(t)} = |C_-^{(t)}|$. The optimal smoothing parameters $h_{N+}^{(t)}$ and $h_{N-}^{(t)}$ are computed either with the aid of (9) assuming Gaussian conditional pdfs, or by iteratively using the exact expressions of $J(\hat{f}_{X+}^{(t)})$ and $J(\hat{f}_{X-}^{(t)})$ followed by (9). The two conditional pdfs are then estimated using (4) and (5). Note that $\sigma_{+}^{(t)}$ and $\sigma_{-}^{(t)}$ correspond to the standard-deviation
of subsets $C_+^{(t)}$ and $C_-^{(t)}$ at iteration $t$.

The proposed Stochastic EM-based unsupervised BER estimation algorithm can now be summarized as in Table I.

V. PERFORMANCE EVALUATION

A. Considered Framework

To evaluate the performance of the proposed unsupervised and non parametric BER estimator, 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

$$r_i = A_1 b_1^{(i)} s_1 + A_2 b_2^{(i)} s_2 + n_i,$$

where $L_{SF}$ denotes the spreading factor, and $s_k \in \{ \pm 1/\sqrt{L_{SF}} \}^{L_{SF}}$ is the spreading code corresponding to user $k$. $A_k$ is the amplitude of user $k = 1, 2$, $b_k^{(i)}$ is the information bit value $\in \{ \pm 1 \}$ of user $k$ at time instant $i$, and $n_i \in \mathbb{R}^{L_{SF}}$ is the temporally and spatially white Gaussian noise, i.e., $n_i \sim \mathcal{N}(0, \sigma^2 I_{L_{SF}})$. The a priori probabilities of information bits are supposed to be identical for both users, i.e., $\pi_+ = P\left[ b_k^{(i)} = +1 \right]$ and $\pi_- = P\left[ b_k^{(i)} = -1 \right] \forall k, i$.

The decision statistic that serves for detecting user 1 at time instant $i$ is $X_i^{(1)} = s_1^T r_i$ [7] and is given as,

$$X_i^{(1)} = A_1 b_1^{(i)} + A_2 b_2^{(i)} \rho + \tilde{n}_i^{(1)},$$

where $\rho$ 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, i.e., $\tilde{n}_i^{(1)} \sim \mathcal{N}(0, \sigma^2)$. The decision about information bit $b_1^{(i)}$ corresponds to the sign of decision statistic $X_i^{(1)}$, i.e., $\hat{b}_1^{(i)} = sgn \left( X_i^{(1)} \right)$. Note that the soft output $X_i^{(1)}$ in (17) contains a mixture of a Gaussian noise and a RV whose pdf is unknown at the receiver. Using (17), we can easily show that the BEP for user 1 is

$$p_{e_1} = 2\pi_+ \pi_- Q \left( \frac{A_1 - A_2 \rho}{\sigma} \right) + (\pi_+^2 + \pi_-^2) Q \left( \frac{A_1 + A_2 \rho}{\sigma} \right).$$

(18)
In the following, we use the two spreading codes

\[ s_1 = \frac{1}{\sqrt{7}} \begin{bmatrix} + & + & + & - & - & - \end{bmatrix}^T, \]  
\[ s_2 = \frac{1}{\sqrt{7}} \begin{bmatrix} - & - & + & + & - & - \end{bmatrix}^T, \]

where the cross-correlation is \( \rho = 0.4286 \). 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 \( \text{SNR} = 1/2\sigma^2 \). Note that under these assumptions (cross-correlation value and equal powers), soft decisions \( X^{(1)}_1, \ldots, X^{(1)}_N \) will be corrupted by severe MAI. This is extremely challenging for the proposed BER estimator since parameter estimation (such as \textit{a priori} probabilities) will be performed in the presence of strong interference. In all simulations, we consider \( T = 6 \) iterations for the Stochastic EM-based parameter estimation while at each iteration \( t = 1, \ldots, 6 \), two iterations are used for computing the optimal smoothing parameters \( h^*_N \) and \( h^*_N \) as mentioned in Subsection IV-C. Note that for all the scenarios we consider in the following, we assume that the proposed unsupervised and non parametric BER estimator has no prior knowledge either about \textit{a priori} probabilities \( \pi_+ \) and \( \pi_- \) or the classification of soft outputs \( X^{(1)}_1, \ldots, X^{(1)}_N \).

**B. Numerical Results**

1) **Performance for Uniform Sources:** First, we consider the case of equiprobable information bits, i.e., \( \pi_+ = \pi_- = 1/2 \). The number of soft outputs that serve for estimating the BER is \( N = 10^4 \) observations. In Fig. 3 and Fig. 4 we present both the theoretical and the estimated conditional pdfs (at the last iteration \( T = 6 \)) for \( \text{SNR} = 0 \)dB and \( 10 \)dB, respectively. We observe that the proposed BER estimator provides accurate estimates of conditional pdfs. Note that for low SNR (Fig. 3), the variance of the MAI plus thermal noise is very high, and therefore the set of soft outputs has a large definition support. This means that in order to achieve smooth pdf estimates, a large number of observations is required. This explains the oscillatory behavior we observe around +1 and −1 for low \( \text{SNR} = 0 \)dB. When the SNR is increased to \( \text{SNR} = 10 \)dB, no oscillations are observed since the MAI plus noise variance is reduced and the definition support of soft observations becomes very tight. In Fig. 5, we provide the BER estimation performance. We notice that the proposed unsupervised and non parametric BER estimator offers the same performance as the MC-aided method where the estimator has perfect knowledge about the transmitted information bits, i.e., perfect classification of soft outputs.
Note that when the proposed estimator knows the transmitted information bits, then its supervised version corresponds only to the computation of the smoothing parameters followed by the BER estimation according to (10) and the iterative Stochastic EM algorithm is not required.

2) Performance for Non Uniform Sources: We now turn to the case when the information bits are not equiprobable. We consider the scenario where \( \pi_+ = 0.75 \) and \( \pi_- = 0.25 \). The number of soft outputs is kept to \( N = 10^4 \). In Fig. 6, we report both the theoretical and estimated weighted conditional pdfs for \( \text{SNR} = 10 \text{dB} \). A quick inspection of the performance graph shows that even if the \textit{a priori} probabilities are not equal, the proposed technique provides reliable pdf estimates. Also, note that as in the previous scenario of equiprobable information bits, the oscillatory behavior is not observed. At the last iteration \( T = 6 \), the estimated \textit{a priori} values are \( \hat{\pi}_+ = 0.752 \) and \( \hat{\pi}_- = 0.248 \). We therefore conclude that the proposed unsupervised and non parametric BER estimator achieves reliable estimates of conditional pdfs for high SNR independently of the distribution of information bits.

3) Performance in the High SNR Region: We now focus on the behavior of the proposed BER estimator at the very high SNR region where it is difficult to achieve a reliable BER estimate when using MC-aided techniques with a limited number of soft observations. In Fig. 7, we report the BER estimation performance with \( \pi_+ = \pi_- = 1/2 \), and using only \( N = 10^3 \) soft observations. The proposed technique provides reliable BER estimates (with respect to the theoretical curve) for SNR values up to \( \text{SNR} = 16 \text{dB} \), while the MC technique fails to do so and stops at \( \text{SNR} = 8 \text{dB} \) because of the very limited number of transmitted information bits. This presents a major strength of the proposed BER estimator, and is very promising for many practical systems where it is required to estimate the BER of the communication link in a real-time fashion based only on a few data frames.

VI. Conclusions

In this paper, we considered the problem of unsupervised BER estimation for any communication system using any modulation/coding or signal processing technique. We proposed a BER estimation algorithm where only soft observations that serve for computing hard decisions about information bits are used to estimate the BER, and no prior knowledge about transmitted information bits is required. 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 (conditioned upon transmitted information bit values). We then proposed a BER computation technique using Gaussian Kernel-based pdf estimation. We provided a theoretical analysis of the estimator, and showed that it is
asymptotically unbiased, and point-wise consistent. Then, we introduced an iterative Stochastic EM technique to compute the parameters that serve for the estimation of conditional pdfs based only of soft observations. The proposed method involves the EM steps to estimate the a priori probabilities of transmitted information bits, and a Stochastic classification step to classify soft observations according to information bits. Finally, we evaluated the performance of the proposed unsupervised BER estimation technique in the framework of CDMA systems to corroborate the theoretical analysis. Interestingly, we showed that when classical MC methods fail to perform BER estimation in the region of high SNR, the proposed estimator provides reliable estimates using only few soft observations.

APPENDIX I
PROOF OF EQUATION (8)

Let us compute the exact value of $J(f_{X}^{b+})$ when the conditional pdf $f_{X}^{b+}$ is assumed to be Gaussian with mean $m_{+}$ and variance $\sigma_{+}^2$. We have,

$$f_{X}^{b+}(x) = \frac{1}{\sqrt{2\pi}\sigma_{+}} \exp\left(-\frac{(x-m_{+})^2}{2\sigma_{+}^2}\right). \tag{I.1}$$

Then, the second derivative of the conditional Gaussian distribution $f_{X}^{b+}$ can be written as,

$$f_{X}^{b+''}(x) = -\frac{1}{\sqrt{2\pi}\sigma_{+}^3} \exp\left(-\frac{(x-m_{+})^2}{2\sigma_{+}^2}\right) \left[1 - \frac{(x-m_{+})^2}{2\sigma_{+}^2}\right]. \tag{I.2}$$

Using the following change of variable, $t = \sqrt{2}(x-m_{+})/\sigma_{+}$, we can write the squared second derivative integral of the conditional Gaussian pdf as,

$$J(f_{X}^{b+}) = \frac{1}{2\pi\sigma_{+}^6} \int_{-\infty}^{+\infty} \left( f_{X}^{b+''}(x) \right)^2 dx = \frac{1}{2\pi\sigma_{+}^6} \int_{-\infty}^{+\infty} \exp\left(-\frac{t^2}{2}\right) \left[1 - \frac{t^2}{2}\right]^2 \left\frac{\sigma_{+}}{\sqrt{2}} \right] dt. \tag{I.3}$$

Let $m_k$ denote the $k$th moment of a zero-mean and unit-variance Gaussian distribution. Then with respect to (I.3), we get,

$$J(f_{X}^{b+}) = \frac{1}{2\sqrt{\pi}\sigma_{+}^4} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \left[1 - t^2 + \frac{t^4}{4}\right] dt = \frac{1}{2\sqrt{\pi}\sigma_{+}^4} \left[m_0 - m_2 + \frac{m_4}{4}\right]. \tag{I.4}$$

As for a zero-mean and unit-variance Gaussian random variable, we have $m_0 = 1$, $m_2 = 1$ and $m_4 = 3m_2^2 = 3$, we finally get,
\[ J(f_X^{b_+}) = \frac{3}{8\sqrt{\pi} \sigma_+^5} \]  (I.5)

**APPENDIX II**

**PROOF OF EQUATION (10)**

By invoking the expression of the BEP (2) and the estimates \( \hat{f}_{X,N+}^{b_+} \) and \( \hat{f}_{X,N-}^{b_-} \) of the two conditional pdfs \( f_X^{b_+} \) and \( f_X^{b_-} \), respectively, we can express the BER estimate as,

\[
\hat{p}_{e,N} = \pi_+ \int_{-\infty}^{0} \hat{f}_{X,N+}^{b_+}(x) \, dx + \pi_- \int_{0}^{+\infty} \hat{f}_{X,N-}^{b_-}(x) \, dx. \tag{II.1}
\]

Given the fact that the two conditional pdfs are estimated using Gaussian Kernels according to (4) and (5), and using the following change of variable, \( t = \frac{x-X_i}{h_{N+}} \) (respectively, \( t = \frac{x-X_i}{h_{N-}} \)) for \( B_+ \) (respectively, \( B_- \)), we get,

\[
B_+ = \int_{-\infty}^{0} \frac{1}{N_+ h_{N+}} \sum_{X_i \in C_+} K \left( \frac{x-X_i}{h_{N+}} \right) \, dx,
\]

\[
= \frac{1}{N_+} \sum_{X_i \in C_+} \int_{-\infty}^{h_{N+}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt,
\]

\[
= \frac{1}{N_+} \sum_{X_i \in C_+} \int_{h_{N+}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt,
\]

\[
= \frac{1}{N_+} \sum_{X_i \in C_+} Q \left( \frac{X_i}{h_{N+}} \right), \tag{II.2}
\]

\[
B_- = \int_{0}^{+\infty} \frac{1}{N_- h_{N-}} \sum_{X_i \in C_-} K \left( \frac{x-X_i}{h_{N-}} \right) \, dx,
\]

\[
= \frac{1}{N_-} \sum_{X_i \in C_-} \int_{h_{N-}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt,
\]

\[
= \frac{1}{N_-} \sum_{X_i \in C_-} Q \left( -\frac{X_i}{h_{N-}} \right). \tag{II.3}
\]

The BER estimate (10) is then obtained by combining (II.1), (II.2), and (II.3).
APPENDIX III

PROOF OF THEOREM 1

Let us recall that the proposed soft BER estimator is given by

$$\hat{p}_{e,N} = \pi_B + \pi_{-B},$$

(III.1)

where,

$$B_+ = \int_{-\infty}^{0} \frac{1}{N_+ h_{N_+}} \sum_{X_i \in C_+} K \left( \frac{x - X_i}{h_{N_+}} \right) dx,$$

(III.2)

$$B_- = \int_{0}^{+\infty} \frac{1}{N_- h_{N_-}} \sum_{X_i \in C_-} K \left( \frac{x - X_i}{h_{N_-}} \right) dx.$$

(III.3)

Let us examine the mathematical expectation of $B_+$, i.e., $\mathbb{E}[B_+]$. With respect to (III.2) we get,

$$\mathbb{E}[B_+] = \int_{-\infty}^{0} \frac{1}{N_+ h_{N_+}} \sum_{X_i \in C_+} \mathbb{E} \left[ K \left( \frac{x - X_i}{h_{N_+}} \right) \right] dx,$$

(III.4)

$$= \int_{-\infty}^{0} \frac{1}{N_+ h_{N_+}} N_+ \mathbb{E} \left[ K \left( \frac{x - X_j}{h_{N_+}} \right) | X_j \in C_+ \right] dx,$$

(III.5)

$$= \int_{-\infty}^{0} \frac{1}{h_{N_+}} \left( \int_{-\infty}^{+\infty} K \left( \frac{x - u}{h_{N_+}} \right) f_{X}^b(u) du \right) dx,$$

(III.6)

where (III.5) is obtained by noting that $\mathbb{E} \left[ K \left( \frac{x - X_i}{h_{N_+}} \right) \right]$ is identical for all $X_i \in C_+$ and can be evaluated only for a particular $X_j \in C_+$. Using the following change of variable $t = \frac{x - u}{h_{N_+}}$, we have,

$$\mathbb{E}[B_+] = \int_{-\infty}^{0} \left( \int_{-\infty}^{+\infty} K(t) f_{X}^b(x - h_{N_+} t) dt \right) h_{N_+} dx,$$

(III.7)

As $f_{X}^b$ is assumed to be a second derivative pdf function, we can write a Taylor series expansion of $f_{X}^b$ as follows,

$$f_{X}^b(x - h_{N_+} t) = f_{X}^b(x) - h_{N_+} t \left( f_{X}^b \right)'(x) + \frac{h_{N_+}^2 t^2}{2} \left( f_{X}^b \right)''(x) + O \left( h_{N_+}^3 t^3 \right).$$

(III.8)

Then, combining (III.7) and (III.8) we get,

$$\mathbb{E}[B_+] = \int_{-\infty}^{0} \left( \int_{-\infty}^{+\infty} K(t) \left[ f_{X}^b(x) - h_{N_+} t \left( f_{X}^b \right)'(x) + \frac{h_{N_+}^2 t^2}{2} \left( f_{X}^b \right)''(x) \right] + O \left( h_{N_+}^3 t^3 \right) \right) dt \right) dx,$$

(III.9)
\[
= \int_{-\infty}^{0} \left( f^b_{X}(x) \int_{-\infty}^{+\infty} K(t) \, dt - \left( f^b_{X}(x) \right)' h_{N+} \int_{-\infty}^{+\infty} tK(t) \, dt \right. \\
\left. + \frac{h_{N+}^2}{2} \left( f^b_{X}(x) \right)'' \int_{-\infty}^{+\infty} t^2 K(t) \, dt \right) \, dx + O(h_{N+}^3). \tag{III.10}
\]

As \( K(.) \) is a zero mean and unit variance Gaussian Kernel, we get \( \int_{-\infty}^{+\infty} K(t) \, dt = 1 \), \( \int_{-\infty}^{+\infty} tK(t) \, dt = 0 \), and \( \int_{-\infty}^{+\infty} t^2 K(t) \, dt = 1 \). Exploiting these properties and the fact that \( \int_{-\infty}^{0} \left( f^b_{X}(x) \right)'' \, dx = \left( f^b_{X}(x) \right)'(0) \), (III.10) can be expressed as,

\[
\mathbb{E}[B_+] = \int_{-\infty}^{0} f^b_{X}(x) \, dx + \frac{h_{N+}^2}{2} \left( f^b_{X}(x) \right)'(0) + O(h_{N+}^3). \tag{III.11}
\]

As \( h_{N+} \to 0 \) when \( N \to +\infty \), we therefore get,

\[
\lim_{N \to +\infty} \mathbb{E}[B_+] = \int_{-\infty}^{0} f^b_{X}(x) \, dx. \tag{III.12}
\]

Similarly, We can show that

\[
\lim_{N \to +\infty} \mathbb{E}[B_-] = \int_{0}^{+\infty} f^b_{X}(x) \, dx. \tag{III.13}
\]

Combining (III.1), (III.12), and (III.13), and by invoking the BEP (2) we get,

\[
\lim_{N \to +\infty} \mathbb{E}[\hat{p}_{e,N}] = \pi_+ \int_{-\infty}^{0} f^b_{X}(x) \, dx + \pi_- \int_{0}^{+\infty} f^b_{X}(x) \, dx \\
= p_e. \tag{III.14}
\]

**APPENDIX IV**

**PROOF OF THEOREM 2**

First, recall that the proposed soft BER estimator is given by,

\[
\hat{p}_{e,N} = \pi_+ B_+ + \pi_- B_-, \tag{IV.1}
\]

where,

\[
B_+ = \frac{1}{N_+ h_{N+}} \sum_{X_i \in C_+} \int_{-\infty}^{0} K \left( \frac{x - X_i}{h_{N+}} \right) \, dx, \tag{IV.2}
\]

\[
B_- = \frac{1}{N_- h_{N-}} \sum_{X_i \in C_-} \int_{0}^{+\infty} K \left( \frac{x - X_i}{h_{N-}} \right) \, dx. \tag{IV.3}
\]
Let us introduce the following quantities,

\[ A_{i+} = \int_{-\infty}^{0} K \left( \frac{x - X_{i}}{h_{N_{+}}} \right) \, dx, \quad \text{where } X_{i} \in C_{+}, \quad (\text{IV}.4) \]

\[ A_{i-} = \int_{0}^{+\infty} K \left( \frac{x - X_{i}}{h_{N_{-}}} \right) \, dx, \quad \text{where } X_{i} \in C_{-}. \quad (\text{IV}.5) \]

As \( X_1, \cdots, X_N \) are independent, the variance of \( \hat{p}_{e,N} \) can be computed as

\[ \text{Var} [\hat{p}_{e,N}] = \pi_{+}^2 \text{Var}(B_{+}) + \pi_{-}^2 \text{Var}(B_{-}) \]

\[ = \frac{\pi_{+}^2}{N_{+} h_{N_{+}}^2} \text{Var}(A_{i+}) + \frac{\pi_{-}^2}{N_{-} h_{N_{-}}^2} \text{Var}(A_{i-}). \quad (\text{IV}.6) \]

From (IV.2) and (III.11), we get,

\[ E \left[ A_{i+} \right] = h_{N_{+}} E \left[ B_{+} \right] = h_{N_{+}} \int_{-\infty}^{0} f^{b_{+}}(x) dx + \frac{h_{N_{+}}^3}{2} \left( f^{b_{+}} \right)'(0) + h_{N_{+}} O(h_{N_{+}}^3). \quad (\text{IV}.7) \]

Now, to determine the analytical expression of (IV.6), we must calculate \( E \left[ A_{i+}^2 \right], E \left[ A_{i-}^2 \right] \) and \( E \left[ A_{i-} \right] \) can be similarly deduced. Using (IV.4), we get,

\[ E \left[ A_{i+}^2 \right] = E \left[ \int_{-\infty}^{0} \int_{-\infty}^{0} K \left( \frac{x - X_{i}}{h_{N_{+}}} \right) K \left( \frac{y - X_{i}}{h_{N_{+}}} \right) \, dx \, dy \mid X_{i} \in C_{+} \right]. \quad (\text{IV}.8) \]

We can easily show that for the chosen Gaussian Kernel, we have,

\[ K \left( \frac{x - X_{i}}{h_{N_{+}}} \right) K \left( \frac{y - X_{i}}{h_{N_{+}}} \right) = K \left( \frac{X_{i} - \frac{x+y}{2}}{\sqrt{2} h_{N_{+}}} \right) K \left( \frac{x - y}{\sqrt{2} h_{N_{+}}} \right). \quad (\text{IV}.9) \]

Using (IV.8), (IV.9), and the following change of variable, \((v, w) = (\frac{x+y}{2}, x - y)\), we have,

\[ E \left[ A_{i+}^2 \right] = E \left[ \int_{-\infty}^{0} \int_{-\infty}^{0} K \left( \frac{X_{i} - \frac{x+y}{2}}{h_{N_{+}}/\sqrt{2}} \right) K \left( \frac{x - y}{\sqrt{2} h_{N_{+}}} \right) \, dx \, dy \mid X_{i} \in C_{+} \right], \quad (\text{IV}.10) \]

\[ = E \left[ \int_{-\infty}^{+\infty} \int_{-\infty}^{0} K \left( \frac{X_{i} - v}{h_{N_{+}}/\sqrt{2}} \right) K \left( \frac{w}{\sqrt{2} h_{N_{+}}} \right) \, dv \, dw \mid X_{i} \in C_{+} \right], \quad (\text{IV}.11) \]

\[ = E \left[ \sqrt{2} h_{N_{+}} \int_{-\infty}^{0} K \left( \frac{X_{i} - v}{h_{N_{+}}/\sqrt{2}} \right) \, dv \mid X_{i} \in C_{+} \right], \quad (\text{IV}.12) \]
where (IV.12) is obtained by noting that \( K(\cdot) \) is a pdf, and therefore \( \int_{\mathbb{R}} K(w)dw = 1 \). It follows from (IV.12) that

\[
\mathbb{E} \left[ A_{i+}^2 \right] = \sqrt{2h_{N+}} \int_{u=-\infty}^{+\infty} \left( \int_{-\infty}^{0} K \left( \frac{u-x}{h_{N+}/\sqrt{2}} \right) dx \right) f_X^{b_+}(u) du. \tag{IV.13}
\]

Using the following change of variable, \( t = \frac{u-x}{h_{N+}/\sqrt{2}} \), we get,

\[
\mathbb{E} \left[ A_{i+}^2 \right] = h_{N+}^2 \int_{t \in \mathbb{R}} \int_{-\infty}^{0} K(t) f_X^{b_+} \left( x + \frac{th_{N+}}{\sqrt{2}} \right) dt dx. \tag{IV.14}
\]

As \( f_X^{b_+} \) is assumed to be a second derivative pdf, a Taylor series expansion of \( f_X^{b_+} \) can be written as,

\[
f_X^{b_+} \left( x + \frac{th_{N+}}{\sqrt{2}} \right) = f_X^{b_+}(x) + \frac{th_{N+}}{\sqrt{2}} \left( f_X^{b_+} \right)'(x) + \frac{t^2h_{N+}^2}{4} \left( f_X^{b_+} \right)''(x) + O \left( t^3h_{N+}^2 \right). \tag{IV.15}
\]

From (IV.14), and (IV.15), and exploiting the fact that \( K(\cdot) \) is a zero mean and unit variance Gaussian Kernel, we get

\[
\mathbb{E} \left[ A_{i+}^2 \right] = h_{N+}^2 \int_{t \in \mathbb{R}} \int_{-\infty}^{0} \left( K(t) f_X^{b_+}(x) + \frac{tK(t)h_{N+}}{\sqrt{2}} \left( f_X^{b_+} \right)'(x) + \frac{t^2K(t)h_{N+}^2}{4} \left( f_X^{b_+} \right)''(x) \right) dt dx,
\]

\[
= h_{N+}^2 \left( p_{e+} + \frac{h_{N+}^2}{4} \left( f_X^{b_+} \right)'(0) \right) + O \left( h_{N+}^2 \right). \tag{IV.16}
\]

Using (IV.7) and (IV.16), we obtain

\[
\frac{1}{N_+h_{N+}^2} Var \left[ A_{i+}^2 \right] = \frac{p_{e+}(1-p_{e+})}{N_+} + \frac{h_{N+}^2}{N_+} \left( f_X^{b_+} \right)'(0) \left( \frac{1}{4} - p_{e+} \right) - \frac{h_{N+}^2}{4N_+} \left( f_X^{b_+} \right)'(0) + \frac{1}{N_+} O \left( h_{N+}^5 \right). \tag{IV.17}
\]

As \( h_{N+} \to 0 \) and \( N_+ \approx \pi_+ N \) when \( N \to +\infty \), it follows that,

\[
\lim_{N \to +\infty} \frac{Var \left[ A_{i+}^2 \right]}{N_+h_{N+}^2} = 0. \tag{IV.18}
\]

Similarly, we can show that \( \lim_{N \to +\infty} \frac{Var \left[ A_{i-}^2 \right]}{N_-h_{N-}^2} = 0 \). By recalling the expression in (IV.6), we finally get, \( \lim_{N \to +\infty} Var \left[ \hat{p}_{e,N} \right] = 0 \).
APPENDIX V

PROOF OF EQUATION (14)

Let us consider the maximization of \( Q(\theta(t)) \) in (13) with respect to (w.r.t) \( \pi_+^{(t)} + \pi_-^{(t)} = 1 \). If we add a Lagrangian Multiplier, we get

\[
L(\theta^{(t)}) = \sum_{i=1}^{N} \left\{ \rho_i^{(t)} \ln \left( \pi_+^{(t)} f_{X,A_+^{(t)}}(X_i) \right) + \rho_i^{(t)} \ln \left( \pi_-^{(t)} f_{X,A_-^{(t)}}(X_i) \right) \right\} - \lambda \left( \pi_+^{(t)} + \pi_-^{(t)} - 1 \right). \tag{V.1}
\]

Setting the two derivatives \( \frac{\partial L(\theta^{(t)})}{\partial \pi_+^{(t)}} \) and \( \frac{\partial L(\theta^{(t)})}{\partial \pi_-^{(t)}} \) to zero, we find

\[
\begin{align*}
\pi_+^{(t)} &= \frac{1}{\lambda} \sum_{i=1}^{N} \rho_i^{(t)}, \\
\pi_-^{(t)} &= \frac{1}{\lambda} \sum_{i=1}^{N} \rho_i^{(t)}.
\end{align*} \tag{V.2}
\]

By invoking the fact that \( \pi_+^{(t)} + \pi_-^{(t)} = 1 \), it follows from (V.2) that \( \lambda = N \).

REFERENCES

[1] M.C. Jeruchim, “Techniques for estimating the bit error rate in the simulation of digital communication systems,” IEEE J. Selec. Areas, Comm., vol. 2, no. 1, pp. 153-170, Jan. 1984.

[2] K.S. Shanmugan and P. Balaban, “A modified Monte Carlo simulation technique for the evaluation of error rate in digital communication systems,” IEEE Trans. Commu., vol. COM-28, pp. 1916–1924, Nov. 1980.

[3] E.J. Gumbel, “Statistics of extremes,” New York, Columbia Univ. Press, 1958.

[4] A. Abedi, M.E. Thompson and A.K. Khandani, “Application of cumulant method in performance evaluation of turbo-like codes,” IEEE Intern. Conf. Commun., ICC’07, 24-28 Jun. 2007 pp. 986 – 989.

[5] S. Saoudi, M. Troudi and F. Ghorbel, “An iterative soft bit error rate estimation of any digital communication systems using a non parametric probability density function,” Eurasip Journal on wireless Communications and Networking, Hindawi Publishing Corporation, Volume 2009, Article ID512192, 9 pages, doi:10.1155/2009/512192.

[6] M-S. Alouini, A. J. Goldsmith, “Capacity of rayleigh fading channels under different adaptive transmission and diversity-combining techniques,” IEEE Trans. Veh. Technol., vol. 48, no. 4, Jul. 1999.

[7] S. Verdu, “Multiuser detection,” Cambridge University Press, Cambridge, UK, August 1998.

[8] M. Tüchter, R. Koetter, and A. Singer, “Turbo equalization: principles and new results,” IEEE Trans. Commun., vol. 50, no. 5, pp. 754–767, May 2002.

[9] S. L. Ariyavisitakul, “Turbo space–time processing to improve wireless channel capacity,” in Proc. IEEE Commun. Conf., ICC, Jun. 2000, vol. 3, pp. 1238–1242.

[10] A. M. Tonello, “MIMO MAP equalization and turbo decoding in interleaved space time coded systems,” IEEE Trans. Commun., vol. 51, no. 2, pp. 155–160, Feb. 2003.
[11] T. Ait-Idir, S. Saoudi and N. Naja, “Space-time turbo equalization with successive interference cancellation for frequency selective MIMO channels,” *IEEE Trans. Veh. Technol.*, vol. 5, no. 5, pp. 2766-2778, Sep. 2008.

[12] T. Ait-Idir and S. Saoudi, “Turbo packet combining strategies for the MIMO-ISI ARQ channel,” *IEEE Trans. Commun.*, (To Appear). available at http://www.inpt.ac.ma/~aitidir/Library/HARQ_TCOMArticle_May2009_final.pdf

[13] M. Rosenblatt, “Remarks on some non-parametric estimates of a density function,” *Ann. Math. Statist.*, vol. 27, Issue 3, pp. 832–837, 1956.

[14] E. Parzen, “On estimation of a probability density function and mode,” *Ann. Math. Statist.*, vol. 33, Issue 3, pp. 1065–1076, 1962.

[15] 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, Sep. 1968.

[16] S. Saoudi, F. Ghorbel and A. Hillion, “Nonparametric probability density function estimation on a bounded support: applications to shape classification and speech coding,” *Applied Stochastic Models and Data Analysis Journal*, John Wiley, vol. 10, no. 3, pp. 215–231, Sep. 1994.

[17] S. Saoudi, F. Ghorbel and A. Hillion, “Some statistical properties of the kernel-diffeomorphism estimator,” *Applied Stochastic Models and Data Analysis Journal*, John Wiley, Vol. 13, Issue 1, pp. 39-58, March 1997.

[18] A. P. Dempster, N. M. Laird, D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society*. Series B, vol. 39, No. 1 (1977), pp. 1–38

[19] P. Masson and W. Pieczynski, “SEM algorithm and unsupervised statistical segmentation of satellite images,” *IEEE Trans. Geoscience and Remote Sensing*, vol. 31, no. 3, May 1993, pp. 618–633.

[20] N. Bouguila and D. Ziou, “A hybrid SEM algorithm for high-dimensional unsupervised learning using a finite generalized Dirichlet mixture,” *IEEE Trans. Image Proc.*, vol. 15, no. 9, (2006), pp. 2657–2668.
Fig. 1. General transmission scheme for any transmitter and receiver with soft outputs $X_1, \cdots, X_N$ and hard decisions $\hat{b}_1, \cdots, \hat{b}_N$. 
Initial classification ($t=1$)
2 classes of $X_i$, $i=1, \ldots, N$

Compute the two a priori probabilities

Compute the two conditional pdfs

$t = t + 1$

Compute the $N$ APPs using a priori prob. and pdfs obtained at iter. $t-1$

Classify $X_1, \ldots, X_N$ with respect to the new APPs

Re-estimate the two a priori probabilities $\pi_+$ and $\pi_-$

Re-estimate the two conditional pdfs

Yes

$t < T$

No

Compute BER

Fig. 2. The diagram of the proposed iterative BER estimation algorithm.
TABLE I

SUMMARY OF THE STOCHASTIC EM-BASED UNSUPERVISED BER ESTIMATION ALGORITHM

1. Initialization ($t = 1$)
   1.1. Classify soft outputs $X_1, \cdots, X_N$ according to their signs, i.e., $C_+^{(1)} = \{ X_i : X_i \geq 0 \}$, and $C_-^{(1)} = \{ X_i : X_i < 0 \}$.
   1.2. Deduce $N_+^{(1)}$ and $N_-^{(1)}$, i.e., $N_+^{(1)} = \left| C_+^{(1)} \right|$ and $N_-^{(1)} = \left| C_-^{(1)} \right|$.
   1.3. Compute standard-deviations $\sigma_+^{(1)}$ and $\sigma_-^{(1)}$ corresponding to $C_+^{(1)}$ and $C_-^{(1)}$, respectively.
   1.4. Compute the smoothing parameters $h_{N_+}^{(1)}$ and $h_{N_-}^{(1)}$ according to (9).
   1.5. Deduce a priori probabilities as $\pi_+^{(1)} = \frac{N_+^{(1)}}{N}$ and $\pi_-^{(1)} = \frac{N_-^{(1)}}{N}$.
   1.6. Compute the two conditional pdfs using $\theta^{(1)}$, (4), and (5).

2. Parameter Estimation at Iteration $t$
   2.1. Estimation Step
      2.1.1. Estimate APPs $\rho_{i+}^{(t)}$ and $\rho_{i-}^{(t)}$ using (12).
   2.2. Maximization Step
      2.2.1. Compute a priori probabilities $\pi_+^{(t)}$ and $\pi_-^{(t)}$ using (14).
   2.3. Classification Step
      2.3.1. Classify subsets $C_+^{(t)}$ and $C_-^{(t)}$ using APP values $\rho_{i+}^{(t)}$ and $\rho_{i-}^{(t)}$ according to the random Bayesian rule (15).
      2.3.2. $N_+^{(t)} = \left| C_+^{(t)} \right|$ and $N_-^{(t)} = \left| C_-^{(t)} \right|$.
      2.3.3. Compute the standard-deviations $\sigma_+^{(t)}$ and $\sigma_-^{(t)}$ corresponding to $C_+^{(t)}$ and $C_-^{(t)}$, resp.
      2.3.4. Compute the smoothing parameters $h_{N_+}^{(t)}$ and $h_{N_-}^{(t)}$ according to (9).
      2.3.5. Compute the two conditional pdfs using $\theta^{(t)}$, (4), and (5).

3. Parameter Estimation
   Compute the BER estimate using $\theta^{(T)}$ and (10).
Fig. 3. Estimated conditional pdfs for $\pi_+ = \pi_- = 1/2$, and SNR = 0dB.
Fig. 4. Estimated conditional pdfs for $\pi_+ = \pi_- = 1/2$, and SNR = 10dB.
Fig. 5. BER performance comparison when \( \pi_+ = \pi_- = 1/2 \), and \( N = 10^4 \) soft observations.
Fig. 6. Estimated conditional pdfs for $\pi_+ = 0.75$, $\pi_- = 0.25$, and SNR = 10dB.
Fig. 7. Behavior of the BER estimation at high SNR with $\pi_+ = \pi_- = 1/2$, and only $N = 10^3$ soft observations.