V. Concluding Remarks

A better diversity gain can be achieved by transmitting copies of the same data symbol through multiple antennas in the closed-loop TAD. To match channel conditions, each antenna can have different weights (i.e., phases and powers). In this correspondence, it is shown that the feedback delay can potentially degrade the performance of closed-loop TAD under a fast fading channel environment. To see the impact of the feedback delay on the performance, an analytical BER expression has been derived and verified by simulation results.

If the channel variation is too fast (e.g., the case of high-speed vehicles), the channel mismatching error is significant and the diversity gain is handful, even though multiple transmit antennas are used. In this case, some alternative approaches (e.g., open-loop TAD) shall be utilized to get good diversity gain.

REFERENCES


Semiblind Maximum-Likelihood Demodulation for CDMA Systems

Joaquín Míguez and Luis Castedo

Abstract—This correspondence addresses the problem of channel estimation and symbol detection in wireless direct-sequence code-division multiple-access (DS-CDMA) communication systems. We introduce a novel multiluser demodulation scheme that proceeds in two steps. First, the multiluser channel parameters are estimated according to a suitable modification of the maximum-likelihood (ML) criterion using the expectation maximization (EM) algorithm. Subsequently, this estimate and other useful side information are employed to perform ML detection of the transmitted symbols with the Viterbi algorithm. Our main contribution is the development of a novel stochastic ML method for channel estimation that takes advantage of all the available statistical information referred to the transmitted signals and channel noise. Additionally, it can incorporate the knowledge of a fraction of the transmitted symbols; hence, the term semiblind. Computer simulation results are presented that show how close-to-optimum performance is achieved in time-dispersive fading channels using remarkably short training sequences.

Index Terms—Channel estimation, code-division multiple access (CDMA), expectation maximization (EM) algorithm, maximum likelihood, multiuser detection, semiblind receivers.

I. INTRODUCTION

Direct-sequence code-division multiple-access (DS-CDMA) schemes used in third-generation mobile communication systems [1] degrade their performance when multiple-access interference (MAI) arising from the simultaneous use of the time and frequency channel resources by several users and intersymbol interference (ISI) due to the time-dispersive nature of typical multipath wireless channels occur. In this context, optimum multiluser demodulation1 is of interest in order to obtain high spectral efficiencies. The optimum multiluser detector performs joint maximum likelihood (ML) estimation of the symbols transmitted by the different users and is implemented with the Viterbi algorithm [2]. The complexity of this algorithm grows exponentially with the number of users and suboptimum procedures based on the expectation maximization (EM) algorithm [3] have been alternatively proposed [4], [5]. Nevertheless, all these detectors rely on a very precise knowledge of the channel responses for all the system users.

When the transmitted data are assumed known at the receiver, the ML criterion for channel estimation reduces to the well-known least squares (LS) method [6]. The EM algorithm has also been exploited to reduce the complexity of this estimator [6]–[9]. A more general approach is to perform joint data detection and channel estimation and several deterministic ML techniques have been proposed to solve this joint problem (see [6], [7], [10], [11], and references therein). Deterministic ML (DML) methods address the maximization of the received signal probability density function (pdf) with respect to (w.r.t.) the channel coefficients and the transmitted symbols, which are all considered as deterministic parameters to be estimated. In practice,
DML approaches lead to decision directed schemes that alternate between channel estimation and data detection; and error propagation phenomena easily appear, leading to performance degradation.

Recently, the problem of optimum demodulation has also been tackled from a Bayesian point of view. In [12], the maximum a posteriori (MAP) estimation of both the channel parameters and the transmitted data is addressed using the space alternating generalized EM (SAGE) algorithm. Unlike in DML schemes, both the channel parameters and the data are dealt with as random variables, the realizations of which must be estimated from the available observations. Hence, an a priori probability model must be chosen for each variable of interest. This is not, in general, an easy task and a wrong choice of the prior probability distribution may affect the performance of the estimation method.

In this correspondence, we introduce a novel method for optimum multiuser demodulation that consists of two stages. First, we take a stochastic ML (SML) approach for the estimation of the multiuser channel parameters. Subsequently, the symbols transmitted by all users are optimally detected (conditioned on the channel estimate) using the Viterbi algorithm. Our main contribution is the development of a new semiblind SML technique for channel estimation in multiuser systems. The main feature of the SML framework is that the unknown data are modeled as random variables that can be integrated out in the channel likelihood function, thus reducing the number of parameters to be estimated and increasing statistical efficiency. Another important characteristic of the proposed estimator is that it easily allows to incorporate the knowledge of known symbols (typically available in currently standardized communication systems that transmit training sequences and/or pilot symbols for diverse purposes); hence, the term semiblind. These known symbols are included in the likelihood function in a systematic way, unlike most semiblind techniques proposed in the literature that are based on ad hoc procedures [13], [14]. Despite the existence of known data, the stochastic likelihood is a nonquadratic function, the maxima of which cannot be found analytically. Hence, we have resorted to the EM methodology in order to practically implement the SML-based estimator.

The remainder of this correspondence is organized as follows. Section II describes the signal model. Section III introduces the channel estimation criterion. In Section IV, we describe the iterative EM algorithm that numerically approximates the estimator. ML multiuser detection using the channel estimate and side information provided by the EM algorithm is briefly described in Section V. Computer simulations are presented in Section VI. Section VII is devoted to the conclusions.

II. SIGNAL MODEL

As shown in [15], a DS-CDMA system with time-dispersive channels can be modeled by the equivalent discrete-time system depicted in Fig. 1. According to this model, when the $i$th user transmits an isolated symbol $s_i$, it is multiplied by a unique binary-valued spreading sequence with $L$ chips per symbol $c_i(k)$, $k = 0, \ldots, L-1$. The resulting signal passes through a linear time-dispersive discrete channel $h_i(k) = 0, \ldots, P-1$ and the received sequence is the superposition of the transmitted signals from the $N$ users plus an additive white Gaussian noise (AWGN) sequence $g(k)$, i.e.,

$$x(k) = \sum_{i=1}^{N} s_i d_i(k) + g(k), \quad k = 0, \ldots, L + P - 2. \quad (1)$$

The sequence $d_i(k) = c_i(k) \ast h_i(k) = \sum_{p=0}^{L-1} h_i(p) c_i(k-p)$, has length $L+P-1$ and will be termed received code. The discrete channel impulse response $h_i(k)$ accounts not only for the actual continuous channel response but also for the relative time delays of the different users (when the system is asynchronous) and the transmit and receive filters.

Due to the channel time-dispersive effect, when a stream of symbols is transmitted, the $i$th user $n$th datum $s_i(n)$ interferes with $s_i(n-1)$, $s_i(n-2), \ldots, s_i(n-m + 1)$, where $m = \lceil (L + P - 1)/L \rceil$ is the channel memory size in symbols ($\lceil \cdot \rceil$ denotes the upper integer part of a real number). Assuming that the received signal is sampled at the chip period, the receiver collects $L$ observations per symbol period and, therefore, the overall received sequence in the $n$th period is given by

$$x(n, k) = x(nL + k) = \sum_{i=1}^{N} \sum_{r=0}^{m-1} d_i(r, k) s_i(n-r) + g(n, k) \quad k = 0, \ldots, L - 1 \quad (2)$$

where $d_i(r, k) = d_i(rL + k)$ and $g(n, k) \equiv g(nL + k)$.

Using vector notation, the observations in (2) can be written as

$$x(n) = \sum_{r=0}^{m-1} D(r) s(n-r) + g(n) = D s(n) + g(n) \quad (3)$$

where $x(n) = [x(n, 0), \ldots, x(n, L - 1)]^T$, $D = [D(r-m+1) \cdots D(0)]$ is the $L \times Nm$ received code matrix, composed of the $L \times N$ submatrices $D(r) = [d_1(r) \cdots d_N(r)]$ and $D(r) = d_i(rL, 0), \ldots, d_i(rL, L-1)]^T$. The transmitted symbol vectors $s(n) = [s_1(n), \ldots, s_N(n)]^T$, $l = n - m + 1, \ldots, n$, are stacked to construct the $N \times m$ received symbol vector $g(n) = [g^T(n-m+1) \cdots g^T(n)]^T$ and $g(n) = [g(n, 0), \ldots, g(n, L-1)]^T$ is a vector of independent and identically distributed (i.i.d.) complex Gaussian random variables with zero mean and variance $\sigma^2$.

In the rest of this correspondence, we will assume that both the number of active users $N$ and the length of the corresponding dispersive channels $P$ are known at the receiver. The first assumption is realistic because the number of active users is always known at the base station for traffic-control purposes. On the other hand, $P$ should be understood as an upper bound for the $N$ different channel lengths [i.e., $h_i(k)$ may have only $P \times K$ nonzero coefficients]. Finally, the AWGN power $\sigma^2$ is also assumed to be known (or estimated) at the receiver. Computer simulations show, however, that large differences between the estimated and the actual noise power do not lead to a noticeable performance loss.

III. MAXIMUM-LIKELIHOOD MULTIUSER CHANNEL ESTIMATION

Let us build the $P \times 1$ th user channel vector $h_i = [h_i(0), \ldots, h_i(P-1)]^T$ and the $P \times N$ multiuser channel matrix $H = [h_1 \ldots h_N]$. The first stage of the proposed ML-based demodulator is to estimate $H$ from a set of $K$ available observation vectors $\mathbf{x}(0), \ldots, \mathbf{x}(K-1)$.
A. Stochastic Maximum Expected-Likelihood Channel Estimation

We propose to estimate the channel matrix $\mathbf{H}$ as

$$\hat{\mathbf{H}} = \arg \max_{\mathbf{H}} \left\{ \mathcal{L}(\mathbf{H}) = \log \left( \frac{\mathcal{F}(\mathbf{H})}{\mathcal{F}(\mathbf{H}(\mathbf{x}(n)))} \right) \right\}$$

(4)

where $\mathcal{F}(\mathbf{H})$ is the pdf of a single observation vector $\mathbf{x}(n)$, which depends on $\mathbf{H}$; $\log(\cdot)$ denotes the natural logarithm; and $\mathcal{F}(\mathbf{H}(\mathbf{x}(n)))$ denotes statistical expectation with respect to (w.r.t.) $\mathbf{x}(n)$. Assuming that the additive white Gaussian noise (AWGN) is statistically independent of the received symbols, the observation probability density function (pdf) turns out to be

$$\mathcal{F}(\mathbf{H}(\mathbf{x}(n))) = \frac{1}{\sigma^2} \mathcal{N}(\mathbf{H}) \exp \left( -\frac{||\mathbf{x}(n) - \mathbf{D}\mathbf{s}(n)||^2}{\sigma^2} \right)$$

(5)

where $\mathcal{N}(\mathbf{H})$ denotes expectation w.r.t. the $\mathcal{N} \times 1$ received symbol vector $\mathbf{s}(n)$, and $||\mathbf{v}||^2_2 = \mathbf{v}^H \mathbf{v}$. Note that the random symbols are marginalized out, so that $\mathcal{F}(\mathbf{H})$ does not depend on $\mathbf{s}(n)$ and, hence, (4) is qualified as an SML method. Substituting (5) into (4), we arrive at the equivalent formulation

$$\hat{\mathbf{H}} = \arg \max_{\mathbf{H}} \left\{ \mathcal{L}(\mathbf{H}) = \mathcal{L}(\mathbf{H}) \log \left( \frac{\mathcal{F}(\mathbf{H})}{\mathcal{F}(\mathbf{H}(\mathbf{x}(n)))} \right) \right\}$$

(6)

where the cost function $\mathcal{L}(\mathbf{H})$ will be termed the expected log-likelihood of the multiuser channel and, hence, $\hat{\mathbf{H}}$ will be referred to as the stochastic maximum expected-likelihood (SME) estimate of $\mathbf{H}$.

The expectation in (6) can be analytically calculated because in digital communications, the symbols are usually modeled as discrete random variables with known finite-alphabet pdf. Indeed, for an arbitrary function $\phi(\cdot)$ of the data

$$\mathcal{F}(\mathbf{H}) = \frac{1}{\sigma^2} \mathcal{N}(\mathbf{H}) \exp \left( -\frac{||\mathbf{x}(n) - \mathbf{D}\mathbf{s}(n)||^2}{\sigma^2} \right)$$

(7)

where $\mathbf{s}_j(n)$ is the $j$th possible realization of $\mathbf{s}(n)$. If $\log_2(V)$ is the number of bits per symbol in the keying format, there are $V^{Mn}$ different terms in (7) and, as a consequence, the complexity of computing $\mathcal{F}(\mathbf{H}(\mathbf{x}(n)))$ is $O(V^{Mn})$, where $m$ is the channel memory size in symbols.

The statistical expectation $\mathcal{E}(\mathbf{x}(n))$, however, cannot be directly computed and must be estimated from the actually received observation vectors. If we consider that each user transmits a block of $K$ symbols, beginning with a training sequence of length $0 \leq M < K$, $\mathcal{S} = [\mathbf{s}(0), \ldots, \mathbf{s}(M-1)]$, the sample mean provides an unbiased estimate of the statistical expectation under the hypotheses of stationarity and ergodicity of the noise and the transmitted data. Hence, the SME channel estimator can be practically obtained as

$$\hat{\mathbf{H}} = \arg \max_{\mathbf{H}} \left\{ \mathcal{L}(\mathbf{H}) = \sum_{j=0}^{M-1} \log \left( \mathcal{F}(\mathbf{H}(\mathbf{x}(n))) \right) \right\}$$

(8)

where $\mathcal{L}(\mathbf{H})$ is an estimate of the expected log-likelihood function.\footnote{When the channels are not dispersive (i.e., when $\mathbf{H}$ is a $1 \times N$ vector) and the modulation format is memoryless, the observation vectors $\mathbf{x}(n)$ are i.i.d. and $\mathcal{L}(\mathbf{H})$ is also the true log-likelihood of the multiuser channel matrix w.r.t. the whole set of observations.}

\begin{table}
\centering
\caption{Iterative EM Channel Estimation Algorithm}
\begin{tabular}{|l|}
\hline
System parameters: \\
$N$: number of system users \\
$L$: spreading sequence length \\
$F$: length of the discrete-time channel impulse responses \\
im: channel memory size in symbol periods \\
$M$: number of training symbols per user \\
$K$: number of observation vectors to be processed \\
\hline
Algorithm initialization: \\
(a) If training data are available ($M > 0$), compute \\
$\mathbf{A}(n)$ and $\mathbf{b}(n)$, for $n = 0, \ldots, M - 1$, as shown in the appendix. \\
Solve the linear system \\
$$\left( \sum_{n=0}^{M-1} \mathbf{A}(n) \right) \mathbf{b}(n) = \sum_{n=0}^{M-1} \mathbf{b}(n)$$
\hline
(b) Else, if no training data are available ($M = 0$): \\
Let $\mathbf{H}_0$ be the channel estimate from the previous block. \\
Iterative step. For $i > 0$, compute \\
(a) For $n = M, M + m - 1$, let \\
$$\mathbf{E}_A(n) = \left( \mathbf{E}_{A}\left(\mathbf{E}_{A}\left(\mathbf{E}_{A}\left(\mathbf{E}_{A}(\mathbf{b}(n) - \mathbf{b}(n))\right)\right)\right)\right)$$
\hline
\end{tabular}
\end{table}

for $M = 0$, totally blind estimation of the multiuser channel matrix is performed.

It is important to remark that conventional ML channel estimation methods are fully supervised, i.e., they rely exclusively on the training sequence to estimate the channel parameters. In such cases, the ML criterion reduces to the classical LS estimation: only the first term in $\mathcal{L}(\mathbf{H})$ is considered, whereas all the statistical knowledge of the unknown symbols [given by the second term in $\mathcal{L}(\mathbf{H})$] is neglected. For this reason, conventional channel estimation algorithms are suboptimum.

We would also like to note the difference between the semiblind estimator (8) and the DML approaches in [6], [7], [10], and [11]. As already mentioned, DML approaches lead to decision-driven schemes where symbol detection and channel estimation are iteratively alternated. Such algorithms are very sensitive to the selection of the initial channel/data estimates and suffer from error propagation phenomena. Equation (8) results from the marginalization of the transmitted symbols in the likelihood of the channel in order to avoid this drawback. As shown in Section IV, the application of the EM approach in this case yields a more robust algorithm that is not subject to error propagation.

IV. EM ALGORITHM

Since it is not possible to find a closed-form solution to (8), we propose to estimate $\mathbf{H}$ using the EM algorithm [3]. Using the standard EM notation, let $\mathbf{x}(n), n = 0, \ldots, K - 1$ be the incomplete (observed)
data and let the extended observations \( \mathbf{x}_n(n) = [\mathbf{s}^T(n) \mathbf{x}^T(n)]^T \), \( n = 0, \ldots, K - 1 \) be the complete data. A derivation of the EM algorithm similar to the one presented in [3] leads to the iterative rule

\[
\mathbf{H}_{i+1} = \arg \max_{\mathbf{H}} \ell_X(\mathbf{H}, \mathbf{H}_i) = \arg \max_{\mathbf{H}} \sum_{n=0}^{K-1} \mathbf{F}_n(n, \mathbf{x}_n(n)) \log(f_{X|H}(\mathbf{x}_n(n))).
\]

where the function \( \ell_X(\cdot, \cdot) \) provides the sufficient statistics for the complete data and can be written as \( \ell_X(\mathbf{H}, \mathbf{H}_i) = \sum_{n=0}^{K-1} \mathbf{F}_n(n, \mathbf{x}_n(n)) \log(f_{X|H}(\mathbf{x}_n(n))) \). The complete data pdf \( f_{X|H}(\mathbf{x}_n(n)) \), in turn, can be easily derived to yield

\[
f_{X|H}(\mathbf{x}_n(n)) = f_{\mathbf{x}_{n}(n)}(\mathbf{x}_n(n)) f_{\mathbf{s}(n)}(\mathbf{s}(n)) = \frac{1}{\pi \sigma_{\mathbf{s}}^2} \exp \left(-\frac{\|\mathbf{x}_n(n) - \mathbf{D}_s(n)\|^2}{\sigma_{\mathbf{s}}^2}\right) f_{\mathbf{s}(n)}(\mathbf{s}(n)).
\]

Substituting (10) into (9) and taking into account the availability of the known symbols in \( \mathbf{S} = [\mathbf{s}(0), \ldots, \mathbf{s}(M-1)] \), we arrive at the iterative EM algorithm for SLM channel estimation

\[
\mathbf{H}_{i+1} = \arg \min_{\mathbf{H}} \left\{ \sum_{n=0}^{K-1} \mathbf{F}_n(n, \mathbf{x}_n(n), \mathbf{H}) \| \mathbf{x}_n(n) - \mathbf{D}_s(n) \|^2 \right\}.
\]

Thus, we have cast (8), which does not have a closed-form solution, into a sequence of quadratic problems that can be analytically solved. In the Appendix, it is demonstrated that the solution to (11) is given by

\[
\text{vec} \left( \mathbf{H}_{i+1} \right) = \left( \sum_{n=0}^{K-1} \mathbf{F}_n(n, \mathbf{x}_n(n), \mathbf{H}) \right)^{-1} \left( \sum_{n=0}^{K-1} \mathbf{F}_n(n, \mathbf{x}_n(n), \mathbf{H}) \mathbf{b}(n) \right)
\]

where vec(\( \mathbf{H} \)) is the \( N \times 1 \) vector built by stacking together the \( N \) column vectors in \( \mathbf{H} \), i.e., vec(\( \mathbf{H} \)) = [\( \mathbf{h}_1^T \ldots \mathbf{h}_N^T \]^T \( \cdot \)). \( \mathbf{A}(n) \) and \( \mathbf{b}(n) \) are an \( N \times N \) matrix and an \( N \times 1 \) vector, respectively, that are obtained from the observations \( \mathbf{x}_n(n) \), the received symbols \( \mathbf{s}(n) \), and the transmitted codes.

The statistical expectations are dropped in (12) for \( n < M \) due to the availability of the training sequence, i.e., \( \mathbf{F}_n(n, \mathbf{x}_n(n), \mathbf{H}) = \mathbf{A}(n) \) and \( \mathbf{F}_n(n, \mathbf{x}_n(n), \mathbf{H}) \mathbf{b}(n) = \mathbf{b}(n) \) when \( n < M \). For \( n \geq M + m - 1 \), the calculation of the expected conditions in (12) requires knowledge of the conditioned a posteriori pdf \( f_{\mathbf{x}_n|\mathbf{H}_i}(\cdot) \) that can be expressed in terms of \( f_{\mathbf{x}_n|\mathbf{H}_i}(\cdot), f_{\mathbf{x}_n|\mathbf{H}_i}(\cdot), \) and \( f_{\mathbf{x}_n}(\cdot) \) using the Bayes theorem. Notice that \( f_{\mathbf{x}_n|\mathbf{H}_i}(\cdot) \) is easily obtained from (5). As a result, the following expression is obtained:

\[
E_f(\mathbf{x}_n|\mathbf{H}_i; \mathbf{H}_i, \Phi(n)) = \frac{E_f(\mathbf{x}_n, \mathbf{s}(n)) \exp \left(-\frac{\|\mathbf{x}_n(n) - \mathbf{D}_s(n, \mathbf{s}(n))\|^2}{\sigma_{\mathbf{s}}^2}\right) \Phi(n)}{E_f(\mathbf{x}_n, \mathbf{s}(n)) \exp \left(-\frac{\|\mathbf{x}_n(n) - \mathbf{D}_s(n, \mathbf{s}(n))\|^2}{\sigma_{\mathbf{s}}^2}\right) \Phi(n)}
\]

where \( \mathbf{D}_s(n, \mathbf{s}(n)) \) is the received code matrix built from \( \mathbf{H}_i \) and \( \Phi(n) \) should be substituted by \( \mathbf{A}(n) \) or \( \mathbf{b}(n) \) as needed to compute (12). Finally, for \( M \leq n < M + m - 1 \), (13) can still be used substituting \( E_f(\mathbf{x}_n(n)) \) by \( E_f(\mathbf{x}_n(n)) \).

Table I outlines the channel estimation procedure. We assume a system receiving bursts of \( IK \) symbols per user, where \( I \) is a positive integer. The first block of \( K \) symbols contains a training sequence of length \( M \), and the semiblind EM algorithm (12) can be applied. For the subsequent \( K \) symbols, the blind EM algorithm, given by (12) with \( M = 0 \), is used. Adequate initialization is provided by the channel estimate obtained from the previous block.

A final word is in order regarding the convergence of the EM algorithm. Under very general regularity conditions, EM algorithms converge monotonically to a stationary point of the likelihood function [3]. Using the results in the Appendix, all the stationary points of \( \ell_X(\mathbf{H}) \) (including the global maximum) can be characterized as vec(\( \mathbf{H} \)) = \( \left( \sum_{n=0}^{K-1} \mathbf{F}_n(n, \mathbf{x}_n(n), \mathbf{H}) \right)^{-1} \left( \sum_{n=0}^{K-1} \mathbf{F}_n(n, \mathbf{x}_n(n), \mathbf{H}) \mathbf{b}(n) \right) \). This expression is analogous to the conventional LS estimator except for the use of the posterior estimates, \( \mathbf{F}_n(n, \mathbf{x}_n(n), \mathbf{H}) \), instead of the true symbols. EM theory also proves that if the algorithm gets trapped at some stationary point not being a local maximum (i.e., a saddle point), a small random perturbation of \( \mathbf{H}_i \) will cause the sequence of iterations to diverge from that point [3].

V. OPTIMUM MULTIUSER DETECTION

When i.i.d. symbols are transmitted, optimum detection results from the joint ML estimation of \( \mathbf{g}(M), \ldots, \mathbf{g}(K-1) \) [2]. Given the multiuser channel estimate \( \mathbf{H} \), ML estimates of the transmitted symbols are obtained as

\[
\hat{\mathbf{g}}_{\mathbf{H}}, = \arg \max_{\mathbf{g}} \left\{ \log(f_{\mathbf{x}_n}(\mathbf{x}_n|\mathbf{g}, \mathbf{H})) \right\}
\]

\[
= \arg \min_{\mathbf{g}} \left\{ \sum_{n=M}^{K-1} \|\mathbf{x}_n(n) - \mathbf{D}_s(n)\|^2 \right\}
\]

where \( f_{\mathbf{x}_n}(\mathbf{x}_n|\mathbf{g}, \mathbf{H}) = f_{\mathbf{x}_n|\mathbf{H}}(\mathbf{x}_n|\mathbf{g}, \mathbf{H}) \) is the joint likelihood of the transmitted symbols. Equation (14) can be efficiently solved using the Viterbi algorithm with branch metrics \( l_n = \|\mathbf{x}_n(n) - \mathbf{D}_s(n)\|^2 \geq 0, n = M, \ldots, K - 1 \). Let us remark that these metrics have already been computed in the last iteration of the EM channel estimation algorithm. Therefore, symbol detection is sped up because it only involves additions and comparisons.

VI. COMPUTER SIMULATIONS

We let consider the uplink of a DS-CDMA communication system with \( N = 3 \) users, binary phase-shift keying symbols, and randomly chosen binary spreading codes with length \( L = 5 \). As described at the end of Section IV, symbols are transmitted in bursts of length \( IK \), where \( K = 26 \) is the block size and \( I \) is a positive integer, and the first block includes \( M = 5 \) known symbols per user. The chip period is \( T_c = 1 \mu s \), so the chip rate is \( R_c = T_c^{-1} = 1 \mathrm{Mc/s} \) and the user bit rate is \( R_b = (L/2)^{-1} = 200 \text{Kbps} \). The iterative EM algorithm in Table I is used for channel estimation, and the Viterbi algorithm as described in Section IV is employed for data detection. We focus our attention in the channel estimation error and the bit error rate (BER) for the initial block containing the training data, since the performance in subsequent blocks depends on the accuracy of the channel estimate provided by the semiblind EM algorithm.

In simulating the channel, we assume a Gaussian wide-sense stationary uncorrelated scattering (US) fading model. Hence, the channel impulse responses are time-varying and can be denoted, at time \( n \), as \( h_i^n = [h_i^n(0), h_i^n(1), \ldots, h_i^n(P-1)]^T \), \( i = 1, 2, \ldots, N \), where the fading coefficients \( h_i^n(p) \) are complex Gaussian random variables with zero mean and variance \( E_{h_i^n}(h_i^n(p)^2) = \sigma_{h_i}^2, \forall n \). According to the US assumption, these coefficients are modeled as statistically independent random variables and, as a consequence, \( E_{h_i^n}(h_i^n(p)^2) = \sigma_{h_i}^2, \forall n \). The Doppler power spectrum is assumed to be the classical Rayleigh, which yields the Clarke’s channel
The autocovariance function \( \phi_{ij}(k) = \sigma_J^2 e^{j2\pi F_d k T_c} \) for the type-C channel in an outdoor-to-indoor pedestrian environment.

Fig. 2. Delay power profiles for the type-C channel in an outdoor-to-indoor pedestrian environment.

Computer simulations have been carried out for the type-C multipath channel defined for the outdoor-to-indoor and pedestrian environments in IMT-2000, according to [17]. It is a hostile frequency-selective channel with a high root mean square (rms) delay spread \((\tau_{\text{max}} = L \tau_s)\). The delay power profile is assumed to be exponentially decreasing [17], [18], normalized to \( \sum_{k=1}^{L} \sigma_J^2 e^{j2\pi F_d k T_c} = 1 \) \( \forall i \), as shown in Fig. 2. If we set a threshold at \( 10 \log_{10}(\sigma_J^2) = -20 \) dB, the maximum length of the discrete-time fading channels is fixed to \( P = 6 \), corresponding to a maximum delay \( \tau_{\text{max}} < 6 \) \( \mu s \). The transmitter motion speed is set to \( v_{\text{m}} = 15 \) Km/h, which yields a maximum Doppler spread \( F_d \approx 28 \) Hz and, therefore, a slow fading model.

We have generated 30 independent realizations of the channel random process and evaluated the performance of the following.

1) The fully supervised LS estimator, termed LS\((K)\). This is also the ML estimator when all the transmitted symbols are known and, therefore, reduces to the proposed EM algorithm in that case. Hence, it provides an adequate performance limit for comparison purposes.

2) The conventional LS estimator, termed LS\((M)\). In this case, only the training sequence \( (M \text{ symbols from each user}) \) is used for channel estimation.

3) The iterative implementation of the generalized ML sequence detection and estimation (GMLSDE) scheme proposed in [11]. The GMLSDE receiver, termed GMLSDE\((M, K)\), is obtained from a DML approach and consists of a decision-directed method that iteratively alternates LS channel estimation and ML symbol detection (using the Viterbi algorithm). The iterative algorithm is initialized with the LS\((M)\) estimate.

4) The proposed semiblind EM algorithm, termed EM\((M, K)\), which is also initialized with the LS\((M)\) estimate.

Performance is measured in terms of the BER and the normalized mean-squared error (NMSE) of the channel estimates. The normalized squared error (NSE) for a single channel realization \( \mathbf{H}(0), \ldots, \mathbf{H}(K-1) \) (using straightforward notation for the time-varying channel process) is defined as

\[
\text{NSE} = \frac{\sum_{n=0}^{K-1} \text{Trace} \left( \left( \mathbf{H}(n) - \mathbf{\hat{H}} \right)^H \left( \mathbf{H}(n) - \mathbf{\hat{H}} \right) \right)}{\sum_{n=0}^{K-1} \text{Trace} \left( \mathbf{H}(n)^H \mathbf{H}(n) \right)}
\]

whereas the NMSE is the average over the set of 30 independent trials. Both BER and NMSE are represented versus the average signal-to-noise ratio (SNR), defined for a particular channel realization as

\[
\text{SNR} = 10 \log_{10} \left( \frac{\sigma_s^2 \sum_{n=0}^{K-1} \text{Trace} \left( \mathbf{D}(n)^H \mathbf{D}(n) \right)}{N K \sigma_B^2} \right) \text{ (dB)}
\]

where \( \sigma_s^2 = E_s(n)s(n) \) \( \forall n \) is the transmitted symbol power.

Since the EM algorithm requires an approximate knowledge of the noise variance \( \sigma_B^2 \), we have assumed that an estimate of this variance \( \hat{\sigma}_B^2 \) is available. The estimation error is \( 70\% \), i.e., \( \sigma_B^2 = 0.3 \sigma_B^2 \). Finally, in order to simulate a near-far scenario [2], the amplitudes of the random channel coefficients have been adequately scaled to yield the power ratios \( 10 \log_{10}(\gamma_i^2/\gamma_0^2) = 5 \) dB and \( 10 \log_{10}(\gamma_i^2/\gamma_0^2) = -5 \) dB, where \( \gamma_i^2 = \sum_{n=0}^{K-1} d_i(n)^H d_i(n) \) is the total received power of user \( i \). \( d_i(n) = [d_i(n)^T, \ldots, d_i(Ln)^T]^T \) and the sequence \( d_i^0 = h^0_i(k) \) is the time-varying received code for user \( i \).

Fig. 3 depicts the NMSE versus the average SNR for the LS\((5)\), LS\((26)\), GMLSDE\((5, 26)\), and EM\((5, 26)\) algorithms. The EM\((5, 26)\) channel estimation algorithm is clearly better than the DML-based GMLSDE\((5, 26)\) scheme and performs close to the LS\((26)\) method. The conventional LS\((5)\) estimator presents a very poor performance.

These results are corroborated by the BER curves depicted in Fig. 4, which were obtained with the two-channel realizations that yielded the least (left plot) and the largest (right plot) channel estimation squared error at SNR = 15 dB. The semiblind EM-based receiver performance is practically optimum, whereas the GMLSDE\((5, 26)\) receiver is very sensitive to error propagation and may present a very poor BER when the initial channel estimate, given by the LS\((5)\) estimator, is not accurate enough. Hence, the degradation in the BER.

Fig. 5 shows the convergence speed of the EM\((5, 26)\) algorithm at SNR = 12 dB. The performance limit is attained quickly, within less than five to six iterations.

VII. CONCLUSION

We have presented a new approach to multuser demodulation in DS-CDMA based on a modification of the ML criterion termed stochastic maximum expected likelihood. The method proceeds in two stages: SME estimation of the multuser channel and ML detection of the transmitted symbols. Estimation of the channel coefficients is iteratively carried out with the EM algorithm, which provides very fast convergence. The ML detector is efficiently implemented using the Viterbi algorithm and taking advantage of some useful side information produced by the EM algorithm, which alleviates the computational load.
Fig. 4. Two sample BER curves for several values of the SNR. Channel realization with the (a) least channel estimation error and (b) largest channel estimation error at SNR = 12 dB.

Fig. 5. Average convergence speed of the EM(5, 26) algorithm.

The main contribution of this correspondence is the SME channel estimation technique. Although channel estimation can be carried out in a blind way, the EM algorithm is known to present local misconvergence problems when it is not adequately initialized. This limitation is circumvented by a semiblind scheme that incorporates training sequences when they are available. Since the role of training data is to ensure the convergence of the EM algorithm, and not to improve the estimation accuracy, they can be remarkably short. Computer simulations that support these results are presented.

APPENDIX

MAXIMIZATION OF THE COMPLETE DATA SUFFICIENT STATISTICS

The solution to (11) is the unique minimum of the quadratic cost function

\[ Q(\mathbf{H}) = \sum_{n=0}^{K-1} E_{w(n)|x(n)} \sum_{i,j} \hat{h}_{ij}^2 |x(n) - \mathbf{D} s(n)|^2. \]

To find it, we need to solve the linear system \( \nabla_{\mathbf{H}} Q(\mathbf{H}) = 0_{p \times N} \), where \( \nabla_{\mathbf{H}} Q(\mathbf{H}) \) is the gradient of the cost function w.r.t. the multiuser channel matrix. Since

\[ \nabla_{\mathbf{H}} Q(\mathbf{H}) = \sum_{n=0}^{K-1} E_{w(n)|x(n)} \sum_{i,j} \nabla_{\mathbf{H}} |x(n) - \mathbf{D} s(n)|^2 \]

the problem of computing the gradient reduces to the derivation of

\[ \nabla_{\mathbf{H}} |x(n) - \mathbf{D} s(n)|^2 \]. It is straightforward to show that

\[
\nabla_{\mathbf{H}} |x(n) - \mathbf{D} s(n)|^2 = -\mathbf{G}_H \left( \sum_{i=0}^{L-1} z(n, I) \sum_{j=1}^{N} d_j^*(iL + l) s_j^*(n - i) \right)
\]

where \( z(n) = [z(n, 0), \ldots, z(n, L - 1)]^T = [x(n) - \mathbf{D} s(n) \end{align} \) and \( \mathbf{G}_H \) is a \( P \times N \) matrix whose element in the \( p \)-th row and \( r \)-th column is

\[
g_{pr} = \sum_{l=0}^{L-1} z(n, l) \sum_{i=0}^{m-1} \sum_{i=0}^{N} \hat{h}_{r, i}(p) d_j^*(iL + l - i) s_j^*(n - i)
\]

where \( w_{n,i}^*(r, l, p) = \sum_{i=0}^{P-1} c_i^*(iL + l - i) s_i(n - i) \) and \( * \) denotes complex conjugation. The previous expression reveals the relationship between the gradient \( \nabla_{\mathbf{H}} Q(\mathbf{H}) \) and the multiuser channel coefficients.

Substituting (15) and (16) into \( \nabla_{\mathbf{H}} Q(\mathbf{H}) = 0_{p \times N} \) leads, after some manipulation, to the system of linear equations

\[
\left( \sum_{n=0}^{K-1} E_{w(n)|x(n)} \mathbf{H} : \mathbf{A}(n) \right) \text{vec}(\mathbf{H}) = \sum_{n=0}^{K-1} E_{w(n)|x(n)} \mathbf{H} : \mathbf{b}(n)
\]

where \( \text{vec}(\mathbf{H}) \) is the \( NP \times 1 \) vector obtained by stacking the \( N \) columns of matrix \( \mathbf{H} \), \( \mathbf{A}(n) \) is an \( NP \times NP \) matrix whose element in the \((i+1)\)-th row and \((j+1)\)-th column is given by \( [\mathbf{A}(n)]_{i+1, j+1} = a_n([i/P], [j/P], j \in P, i \in P) \), and \( \mathbf{b}(n) \) is an \( NP \times 1 \) vector whose \((i+1)\)-th element is...
quantities can be decomposed as

\[ a(i, p, j, q) = \sum_{l=0}^{j-1} \sum_{r=0}^{m-1} c(l \cdot r + l - q) s_j(n - r) n^P_{\gamma}(i, l, p) \]

\[ b_n(i, p) = \sum_{j=1}^{N} x(n, l) n^P_{\gamma}(i, l, p) i_s \]

Finally, the solution to (18) is

\[ \text{vec}(H_n) = \left( \sum_{n=0}^{K-1} F_{\mu(n)}[x(n); H, A(n)] \right)^{-1} \cdot \left( \sum_{n=0}^{K-1} F_{\mu(n)}[x(n); H, b(n)] \right). \]