

Maximum Likelihood Turbo Iterative Channel Estimation for Space-Time Coded Systems and Its Application to Radio Transmission in Subway Tunnels

Miguel González-López

*Departamento de Electrónica y Sistemas, Universidade da Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain
Email: miguelgl@udc.es*

Joaquín Míguez

*Departamento de Electrónica y Sistemas, Universidade da Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain
Email: jmiguez@udc.es*

Luis Castedo

*Departamento de Electrónica y Sistemas, Universidade da Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain
Email: luis@udc.es*

Received 31 December 2002; Revised 31 July 2003

This paper presents a novel channel estimation technique for space-time coded (STC) systems. It is based on applying the maximum likelihood (ML) principle not only over a known pilot sequence but also over the unknown symbols in a data frame. The resulting channel estimator gathers both the deterministic information corresponding to the pilot sequence and the statistical information, in terms of a posteriori probabilities, about the unknown symbols. The method is suitable for Turbo equalization schemes where those probabilities are computed with more and more precision at each iteration. Since the ML channel estimation problem does not have a closed-form solution, we employ the expectation-maximization (EM) algorithm in order to iteratively compute the ML estimate. The proposed channel estimator is first derived for a general time-dispersive MIMO channel and then is particularized to a realistic scenario consisting of a transmission system based on the global system mobile (GSM) standard performing in a subway tunnel. In this latter case, the channel is nondispersive but there exists controlled ISI introduced by the Gaussian minimum shift keying (GMSK) modulation format used in GSM. We demonstrate, using experimentally measured channels, that the training sequence length can be reduced from 26 bits as in the GSM standard to only 5 bits, thus achieving a 14% improvement in system throughput.

Keywords and phrases: STC, turbo equalization, turbo channel estimation, maximum likelihood channel estimation, GSM, subway tunnels.

1. INTRODUCTION

Recently, the so-called Turbo codes [1, 2, 3] have revealed themselves as a very powerful coding technique able to approach the Shannon limit in AWGN channels. A Turbo code is made up of two component codes (block or convolutional) parallelly or serially concatenated via an interleaver. This simple coding scheme produces very long codewords, so each source information bit is highly spread through the transmitted coded sequence. At reception, optimum maximum likelihood (ML) decoding can be carried out by considering the hypertext associated with the concatenation of the two

component codes. Obviously, such a decoding approach becomes impractical in most situations. The key idea behind Turbo coding is to overcome this problem by employing a suboptimal, but very powerful, decoding scheme termed *iterative maximum a posteriori* (MAP) decoding [3, 4]. Basically, the method relies on independently decoding each of the component codes and exchanging in an iterative fashion the statistical information, that is, the a posteriori probabilities about symbols, obtained in each decoding module.

The same decoding principle has also been successfully applied, under the term *Turbo equalization* [5], to effectively compensate the ISI induced by the channel and/or the

modulation scheme. This technique exploits the fact that ISI can be viewed as a form of rate-1, nonrecursive coding. So, whatever coding scheme is used, if an interleaver is located prior to the channel, the overall effect of coding and ISI can be treated as a concatenated code and therefore, iterative MAP decoding can be applied. Luschi et al. [6] present an in-depth review of this technique and further improvements can be found in [7, 8, 9, 10]. In general, iterative MAP processing can be applied to a variety of situations where the overall system can be viewed as a concatenation of modules whose input/output relationship can be described as a (hidden) Markov chain. Several works have appeared in the last years exploiting this idea. For instance, Görtz [11], Garcia-Frias and Villasenor [12], and Guyader et al. [13] worked on the problem of joint source-channel decoding and Zhang and Burr [14] addressed the problem of symbol timing recovery.

In practical receivers, where the channel impulse response has to be estimated, it is convenient to have channel estimators capable of benefiting from the high performance of Turbo equalizers [15, 16, 17]. Moreover, second- and third-generation mobile standards consider the transmission of pilot sequences known by the receiver for channel estimation purposes. In the *global system mobile* (GSM) standard, this sequence is 26 bits long, which represents 17.6% of the total frame length (148 bits) [18]. Such a long training sequence is necessary if classical estimation techniques, such as least squares (LS), are used. Employing more refined channel estimators, such as the one presented in this paper, we can dramatically decrease the necessary length of the training sequence and therefore increase the overall system throughput. In [19], an ML-based channel estimator is presented where the ML principle is applied not only to the pilot sequence, but also to the whole data frame. Since the involved optimization problem had no analytical solution, the expectation-maximization (EM) algorithm [20] was used for iteratively obtaining the solution.

Also, wireless communications research has been very influenced by the discovery of the potentials of communicating through multiple-input multiple-output (MIMO) channels, which can be carried out using antenna diversity not only at reception, as classical space-diversity techniques have been doing, but also at transmission. MIMO techniques have the advantage to provide high data rate wireless services at no extra bandwidth expansion or power consumption. Telatar [21] calculated the capacity associated with a MIMO channel that in certain cases grows linearly with the number of antennas [22]. More recent progress in information theoretical properties of multiantenna channel can be found in [23].

Although MIMO channel capacity can be really high, it can only be successfully exploited by proper coding and modulation schemes. The term space-time Coding (STC) [24, 25] has been adopted for such techniques. Special efforts have been made in code design [24, 26] and several decoding approaches have been developed for these codes. In both fields, the Turbo principle has been applied in profusion. Turbo ST codes designs can be found in [27, 28, 29] and various Turbo decoding schemes are exposed in [30, 31].

As in single-antenna systems, practical ST receivers must perform the operation of channel estimation. Having efficient and robust estimators is crucial to guarantee that the system performance degradation due to the channel estimation error is minimized. In this paper, we present a novel channel estimation technique that gathers both the deterministic information corresponding to the pilot sequence and the statistical information, in terms of a posteriori probabilities, about the unknown symbols. The method is suitable for Turbo equalization schemes where those probabilities are computed with more and more precision at each iteration. We derive the channel estimator for general MIMO time-dispersive channels and analyze its performance in a multiple-antenna communication system based on the GSM standard operating inside subway tunnels.

The main motivation for developing a multiple-antenna GSM-based communication system is the following. GSM is, by far, the most widely deployed radio-communication system. Since 1993, its radio interface (GSM-R) has been adopted by the European railway digital radio-communication systems. Due to the conservative nature of its market, it is expected that railway radio-communication systems will employ GSM-R for the long-term future. For this reason, when subway operators wish to deploy advanced, high data rate, digital services for security or entertainment purposes, it is very likely that they will prefer to increase the capacity of the existing GSM-R system rather than switch to another radio standard. STC and Turbo equalization are very promising ways of achieving this capacity growth [32]. In this specific application, we will show that the proposed iterative MLMIMO channel estimation method has large benefits over traditional channel estimation approaches.

The rest of the paper is organized as follows. Section 2 presents the signal model and Section 3 describes the Turbo equalization scheme for STC systems. Next, in Section 4, we derive the ML channel estimator for a general time-dispersive MIMO channel. Since direct application of the ML principle leads to an optimization problem without closed-form solution, the EM algorithm is applied for computing the actual value of the solution, resulting in the so-called ML-EM estimator. The application of the proposed channel estimator to a STC GSM-based system operating in subway tunnels is detailed in Section 5. Section 6 presents the results of computer experiments for both the general case and experimental measurements of subway tunnel MIMO channels. Finally, Section 7 is devoted to the conclusions.

2. SIGNAL MODEL

We consider the transmitter signal model corresponding to an STC system shown in Figure 1. The original bit sequence $u(k)$ feeds an ST encoder whose output is a sequence of vectors $\mathbf{c}(k) = [c_1(k) \ c_2(k) \ \cdots \ c_N(k)]^T$, with N being the number of transmitting antennas. The specific spatio-temporal structure of the sequence of vectors $\mathbf{c}(k)$ depends on the particular STC technique employed. Any of the several STC methods that have been proposed in the literature could be used in our scheme. However, we have focused on ST

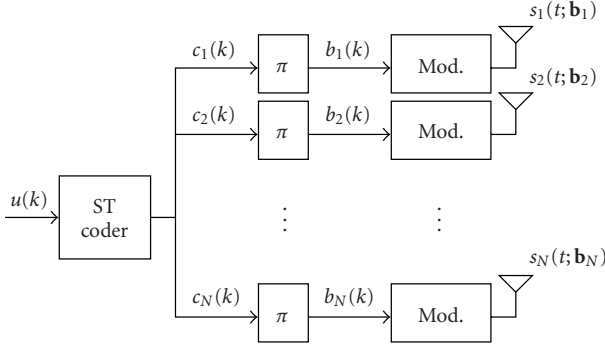


FIGURE 1: Transmitter model.

trellis codes [24, 25] to elaborate our simulation results. Each component of $\mathbf{c}(k)$ is independently interleaved to produce a new symbol vector $\mathbf{b}(k) = [b_1(k) \ b_2(k) \ \dots \ b_N(k)]^T$ and these are the symbols that are afterwards modulated (waveform encoded) to yield the signals $s_i(t; \mathbf{b}_i)$ $i = 1, 2, \dots, N$ that will be transmitted along the radio channel. Without loss of generality, we will assume that the modulation format is linear and that the channel suffers from time-dispersive multipath fading with memory length m . It is well known that at reception, matched-filtering and symbol-rate sampling can be used to obtain a set of sufficient statistics for the detection of the transmitted symbols. Using vector notation, this set of statistics will be grouped in vectors $\mathbf{x}(k) = [x_1(k) \ x_2(k) \ \dots \ x_L(k)]^T$, $k = 0, 1, \dots, K - 1$, where L is the number of receiving antennas and K is the number of total transmitted symbol vectors in a data frame. Elaborating the signal model, it can be easily shown that the sufficient statistics $\mathbf{x}(k)$ can be expressed as

$$\mathbf{x}(k) = \mathbf{H}\mathbf{z}(k) + \mathbf{v}(k), \quad (1)$$

where matrix $\mathbf{H} = [H(m-1) \ H(m-2) \ \dots \ H(0)]$ represents the overall dispersive MIMO channel with memory length m . Each submatrix

$$H(i) = \begin{bmatrix} h_{11}(i) & h_{12}(i) & \dots & h_{1N}(i) \\ h_{21}(i) & h_{22}(i) & \dots & h_{2N}(i) \\ \vdots & \vdots & \ddots & \vdots \\ h_{L1}(i) & h_{L2}(i) & \dots & h_{LN}(i) \end{bmatrix} \quad (2)$$

contains the fading coefficients that affect the symbol vector $\mathbf{b}(k-i)$. Vector $\mathbf{z}(k)$ results from stacking the source vectors $\mathbf{b}(k)$, that is,

$$\mathbf{z}(k) = [\mathbf{b}^T(k-m+1) \ \mathbf{b}^T(k-m+2) \ \dots \ \mathbf{b}^T(k)^T]. \quad (3)$$

Finally, the noise component $\mathbf{v}(k)$ is a vector of mutually independent complex-valued, circularly symmetric Gaussian random processes, that is, the real and imaginary parts are zero-mean, mutually independent Gaussian random processes having the same variance. We will also assume that the noise is temporally white with variance σ_v^2 .

3. ST TURBO DETECTION

Figure 2 shows the block diagram of an ST Turbo detector. The MAP equalizer [4] computes $L[\mathbf{b}(k)|\bar{\mathbf{x}}]$ which are the *a posteriori* log-probabilities of the input symbols $\mathbf{b}(k)$ based on the available observations $\bar{\mathbf{x}} = [\mathbf{x}^T(0) \ \mathbf{x}^T(1) \ \dots \ \mathbf{x}^T(K-1)]^T$. Due to its time-dispersive nature, it is convenient to represent our MIMO channel by means of a finite-state machine (FSM) having $2^{N(m-1)}$ states. This FSM has 2^N transitions per state which implies that there is a total number of 2^{Nm} transitions between two time instants. Let $e_k = (s_{k-1}, \mathbf{b}(k), \mathbf{s}(k), s_k)$ be one of the 2^{Nm} possible transitions at time k of this FSM. This transition depends on four parameters: the incoming state s_{k-1} , the outgoing state s_k , the input symbol vector $\mathbf{b}(k)$, and the output symbol vector without noise $\mathbf{s}(k) = \mathbf{H}\mathbf{z}(k)$. It is important to point out that the incoming state is determined by the $m-1$ previous symbol vectors, that is, $s_{k-1} = (\mathbf{b}(k-m+1), \mathbf{b}(k-m+2), \dots, \mathbf{b}(k-1))$. On the other hand, the outgoing state is a function of the previous state and the current input symbols, that is, $s_k = f_{\text{next}}(s_{k-1}, \mathbf{b}(k))$. For a better description of the MAP equalizer, we are going to introduce the notation $\mathbf{b}(k) = L_{\text{in}}(e_k)$ and $\mathbf{s}(k) = L_{\text{out}}(e_k)$ to represent the input and output symbol vectors associated to the transition e_k , respectively. Note that the output vector does not depend on the outgoing state s_k , so we will slightly change our notation and write

$$\begin{aligned} \mathbf{s}(k) &= L_{\text{out}}(e_k) = L_{\text{out}}(s_{k-1}, \mathbf{b}(k)) \\ &= L_{\text{out}}(\mathbf{z}(k)) = \mathbf{H}\mathbf{z}(k). \end{aligned} \quad (4)$$

The *a posteriori* log-probabilities $L[\mathbf{b}(k)|\bar{\mathbf{x}}]$ can be recursively computed by means of the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm [3, 4] which is summarized in the sequel. The first stage when computing the *a posteriori* log-probabilities is noting that

$$L[\mathbf{b}(k)|\bar{\mathbf{x}}] = L[\mathbf{b}(k), \bar{\mathbf{x}}] + h_b, \quad (5)$$

where h_b is the constant that makes $P[\mathbf{b}(k)|\bar{\mathbf{x}}]$ a probability mass function and

$$L[\mathbf{b}(k), \bar{\mathbf{x}}] = \log \sum_{e_k: L_{\text{in}}(e_k) = \mathbf{b}(k)} \exp L[e_k, \bar{\mathbf{x}}] \quad (6)$$

is the joint log-probability of the transition e_k and the set of available observations $\bar{\mathbf{x}}$. This joint log-probability can be expressed as

$$L[e_k, \bar{\mathbf{x}}] = \alpha_{k-1}[s_{k-1}] + \gamma_k[e_k] + \beta_k[s_k], \quad (7)$$

where

$$\begin{aligned} \alpha_k[s] &= L[s_{k-1}, \bar{\mathbf{x}}_k^-], \\ \gamma_k[e_k] &= L[\mathbf{b}(k)] + L[\mathbf{x}(k)|\mathbf{s}(k)], \\ \beta_k[s] &= L[\bar{\mathbf{x}}_k^+ | s_k], \end{aligned} \quad (8)$$

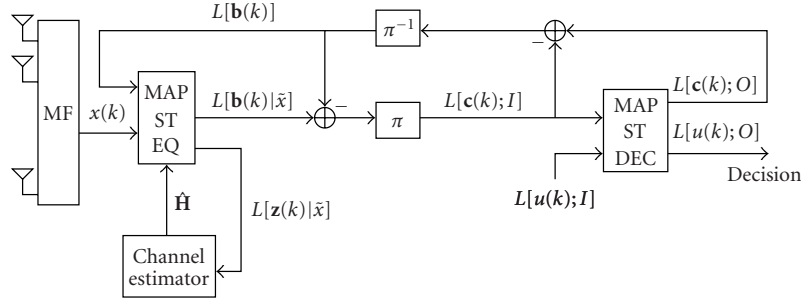


FIGURE 2: Receiver model.

with

$$L[\mathbf{x}(k)|\mathbf{s}(k)] = -\frac{1}{\sigma_v^2} \|\mathbf{x}(k) - \mathbf{H}\mathbf{z}(k)\|^2, \quad (9)$$

$$\tilde{\mathbf{x}}_k^- = [\mathbf{x}^T(0) \ \mathbf{x}^T(1) \ \cdots \ \mathbf{x}^T(k-1)], \quad (10)$$

$$\tilde{\mathbf{x}}_k^+ = [\mathbf{x}^T(k+1) \ \mathbf{x}^T(k+2) \ \cdots \ \mathbf{x}^T(K-1)]. \quad (11)$$

Note that the noise variance σ_v^2 is needed in (9). Our simulation results assume this parameter as known. However, it could be estimated and, in particular, it can be considered as another parameter to be estimated by the ML estimator described in Section 4, as shown in [33], for the case of a decision feedback-equalizer (DFE) instead of a MAP detector. The computation of the quantities $\alpha_k[s]$, $\gamma_k[e_k]$, and $\beta_k[s]$ can be carried out recursively by first performing a forward recursion

$$\begin{aligned} & \alpha_{k-1}[s_{k-1}] \\ &= \log \sum_{\substack{\mathbf{b}(k), s_{k-2}: \\ f_{\text{next}}(s_{k-2}, \mathbf{b}(k-1)) = s_{k-1}}} \exp \left\{ \alpha_{k-2}[s_{k-2}] + L[\mathbf{b}(k-1)] \right. \\ & \quad \left. + L[\mathbf{x}(k)|\mathbf{s}(k)] \right\} \end{aligned} \quad (12)$$

with initial values $\alpha_0[s=0] = 0$ and $\alpha_0[s \neq 0] = -\infty$, and then proceeding with a backward recursion

$$\begin{aligned} \beta_k[s_k] &= \log \sum_{\substack{\mathbf{b}(k+1), s_{k+1}: \\ f_{\text{next}}(s_k, \mathbf{b}(k+1)) = s_{k+1}}} \exp \left\{ \beta_{k+1}[s_{k+1}] + L[\mathbf{b}(k+1)] \right. \\ & \quad \left. + L[\mathbf{x}(k+1)|\mathbf{s}(k+1)] \right\} \end{aligned} \quad (13)$$

using as initial values $\beta_{K-1}[s = s_{K-1}] = 0$ and $\beta_{K-1}[s \neq s_{K-1}] = -\infty$.

Similarly, the decoder has to compute the *a posteriori* log-probabilities of the original symbols $L[u(k); O]$ from their *a priori* log-probabilities $L[u(k); I] = \log(0.5)$ and the *a priori* log-probabilities $L[\mathbf{c}(k); I]$ which come from the detector. Again, the BCJR algorithm applies [3, 4]. It also computes the *a posteriori* log-probabilities of the transmitted symbols

$L[\mathbf{c}(k); O]$ using

$$\begin{aligned} & L[\mathbf{c}(k); O] \\ &= \log \sum_{e_k: L_{\text{out}}(e_k) = \mathbf{c}(k)} \exp \{ \alpha_{k-1}[s_{k-1}] + \gamma_k[s_k] + \beta_k[s_k] \}, \end{aligned} \quad (14)$$

where $L[\mathbf{c}(k); I]$ is utilized as branch metric. These computed log-probabilities are then fed back to the detector to act as the *a priori* log-probabilities $L[\mathbf{b}(k)]$. As reflected in Figure 2, note that it is always necessary to subtract the *a priori* component from the computed log-probabilities before forwarding them to the other module in order to avoid statistical dependence with the results of the previous iteration.

4. MAXIMUM LIKELIHOOD CHANNEL ESTIMATION

Channel estimation is often mandatory when practically implementing ST detection strategies, unless we deal with some kind of blind processing techniques. In this section, we will present a novel channel estimation method that will enable us to take full advantage from the Turbo detection scheme presented in the Section 3.

When developing our channel estimation approach, we will exploit the fact that transmitted data frames in most practical systems contain a deterministic known pilot sequence of length M for the purpose of estimating the channel at reception. For instance, in GSM, this sequence is $M = 26$ bits long [18]. Let $\tilde{\mathbf{b}}_f = [\tilde{\mathbf{b}}_t^T \ \tilde{\mathbf{b}}^T]^T$ denote the overall data frame, which includes $\tilde{\mathbf{b}}_t = [\mathbf{b}_t^T(0) \ \mathbf{b}_t^T(1) \ \cdots \ \mathbf{b}_t^T(M-1)]^T$ as the training sequence and $\tilde{\mathbf{b}} = [\mathbf{b}^T(M) \ \mathbf{b}^T(M+1) \ \cdots \ \mathbf{b}^T(K-1)]^T$ as the information sequence. Analogously, $\tilde{\mathbf{x}}_f = [\tilde{\mathbf{x}}_t^T \ \tilde{\mathbf{x}}^T]^T$ are the observations corresponding to one data frame, where $\tilde{\mathbf{x}}_t = [\mathbf{x}_t^T(0) \ \mathbf{x}_t^T(1) \ \cdots \ \mathbf{x}_t^T(M-1)]^T$ represents the pilot sequence and $\tilde{\mathbf{x}} = [\mathbf{x}(M) \ \mathbf{x}(M+1) \ \cdots \ \mathbf{x}(K-1)]^T$ corresponds to the information sequence. The ML estimator is thus given by

$$\hat{\mathbf{H}} = \arg \max_{\mathbf{H}} f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}}; \mathbf{H}}(\tilde{\mathbf{x}}), \quad (15)$$

where $f_{\tilde{\mathbf{x}}|\tilde{\mathbf{b}}; \mathbf{H}}$ is the probability density function (pdf) of the observations conditioned on the available information (the training sequence $\tilde{\mathbf{b}}_t$) and the parameters to be estimated

(the channel matrix \mathbf{H}). Although, this is a problem without closed-form solution, the EM algorithm [20] can be employed to iteratively solve (15). The EM algorithm relies on defining a so-called “complete data” set formed by the observable variables and by additional unobservable variables. At each iteration of the algorithm, a more refined estimate is computed by averaging the log-likelihood of the complete data set with respect to the pdf of the unobservable variables conditioned on the available set of observations. Using the EM terminology, we define the union of the observations (which are the observable variables) and the transmitted bit sequence (which are the unobservable variables) $\tilde{\mathbf{x}}_e = [\tilde{\mathbf{b}}_f^T \tilde{\mathbf{x}}_f^T]^T$ as the complete data set, whereas the observations $\tilde{\mathbf{x}}_f$ are the incomplete data set. The relationship between $\tilde{\mathbf{x}}_e$ and $\tilde{\mathbf{x}}_f$ must be given by a noninvertible linear transformation, that is, $\tilde{\mathbf{x}}_f = \mathbf{T}\tilde{\mathbf{x}}_e$. It can be easily seen that in our case, this transformation is given by $\mathbf{T} = [\mathbf{0}_{L(M+K) \times N(M+K)} \mathbf{I}_{L(M+K)}]$. With these definitions in mind, the estimate of the channel at the $i + 1$ th iteration is obtained by solving

$$\hat{\mathbf{H}}_{i+1} = \arg \max_{\mathbf{H}} E_{\tilde{\mathbf{x}}_e | \tilde{\mathbf{x}}_f, \tilde{\mathbf{b}}_f; \hat{\mathbf{H}}_i} \{ \log f_{\tilde{\mathbf{x}}_e | \tilde{\mathbf{b}}_f; \mathbf{H}}(\tilde{\mathbf{x}}_e) \}, \quad (16)$$

where $E_f\{\cdot\}$ denotes the expectation operator with respect to the pdf $f(x)$. Expanding the previous expression, we have

$$\begin{aligned} \hat{\mathbf{H}}_{i+1} &= \arg \max_{\mathbf{H}} E_{\tilde{\mathbf{b}}_f | \tilde{\mathbf{x}}_f; \hat{\mathbf{H}}_i} \{ \log [f_{\tilde{\mathbf{x}}_f | \tilde{\mathbf{b}}_f; \mathbf{H}}(\tilde{\mathbf{x}}_f) f_{\tilde{\mathbf{b}}_f}(\tilde{\mathbf{b}}_f)] \} \\ &= \arg \max_{\mathbf{H}} E_{\tilde{\mathbf{b}}_f | \tilde{\mathbf{x}}_f; \hat{\mathbf{H}}_i} \{ \log [f_{\tilde{\mathbf{x}}_f | \tilde{\mathbf{b}}_f; \mathbf{H}}(\tilde{\mathbf{x}}_f) f_{\tilde{\mathbf{x}}_f | \tilde{\mathbf{b}}_f; \mathbf{H}}(\tilde{\mathbf{x}}_f)] \} \\ &= \arg \max_{\mathbf{H}} \log f_{\tilde{\mathbf{x}}_f | \tilde{\mathbf{b}}_f; \mathbf{H}}(\tilde{\mathbf{x}}_f) + E_{\tilde{\mathbf{b}}_f | \tilde{\mathbf{x}}_f; \hat{\mathbf{H}}_i} \{ \log f_{\tilde{\mathbf{x}}_f | \tilde{\mathbf{b}}_f; \mathbf{H}}(\tilde{\mathbf{x}}_f) \} \\ &= \arg \min_{\mathbf{H}} \sum_{k=0}^{M-1} \|\mathbf{x}_t(k) - \mathbf{H}\mathbf{z}_t(k)\|^2 \\ &\quad + E_{\tilde{\mathbf{b}}_f | \tilde{\mathbf{x}}_f; \hat{\mathbf{H}}_i} \left\{ \sum_{k=M}^{K-1} \|\mathbf{x}(k) - \mathbf{H}\mathbf{z}(k)\|^2 \right\}, \end{aligned} \quad (17)$$

where the last equality follows from the fact that, as far as we assume AWGN, the pdf of the observations conditioned on the transmitted symbols $f_{\tilde{\mathbf{x}}_f | \tilde{\mathbf{b}}_f; \mathbf{H}}$ is Gaussian. This leads to the following quadratic optimization problem:

$$\begin{aligned} \hat{\mathbf{H}}_{i+1} &= \arg \min_{\mathbf{H}} \sum_{k=0}^{M-1} \|\mathbf{x}_t(k) - \mathbf{H}\mathbf{z}_t(k)\|^2 \\ &\quad + \sum_{k=M}^{K-1} E_{\tilde{\mathbf{b}}_f | \tilde{\mathbf{x}}_f; \hat{\mathbf{H}}_i} \left\{ \|\mathbf{x}(k) - \mathbf{H}\mathbf{z}(k)\|^2 \right\} \end{aligned} \quad (18)$$

with the closed-form solution¹

$$\hat{\mathbf{H}}_{i+1} = (\mathbf{R}_{\mathbf{z}_t, t} + \mathbf{R}_{\mathbf{z}_z}) \times (\mathbf{R}_{\mathbf{z}_t, t} + \mathbf{R}_{\mathbf{z}_z})^{-1}, \quad (19)$$

¹Since the expectation operator is linear, the derivation leading to (19) follows, step by step, the usual optimization procedure to find the LS estimate of a linear system given a set of noisy observations (see, e.g., [34]). Such a procedure includes the calculation of the gradient with respect to the system coefficients and then solving for the points where the gradient vanishes. Hence, solving (17) is tedious, since derivatives have to be computed for the coefficients in matrix \mathbf{H} , but conceptually straightforward.

where

$$\mathbf{R}_{\mathbf{z}_t, t} = \sum_{k=0}^{M-1} \mathbf{x}_t(k) \mathbf{z}_t^H(k), \quad (20)$$

$$\mathbf{R}_{\mathbf{z}_t} = \sum_{k=0}^{M-1} \mathbf{z}_t(k) \mathbf{z}_t^H(k), \quad (21)$$

$$\mathbf{R}_{\mathbf{z}_z} = \sum_{k=M}^{K-1} E_{\tilde{\mathbf{b}}_f | \tilde{\mathbf{x}}_f; \hat{\mathbf{H}}_i} \{ \mathbf{x}(k) \mathbf{z}^H(k) \}, \quad (22)$$

$$\mathbf{R}_z = \sum_{k=M}^{K-1} E_{\tilde{\mathbf{b}}_f | \tilde{\mathbf{x}}_f; \hat{\mathbf{H}}_i} \{ \mathbf{z}(k) \mathbf{z}^H(k) \}. \quad (23)$$

Note that for computing (22) and (23), it is necessary to know the probability mass function $p_{\mathbf{z}(k) | \tilde{\mathbf{x}}_f; \hat{\mathbf{H}}_i}$. Towards this aim, we take benefit from the Turbo equalization process because

$$L[\mathbf{z}(k) | \tilde{\mathbf{x}}; \hat{\mathbf{H}}_i] = L[\mathbf{z}(k), \tilde{\mathbf{x}}; \hat{\mathbf{H}}_i] + h_z = L[e_k, \tilde{\mathbf{x}}] + h_z, \quad (24)$$

where h_z is the constant that makes $p_{\mathbf{z}(k) | \tilde{\mathbf{x}}; \hat{\mathbf{H}}_i}$ a probability mass function and $L[e_k, \tilde{\mathbf{x}}]$ is the joint log-probability of the transition e_k and the set of available observations. Notice that this quantity has already been computed in the Turbo equalization process (see (7)). This fact makes the proposed channel estimator very suitable to be used within a Turbo equalization structure.

5. APPLICATION TO AN STC SYSTEM FOR SUBWAY ENVIRONMENTS

We focus now on the application of the ML-EM channel estimator described in Section 4 to an STC GSM-like system for underground railway transportation systems. Some practical considerations follow. In subway tunnel environments, propagation conditions result in flat multipath fading because its delay spread is small when compared to the GSM symbol period [35]. Nevertheless, the modulation employed by the GSM standard, Gaussian minimum shift keying (GMSK), induces controlled ISI and thus Turbo ST Equalization can be employed for the purpose of joint demodulating and decoding. In addition, experimental measurements [36] have revealed that in this environment, there exist strong spatial correlations between subchannels. These spatial correlations will be taken into account when evaluating the receivers' performance in the following section because we will use, in the computer simulations, experimental measurements of MIMO channel impulse responses obtained in subway tunnels. These field measurements have been carried out in the framework of the European project “ESCORT” [37]. We will show how the proposed channel estimator allows to reduce the necessary length of the training sequence from 26 bits in the GSM standard up to only 5 bits, while performance is maintained very close to the optimum (i.e., the bit error rate (BER) obtained when the channel is perfectly known at reception) which clearly implies a very high gain in the overall system throughput.

Figure 1 can be useful again for modeling the STC transmitter under consideration (for the sake of clarity, we refer the reader to Appendix A for a detailed description). This model can be summarized as follows. Each component of $\mathbf{b}(k)$ is independently modulated using the GMSK modulation format. GMSK is a partial response continuous phase modulation (CMP) signal and thus a nonlinear modulation format. Nevertheless, it can be expressed in terms of its Laurent expansion [38, 39, 40] as the sum of 2^{p-1} PAM signals, where p is the memory induced by the modulation. For the GMSK format in the GSM standard, $p = 3$ but the first PAM component contains 99.63% of the total GMSK signal energy [39, 40], so we can approximate the signal radiated by the i th antenna as

$$s_i(t; \mathbf{b}_i) \approx \sqrt{\frac{2E_b}{T}} \sum_{k=-\infty}^{\infty} a_i(k)h(t - kT), \quad (25)$$

where E_b is the bit energy, T the symbol period, $a_i(k) = ja_i(k-1)b_i(k)$ are the transmitted symbols which belong to a QPSK constellation, $\mathbf{b}_i = \{b_i(k)\}_{k=-\infty}^{\infty}$ is the bit sequence to be modulated, and $h(t)$ is a pulse waveform that spans along the interval $[0, pT]$, where p is the memory of the modulation. It is demonstrated in [38] that the transmitted symbols $a_i(k)$ are uncorrelated and have unit variance. In order to simplify the detection process at the receiver, we will assume that a differential precoder is employed prior to modulation, that is, $d_i(k) = b_i(k-1)b_i(k)$ because we have then $a_i(k) = ja_i(k-1)d_i(k) = j^k b_i(k)$.

Considering that the transmission channel inside subway tunnels suffers from flat multipath fading [35], the signal received at the l th antenna is

$$y_l(t) = \sum_{i=1}^N h_{li}s_i(t; \mathbf{b}_i) + n_l(t), \quad (26)$$

where h_{li} is the fading observed between the i th transmitting antenna and the l th receiving antenna and $n_l(t)$ is a continuous-time complex-valued white Gaussian process with power spectral density $N_0/2$.

The received signals $y_l(t)$ are passed through a bank of filters matched to the pulse waveform $h(t)$ and sampled at the symbol rate in order to obtain a set of sufficient statistics for the detection of the transmitted symbols. Because $h(t)$ does not satisfy the zero-ISI condition, a discrete-time whitening filter [41, 42] is located after sampling. In addition, the rotation j^k induced by the GMSK modulation is compensated by multiplying the received signal by j^{-k} , resulting in the following expression for the observations:

$$\begin{aligned} x_l(k) &= \sum_{i=1}^N h_{li} \sum_{m=0}^{p-1} f(m)b_i(k-m) + v_l(k) \\ &= \sum_{i=1}^N h_{li}s_i(k) + v_l(k), \end{aligned} \quad (27)$$

where $v_l(k)$ represents the complex-valued AWGN with variance σ_v^2 and $f(m) = [0.8053, -0.5853j, -0.0704]$ is the equivalent discrete-time impulse response that takes into ac-

count the transmitting, receiving, and whitening filters, and the derotation operation. Using vector notation, the output of the whitening filters after the derotation can be expressed as

$$\mathbf{x}(k) = \mathcal{H}\mathbf{s}(k) + \mathbf{v}(k), \quad (28)$$

where $\mathbf{x}(k) = [x_1(k) \ x_2(k) \ \cdots \ x_L(k)]^T$ and

$$\mathcal{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22} & \cdots & h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{L1} & h_{L2} & \cdots & h_{LN} \end{bmatrix}. \quad (29)$$

Equation (28) can be rewritten in the form of (1) as

$$\begin{aligned} \mathbf{x}(k) &= [f(0)\mathcal{H} \ f(1)\mathcal{H} \ f(2)\mathcal{H}] \begin{bmatrix} \mathbf{b}(k-2) \\ \mathbf{b}(k-1) \\ \mathbf{b}(k) \end{bmatrix} + \mathbf{v}(k) \\ &\equiv \mathbf{H}\mathbf{z}(k) + \mathbf{v}(k). \end{aligned} \quad (30)$$

However, this signal model for the observations does not emphasize that the ISI comes from the GMSK modulation format instead of the time-dispersion of the multipath channel. As a consequence, we prefer to rewrite (28) as

$$\mathbf{x}(k) = \mathcal{H}\mathbf{B}(k)\mathbf{f} + \mathbf{v}(k), \quad (31)$$

where

$$\begin{aligned} \mathbf{B}(k) &= [\mathbf{b}(k) \ \mathbf{b}(k-1) \ \mathbf{b}(k-2)], \\ \mathbf{f} &= [0.8053, -0.5853j, -0.0704]^T. \end{aligned} \quad (32)$$

5.1. ML channel estimation for STC GSM-like systems with flat fading

Estimating the channel according to (30) and directly applying the method described in the previous section is highly inefficient because we have to estimate an unnecessarily large number of parameters. In addition, this way we do not take into account the knowledge at reception of the controlled ISI introduced by the modulator, given by $f(m)$. Equation (31) is preferable because it enables us to formulate the estimation of only the unknown channel coefficients h_{li} , as it is explained in the sequel. Again, we assume that the transmitted data frames contain a known pilot sequence of length M . The ML estimator of the channel is given by

$$\widehat{\mathcal{H}} = \arg \max_{\mathcal{H}} f_{\mathbf{x}|\bar{\mathbf{b}}_i; \mathcal{H}}(\bar{\mathbf{x}}). \quad (33)$$

This is a problem without closed-form solution, so we will apply the EM algorithm in a similar way to the general case explored in Section 4. We define the complete and incomplete data sets as $\bar{\mathbf{x}}_e = [\bar{\mathbf{b}}_f^T \ \bar{\mathbf{x}}_f^T]^T$ and $\bar{\mathbf{x}}_f$, respectively. Both sets are related through the linear transformation $\bar{\mathbf{x}}_f = \mathbf{T}\bar{\mathbf{x}}_e$, where $\mathbf{T} = [\mathbf{0}_{L(M+K) \times N(M+K)} \ \mathbf{I}_{L(M+K)}]$. Using the latter definitions, the $i+1$ th estimate of the channel is computed using

the EM method as

$$\widehat{\mathcal{H}}_{i+1} = \arg \max_{\mathcal{H}} E_{\tilde{\mathbf{x}}_e | \tilde{\mathbf{x}}_f, \tilde{\mathbf{b}}_i; \widehat{\mathcal{H}}_i} \{ \log f_{\tilde{\mathbf{x}}_e | \tilde{\mathbf{b}}_i; \mathcal{H}}(\tilde{\mathbf{x}}_e) \}. \quad (34)$$

Making similar manipulations to those made for the time-dispersive MIMO channel, we arrive at the following optimization problem:

$$\begin{aligned} \widehat{\mathcal{H}}_{i+1} = \arg \min_{\mathcal{H}} & \sum_{k=0}^{M-1} \|\mathbf{x}_t(k) - \mathcal{H} \mathbf{B}_t(k) \mathbf{f}\|^2 \\ & + \sum_{k=M}^{K-1} E_{\mathbf{b}(k) | \tilde{\mathbf{x}}; \widehat{\mathcal{H}}_i} \{ \|\mathbf{x}(k) - \mathcal{H} \mathbf{B}(k) \mathbf{f}\|^2 \} \end{aligned} \quad (35)$$

which is also a quadratic optimization problem whose solution is

$$\widehat{\mathcal{H}}_{i+1} = (\mathbf{R}_{x_b, t} + \mathbf{R}_{x_b}) \times (\mathbf{R}_{b, t} + \mathbf{R}_b)^{-1}, \quad (36)$$

where

$$\begin{aligned} \mathbf{R}_{x_b, t} &= \sum_{k=0}^{M-1} \mathbf{x}_t(k) (\mathbf{B}_t(k) \mathbf{f})^H, \\ \mathbf{R}_{b, t} &= \sum_{k=0}^{M-1} (\mathbf{B}_t(k) \mathbf{f}) (\mathbf{B}_t(k) \mathbf{f})^H, \\ \mathbf{R}_{x_b} &= \sum_{k=M}^{K-1} E_{\mathbf{b}(k) | \tilde{\mathbf{x}}; \widehat{\mathcal{H}}_i} \{ \mathbf{x}(k) (\mathbf{B}(k) \mathbf{f})^H \}, \\ \mathbf{R}_b &= \sum_{k=M}^{K-1} E_{\mathbf{b}(k) | \tilde{\mathbf{x}}; \widehat{\mathcal{H}}_i} \{ (\mathbf{B}(k) \mathbf{f}) (\mathbf{B}(k) \mathbf{f})^H \}. \end{aligned} \quad (37)$$

Here we need to average with respect to the pdf $f_{\mathbf{B}(k) | \tilde{\mathbf{x}}; \widehat{\mathcal{H}}_i}$. Again, we take benefit from the Turbo equalization process because

$$L[\mathbf{B}(k) | \tilde{\mathbf{x}}; \widehat{\mathcal{H}}_i] = L[\mathbf{B}(k), \tilde{\mathbf{x}}; \widehat{\mathcal{H}}_i] + h_B = L[e_k, \tilde{\mathbf{x}}] + h_B, \quad (38)$$

where h_B is the constant that makes $p_{\mathbf{B}(k) | \tilde{\mathbf{x}}; \widehat{\mathcal{H}}_i}$ a probability mass function and $L[e_k, \tilde{\mathbf{x}}]$ is a quantity already computed in the Turbo equalization process.

6. SIMULATION RESULTS

6.1. Rayleigh MIMO channel

Computer simulations were carried out to illustrate the performance of the proposed channel estimator. Figure 3 plots the BER after decoding obtained for a 2×2 STC system over a nondispersive channel. Data are transmitted in blocks of 218 bits out of which the pilot sequence occupies $M = 10$ bits. The performance curves for both the LS method and when the channel is perfectly known are also shown for comparison. Note that there is no iteration gain when the channel is known because there is no ISI and, therefore, no ‘‘inner coding’’ for the Turbo processing. Nevertheless, this is not true when the ML-EM channel estimator is used because the channel is reestimated at each iteration of the Turbo equalization process. The ST encoder is a rate 1/2 full diversity convolutional binary code with generating matrix

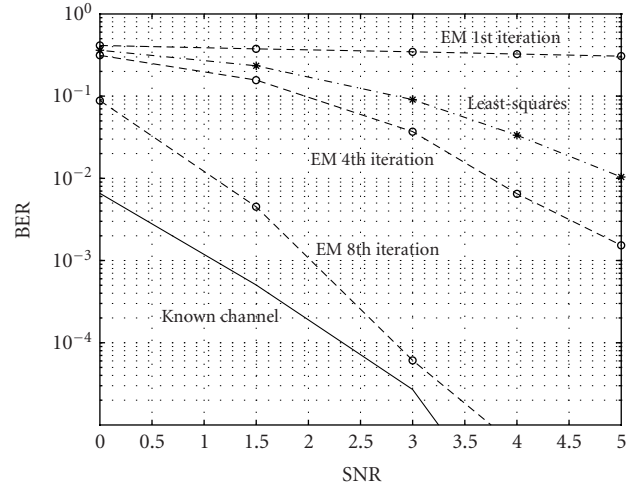


FIGURE 3: Performance results for ST coded data over a nondispersive channel.

$\mathbf{G} = [46, 72]$ in octal representation [26]. The independent interleavers are 20800 bits long each. The modulation format is BPSK and each channel coefficient is modeled as a zero-mean, complex-valued, circularly invariant Gaussian random process. Consequently, their magnitudes are Rayleigh distributed. We have also assumed that the channel coefficients are both temporally and spatially independent, having variance $\sigma_h^2 = 1/2$ per complex dimension. The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = \frac{E\{(\mathbf{H}\mathbf{z}(k))^H (\mathbf{H}\mathbf{z}(k))\}}{E\{\mathbf{v}^H(k) \mathbf{v}(k)\}} = \frac{\text{Tr}\{\mathbf{H}\mathbf{H}^H\}}{L\sigma_v^2}, \quad (39)$$

where $\text{Tr}\{\cdot\}$ denotes the trace operator. The channel changes at each transmitted block. Figure 3 shows that, even if its result for the first iteration is very poor, the ML-EM channel estimator outperforms the classical LS method from the fourth iteration.

The bad performance obtained by the ML-EM estimator at the first iteration comes from the fact that the Turbo equalizer is using an uninformative initial estimate of the channel. Specifically, (19) can be viewed as an LS estimator, where the correlation matrices $\mathbf{R}_{x_z, t}$ and $\mathbf{R}_{z, t}$ have been modified by the addition of the matrices \mathbf{R}_{x_z} and \mathbf{R}_z , respectively. In the first iteration, these matrices are computed by assuming that $p_{\mathbf{z}(k) | \tilde{\mathbf{x}}; \widehat{\mathbf{H}}_i}$ is a uniform probability mass function (therefore, independent of the initial channel estimate $\widehat{\mathbf{H}}_0$) in (22) and (23). This results in a degradation of the pure LS estimator and a very high symbol error rate (SER) after decoding. Such a high SER (around 0.4) can never lead the Turbo equalization process to convergence. However, in our case, convergence is achieved because, in the next iterations, a substantial improvement is obtained in channel estimation from the EM algorithm (not from the Turbo structure itself). Notice that one iteration of the EM algorithm (19) is performed only after one complete equalization and decoding step. Anyway, once the channel estimate is good enough for the Turbo

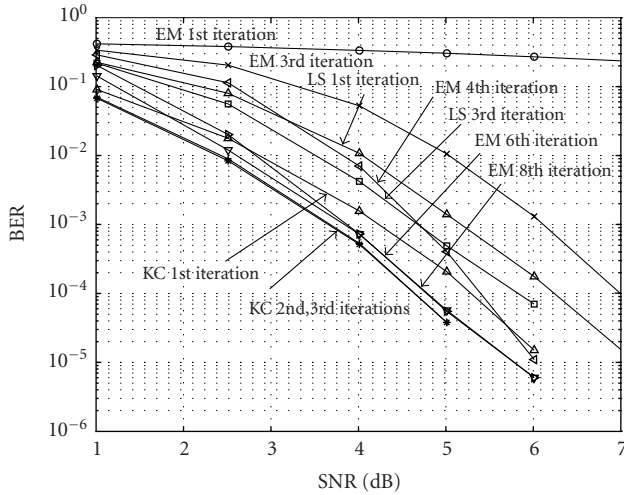


FIGURE 4: Performance results for ST coded data over a dispersive channel with memory $m = 2$.

equalization structure to lie in its convergence region, both the EM algorithm and the Turbo iterative process help in reducing the error rate. Figure 3 also shows that at the eighth iteration, the performance is very close to the optimum, that is, known channel case. Only 0.5 dB separates the two curves at a BER of 10^{-4} .

Figure 4 shows the results (BER after decoding) obtained when a time-dispersive MIMO channel with memory $m = 2$ is considered. The simulation parameters are the same as in Figure 3. In particular, note that, again, each channel coefficient has variance $\sigma_h^2 = 1/2$ per complex dimension. It is apparent that at the fourth iteration, the ML-EM estimator performs very similar to the LS method, which does not improve significantly through the iterations. At the eighth iteration, the performance of the ML-EM estimator is again very close to the known channel case.

6.2. GSM-based transmission over subway tunnel MIMO channels

The performance of the proposed GSM-based transmission system with a Turbo STC receiver in subway tunnel environments has also been tested through computer simulations. The channel matrices \mathcal{H} result from experimental measurements (carried out within the framework of the European project "ESCORT") of the MIMO channel impulse response present in a subway tunnel. The experimental setup consisted of four transmitting antennas, each one having a 12 dBi gain, located at the station platform, and four patch antennas located behind the train windscreen. The complex impulse responses were measured with a channel sounder having a bandwidth of 35 MHz by switching successively the antennas and stopping the train approximately each 2 m. From the whole set of 4×4 measured subchannels, only those corresponding to the furthest antennas were picked up for constructing a 2×2 system. In [35], it was demonstrated that the mean capacity of the measured channel is less than the ca-

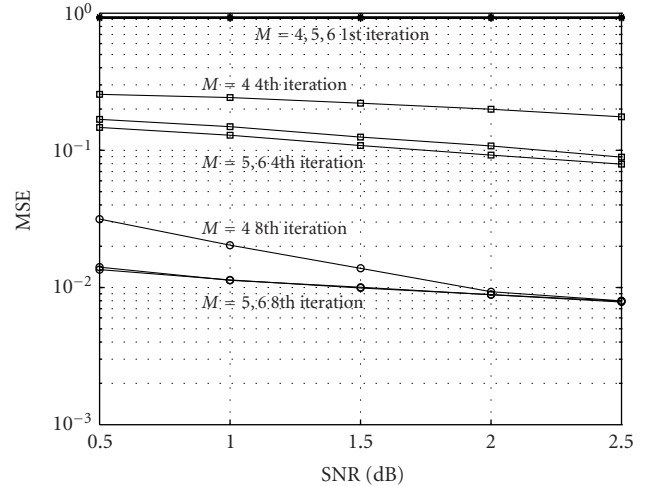


FIGURE 5: MSE for several lengths of the training sequence.

capacity of Rayleigh fading channels, this difference being more remarkable in the case of a 4×4 system.

The ability of our channel estimation technique to combine the deterministic information of the pilot symbols and the statistical information from the unknown symbols, thanks to the ST Turbo detector, enables us to considerably reduce the size of the training sequence in GSM systems. Indeed, by means of computer simulations, we have determined the minimum length of the training sequence for the considered GSM-based MIMO system. Figure 5 shows the channel estimation mean square error (MSE) for several values of the training sequence length ($M = 4, 5$, and 6 bits). The channel code is the same as in the previous simulations. The interleaver size is 20800 bits and the frame length is 148, as established in the GSM standard. There is a significant difference in the estimation error between using $M = 4$ bits and $M = 5$ bits, whereas the gap between $M = 5$ and $M = 6$ is very small. This points out that $M = 5$ bits is the minimum length for the training sequence. This assumption can also be corroborated in Figure 6, where the SER at the output of the decoder is plotted versus the required SNR.

Next, we compare the results obtained with the proposed estimator using a training sequence of $M = 5$ bits and those obtained with classical LS using a training sequence of $M = 26$ bits (the length standardized in GSM). The results obtained when the receiver perfectly knows the channel are also plotted for comparison. As it is shown in Figure 7, the proposed method (ML-EM) with $M = 5$ bits performs better than the LS with $M = 26$ bits beyond the sixth iteration, achieving a performance very close to the known channel case beyond the seventh iteration.

7. CONCLUSIONS

In this paper, we propose a novel ML-based time-dispersive MIMO channel estimator for STC systems that employ

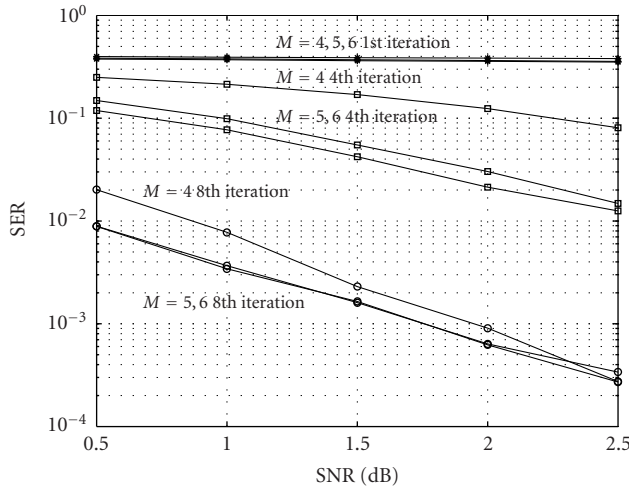


FIGURE 6: SER versus SNR at the output of the decoder for several lengths of the training sequence.

Turbo ST receivers. We formulate the ML estimation problem that takes into account the deterministic symbols corresponding to the training sequence and the statistics of the unknown symbols. These statistics can be obtained and successively refined if an ST Turbo equalizer is used at reception. This full exploitation of all the available statistical information at reception renders an extremely powerful channel estimation technique that outperforms conventional approaches based only on the training sequence. Since the involved optimization problem has no closed-form solution, the EM algorithm is employed in order to iteratively obtain the solution. The main limitation of our approach is that the computational complexity of the channel estimator grows exponentially with the number of transmitting antennas and the channel memory size, hence it is only practical for a moderate size of the transmitter antenna array. Note, however, that this complexity is inherent to the problem of optimal detection and estimation in MIMO systems.

The method has been particularized for a realistic scenario in which an STC system based on the GSM standard transmits along railway subway tunnels. Simulation results show how our channel estimation technique enables us to diminish the training sequence length up to only 5 bits, instead of the 26 bits considered in the GSM standard, thus achieving a 14% increase in the system throughput.

APPENDICES

A. SIGNAL MODEL OF AN STC GSM SYSTEM

The transmitter model depicted in Figure 1 is valid for an STC GSM system. The signal radiated by i th antenna is given by [38, 40]

$$s_i(t; \mathbf{b}_i) = \sqrt{\frac{2E_b}{T}} \exp \left[j\pi \sum_{k=-\infty}^{\infty} b(k)q(t - kT) \right], \quad (\text{A.1})$$

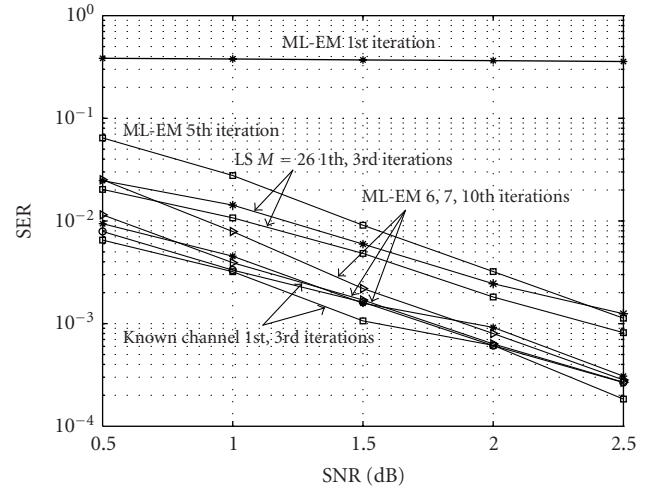


FIGURE 7: Performance comparison between ML-EM ($M = 5$ bits), LS ($M = 26$ bits), and known channel.

where E_b is the bit energy, T the symbol period, $\mathbf{b}_i = \{b_i(k)\}_{k=-\infty}^{\infty}$ the bit sequence to be modulated, and

$$q(t) = \int_{-\infty}^t g(\tau) d\tau, \quad (\text{A.2})$$

where $g(t)$ is the convolution between a Gaussian-shaped pulse and a rectangular-shaped pulse centered at the origin [43, 44], that is,

$$g(t) = u(t) * \text{rect}\left(\frac{t}{T}\right), \quad (\text{A.3})$$

where

$$\text{rect}\left(\frac{t}{T}\right) = \begin{cases} \frac{1}{2T}, & |t| \leq \frac{T}{2}, \\ 0, & \text{otherwise,} \end{cases} \quad (\text{A.4})$$

$$u(t) = \frac{1}{\sqrt{2\pi}\sigma_u} \exp\left\{-\frac{1}{2}\left(\frac{t}{\sigma_u}\right)^2\right\},$$

with

$$\sigma_u = \frac{\sqrt{\log 2}}{2\pi B}, \quad (\text{A.5})$$

where B is the 3 dB bandwidth of $u(t)$. It is possible to derive a closed-form expression for $g(t)$ given by [38, 40]

$$g(t) = \frac{1}{2T} \left[Q\left(\frac{t - T/2}{\sigma_u}\right) - Q\left(\frac{t + T/2}{\sigma_u}\right) \right], \quad (\text{A.6})$$

where

$$Q(t) = \frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-\tau^2/2} d\tau \quad (\text{A.7})$$

is the Gaussian complementary error function. With the aim

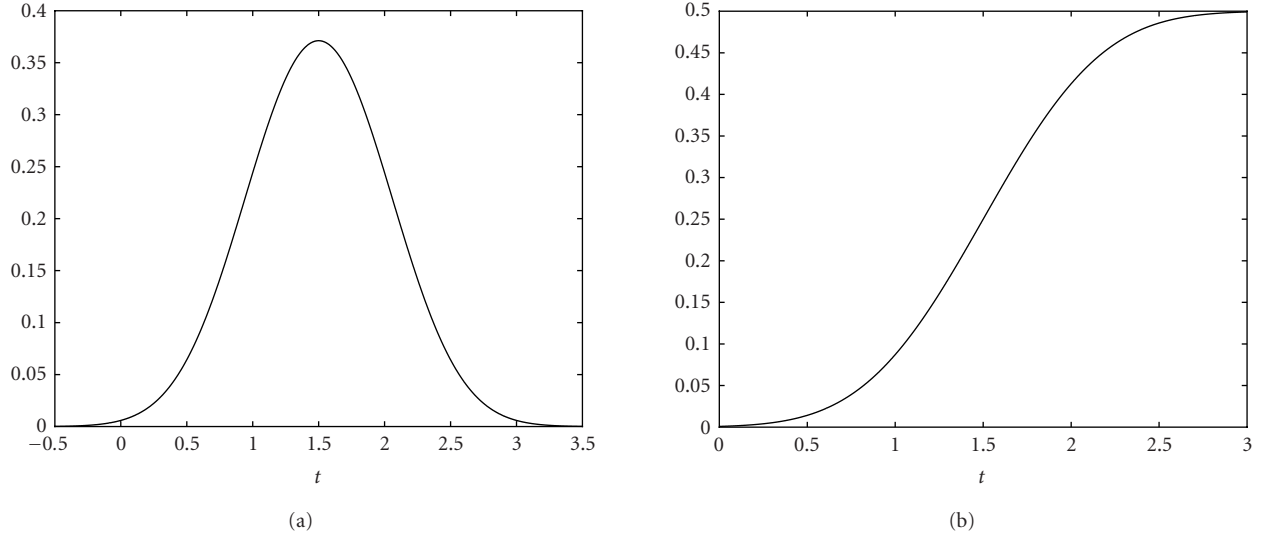


FIGURE 8: (a) Shifted GSMK pulse, $g(t - 1.5T)$, for $p = 3$. (b) GSMK phase pulse, $q(t)$.

of simplifying subsequent analysis, we redefine $g(t) \equiv g(t - p/2T)$, so it is limited to the interval $[0, pT]$, where p is the number of symbol periods where the signal has significant values. For GSM ($B = 0.3$), a value of $p = 3$ is reasonable [40], as it can be verified in Figure 8, that plot the properly shifted versions of $g(t)$ and $q(t)$ when $B = 0.3$.

Since GSMK is a partial response CPM, it can be expressed in terms of its Laurent expansion [38, 39, 40], formed by the sum of 2^{p-1} PAM signals, where p is the memory induced by the modulation. Since in GSM, the first PAM component contains 99.63% of the total GSMK signal energy [39, 40], we can approximate the signal radiated by the i th antenna by

$$s_i(t; \mathbf{b}_i) \approx \sqrt{\frac{2E_b}{T}} \sum_{k=-\infty}^{\infty} a_i(k)h(t - kT), \quad (\text{A.8})$$

where $a_i(k) = ja_i(k-1)b_i(k)$ are the transmitted symbols, which belong to a QPSK constellation, are uncorrelated and have unit variance [38]. In order to simplify the detection process at the receiver, we will assume that a differential precoder is employed prior to modulation, that is, $d_i(k) = b_i(k-1)b_i(k)$ because then we have $a_i(k) = ja_i(k-1)d_i(k) = j^k b_i(k)$. The pulse waveform $h(t)$ is equal to $C(t-3T)C(t-2T)C(t-T)$, where $C(t) = \cos(\pi q(|t|))$. Figure 9a shows that it takes significant values over the interval $[0.5T, 3.5T]$ because the actual and the linearized GSMK waveforms are shifted by half a symbol period.

In order to detect the transmitted symbols, $s_i(t; \mathbf{b}_i)$ is passed through a filter matched to the pulse waveform $h(t)$ and then sampled at the symbol rate. The output of the matched filter is given by

$$\begin{aligned} r_i(t) &= a_i(t) * h(t) * h^*(-t) + n(t) * h^*(-t) \\ &= a_i(t) * R_h(t) + g(t), \end{aligned} \quad (\text{A.9})$$

where

$$a_i(t) = \sqrt{\frac{2E_b}{T}} \sum_{k=-\infty}^{\infty} a_i(k)\delta(t - kT) \quad (\text{A.10})$$

and $R_h(t)$ (see Figure 9b) denotes the autocorrelation function of $h(t)$. After sampling, we have

$$r_i(k) \equiv r_i(t = kT) = a_i(k) * R_h(k) + g(k), \quad (\text{A.11})$$

where the autocorrelation function of $g(k)$ is $R_g(k) = (N_0/2)R_h(k)$. Clearly, the noise $g(k)$ is colored because $h(t)$ does not satisfy the zero-ISI condition. Since it is more comfortable to perform detection assuming white noise, a discrete-time whitening filter [41, 42] is located after sampling

$$W(z) = \frac{1}{F^*(z^{-1})}, \quad (\text{A.12})$$

where $F^*(z^{-1})$ comes from the factorization of the autocorrelation function $R_h(k) = F(z)F^*(z^{-1})$. This expression for the whitening filter leads to an overall system response given by $F(z)$. In Appendix B, we demonstrate that the maximum phase $F(z)$ polynomial is given by

$$\begin{aligned} F(z) &= \sqrt{\frac{r_2}{\rho_1 \rho_2}} (1 - \rho_1 z^{-1})(1 - \rho_2 z^{-1}) \\ &= 0.8053 + 0.5853z^{-1} + 0.0704z^{-2}, \end{aligned} \quad (\text{A.13})$$

where $\rho_1 = -0.1522$, $\rho_2 = -0.5746$, and $r_2 = R_h(-2)$.

In addition, the rotation j^k induced by the GSMK is compensated by multiplying the received signal by j^{-k} , resulting

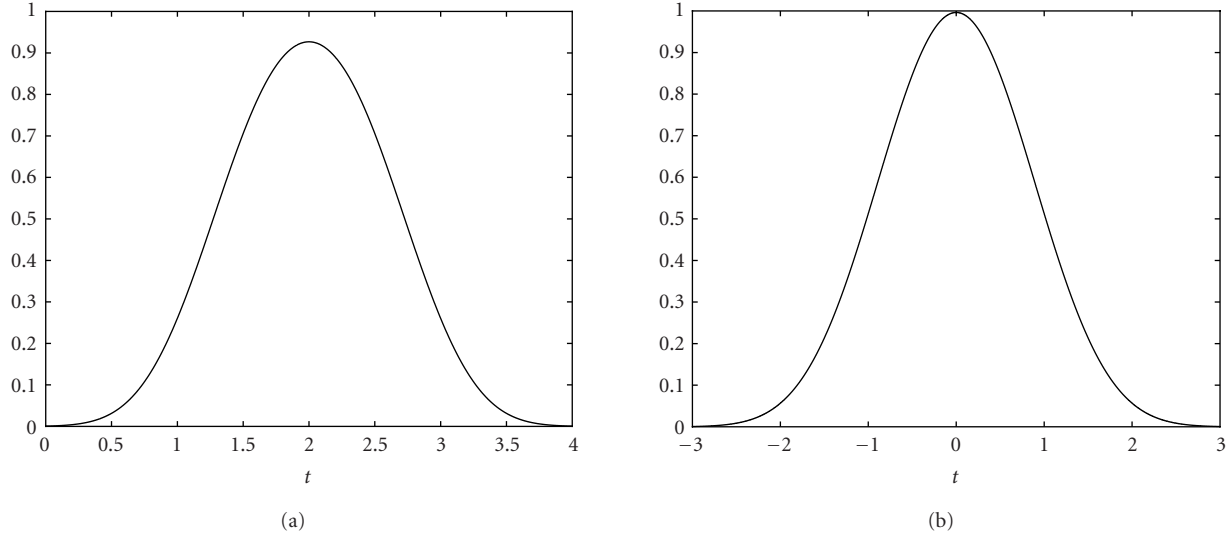


FIGURE 9: (a) Pulse shape, $h(t)$. (b) Autocorrelation function, $R_h(t)$.

in the following expression for the observations:

$$\begin{aligned} x_l(k) &= \sum_{i=1}^N h_{li} \sum_{m=0}^{p-1} f(m) b_i(k-l) + v_l(k) \\ &= \sum_{i=1}^N h_{li} s_i(k) + v_l(k), \end{aligned} \quad (\text{A.14})$$

where $v_l(k)$ is AWGN with variance σ_v^2 and $f(m) = [0.8053 \ -0.5853j \ -0.0704]$ is the equivalent discrete-time impulse response that takes into account the transmitting, receiving, and whitening filters, and the derotation operation. Using vector notation, the output of the whitening filters after the derotation can be expressed as in (28), where $\mathbf{x}(k) = [x_1(k) \ x_2(k) \ \cdots \ x_L(k)]^T$, $\mathbf{s}(k) = [s_1(k) \ s_2(k) \ \cdots \ s_N(k)]^T$, and is as in (29).

B. COMPUTATION OF THE DISCRETE-TIME WHITENING FILTER

First, it is important to note that there are 2^p choices of $F(z)$ that satisfy the desired factorization $R_h(z) = F(z)F^*(z^{-1})$. The different choices yield filters $1/F^*(z^{-1})$ that have the same magnitude but different phase response. One possible choice is to select $F^*(z^{-1})$ so that it is a minimum phase, that is, with all its roots inside the unit circle. In this way, $1/F^*(z^{-1})$ is a realizable causal and stable discrete system. The problem of this selection is that the overall impulse response $F(z)$ will be the maximum phase and anticausal, and the resulting ISI will be difficult to compensate. To overcome this limitation, we choose $F^*(z^{-1})$ to be the maximum phase and thus the whitening filter $1/F^*(z^{-1})$ will be stable only if it is considered anticausal. Nevertheless, anticausal filters can be implemented if a sufficient large delay is introduced. The advantage of this approach is that now the overall impulse response $F(z)$ is causal and minimum phase.

Considering that $R_h(t)$ takes significant values only over the interval $[-2, 2]$ (see Figure 9b), we have

$$\begin{aligned} R_h(k) &= \{r_{-2}, r_{-1}, r_0, r_1, r_2\} = \{r_2^*, r_1^*, r_0, r_1, r_2\} \\ &= \{0.0567, 0.5127, 0.9963, 0.5127, 0.0567\}. \end{aligned} \quad (\text{B.1})$$

As mentioned before, note that $h(t)$ contains 99.63% of the actual GMSK total energy because $R_h(0) = 0.9963$. The Z-transform of $R_h(k)$ is

$$R_h(z) = r_{-2}z^2 + r_{-1}z + r_0 + r_1z^{-1} + r_2z^{-2} \quad (\text{B.2})$$

that we can express as

$$R_h(z) = r_2 \left(\frac{z-1}{\rho_1^*} \right) \left(\frac{z-1}{\rho_2^*} \right) (1 - \rho_1 z^{-1})(1 - \rho_2 z^{-1}). \quad (\text{B.3})$$

Forcing $|\rho_1|, |\rho_2| \leq 1$ in order that the resulting whitening filter exists and be stable, we have $\rho_1 = -0.1522$ and $\rho_2 = -0.5746$. Taking into account that ρ_1 and ρ_2 are real valued, we arrive at

$$F(z) = \sqrt{\frac{r_2}{\rho_1 \rho_2}} (1 - \rho_1 z^{-1})(1 - \rho_2 z^{-1}) \quad (\text{B.4})$$

and thus the whitening filter is given by

$$W(z) = \frac{1}{F^*(z^{-1})} = \frac{\sqrt{\rho_1 \rho_2 / r_2}}{(1 - \rho_1 z)(1 - \rho_2 z)} \quad (\text{B.5})$$

whose inverse Z-transform is

$$w(k) = \{w_k\}_{k=-\infty}^0 = \left\{ \frac{\sqrt{\rho_1 \rho_2 / r_2}}{\rho_2 - \rho_1} (\rho_2^{-k+1} - \rho_1^{-k+1}) \right\}. \quad (\text{B.6})$$

Since $\{|w_k|\}_{k=0}^{-\infty}$ is a strictly decreasing series, we can consider only the first significant w_k coefficients. Taking into account

that $|w_{-20}| < 10^{-4}$, we can implement $w(k)$ as an anticausal FIR filter:

$$\begin{aligned} w(k) &\approx \{w_{-19}, w_{-18}, \dots, w_{-1}, w_0\} \\ &= \{0.0001, -0.0001, 0.0002, -0.0004, 0.0007, \\ &\quad -0.0013, 0.0022, -0.0038, 0.0066, -0.0115, \\ &\quad 0.0201, -0.0349, 0.0608, -0.1058, 0.1839, \\ &\quad 0.3189, 0.5473, -0.9025, 1.2417\}. \end{aligned} \quad (\text{B.7})$$

ACKNOWLEDGMENT

This work has been supported by the European Commission under Contract no. IST-1999-20006 (ESCORT project) and by Ministerio de Ciencia y Tecnología of Spain and FEDER funds from the European Union under Grant no. TIC2001-0751-C04-01.

REFERENCES

- [1] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Trans. Communications*, vol. 44, no. 10, pp. 1261–1271, 1996.
- [2] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 429–445, 1996.
- [3] C. Heegard and S. B. Wicker, *Turbo Coding*, Kluwer Academic Publishers, Boston, Mass, USA, 1999.
- [4] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "A soft-input soft-output APP module for iterative decoding of concatenated codes," *IEEE Communications Letters*, vol. 1, no. 1, pp. 22–24, 1997.
- [5] C. Douillard, A. Picart, P. Didier, M. Jézéquel, C. Berrou, and A. Glavieux, "Iterative correction of intersymbol interference: turbo-equalization," *European Transactions on Telecommunications*, vol. 6, no. 5, pp. 507–511, 1995.
- [6] C. Luschi, M. Sandell, P. Strauch, et al., "Advanced signal-processing algorithms for energy-efficient wireless communications," *Proceedings of the IEEE*, vol. 88, no. 10, pp. 1633–1650, 2000.
- [7] X. Wang and R. Chen, "Blind turbo equalization in Gaussian and impulsive noise," *IEEE Trans. Vehicular Technology*, vol. 50, no. 4, pp. 1092–1105, 2001.
- [8] C. Laot, A. Glavieux, and J. Labat, "Turbo equalization: adaptive equalization and channel decoding jointly optimized," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1744–1752, 2001.
- [9] Z. Yang and X. Wang, "Turbo equalization for GMSK signaling over multipath channels based on the Gibbs sampler," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1753–1763, 2001.
- [10] B. L. Yeap, T. H. Liew, J. Hámorský, and L. Hanzo, "Comparative study of turbo equalization schemes using convolutional, convolutional turbo, and block-turbo codes," *IEEE Transactions on Wireless Communications*, vol. 1, no. 2, pp. 266–273, 2002.
- [11] N. Görtz, "On the iterative approximation of optimal joint source-channel decoding," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1662–1670, 2001.
- [12] J. Garcia-Frias and J. D. Villasenor, "Joint turbo decoding and estimation of hidden Markov sources," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1671–1679, 2001.
- [13] A. Guyader, E. Fabre, C. Guillemot, and M. Robert, "Joint source-channel turbo decoding of entropy-coded sources," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1680–1696, 2001.
- [14] L. Zhang and A. Burr, "APPA symbol timing recovery scheme for turbo-codes," in *Proc. 13th IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications Conference*, Lisbon, Portugal, September 2002.
- [15] M. C. Valenti and B. D. Woerner, "Iterative channel estimation and decoding of pilot symbol assisted turbo codes over flat-fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1697–1705, 2001.
- [16] C. Komninakis and R. D. Wesel, "Joint iterative channel estimation and decoding in flat correlated Rayleigh fading," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1706–1717, 2001.
- [17] K.-D. Kammeyer, V. Kühn, and T. Petermann, "Blind and nonblind turbo estimation for fast fading GSM channels," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1718–1728, 2001.
- [18] A. Mehrotra, *GSM System Engineering*, Artech House, Boston, Mass, USA, 1997.
- [19] A. O. Berthet, B. S. Ünäl, and R. Visoz, "Iterative decoding of convolutionally encoded signals over multipath Rayleigh fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1729–1743, 2001.
- [20] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, John Wiley & Sons, New York, NY, USA, 1997.
- [21] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.
- [22] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.
- [23] A. Grant, "Rayleigh fading multi-antenna channels," *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 3, pp. 316–329, 2002.
- [24] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.
- [25] A. F. Naguib, N. Seshadri, and A. R. Calderbank, "Increasing data rate over wireless channels," *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 76–92, 2000.
- [26] A. R. Hammons Jr. and H. El Gamal, "On the theory of space-time codes for PSK modulation," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 524–542, 2000.
- [27] X. Lin and R. S. Blum, "Improved space-time codes using serial concatenation," *IEEE Communications Letters*, vol. 4, no. 7, pp. 221–223, 2000.
- [28] H.-J. Su and E. Geraniotis, "Space-time turbo codes with full antenna diversity," *IEEE Trans. Communications*, vol. 49, no. 1, pp. 47–57, 2001.
- [29] Y. Liu, M. P. Fitz, and O. Y. Takeshita, "Full rate space-time turbo codes," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 5, pp. 969–980, 2001.
- [30] S. Lek, "Turbo space-time processing to improve wireless channel capacity," *IEEE Trans. Communications*, vol. 48, no. 8, pp. 1347–1359, 2000.
- [31] A. Stefanov and T. M. Duman, "Turbo-coded modulation for systems with transmit and receive antenna diversity over block fading channels: system model, decoding approaches, and practical considerations," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 5, pp. 958–968, 2001.

- [32] M. González-López, A. Dapena, and L. Castedo, "MAP space-time receivers for GSM in subway tunnel environments," in *Proc. 11th European Signal Processing Conference*, Toulouse, France, September 2002.
- [33] M. González-López, J. Míguez, and L. Castedo, "Decision feedback Turbo equalization for space-time coded systems," in *Proc. IEEE 28th Int. Conf. Acoustics, Speech, Signal Processing*, Hong Kong, China, April 2003.
- [34] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1992.
- [35] M. González-López, A. Dapena, and L. Castedo, "Space-time coding for GSM systems in subway tunnel environments," in *Proc. IEEE 27th Int. Conf. Acoustics, Speech, Signal Processing*, Orlando, Fla, USA, May 2002.
- [36] J. Baudet, M. González-López, D. Degardin, et al., "Performance of space time coding in subway tunnel environments," in *Proc. IEE Technical Seminar on MIMO Communication Systems: from Concept to Implementation*, pp. 2/1–2/6, London, UK, December 2001.
- [37] ESCORT, "Enhanced diversity and space-time coding for metro and railway transmission," Final Tech. Rep. D 6021, France, 2002.
- [38] U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*, Plenum Press, New York, NY, USA, 1997.
- [39] N. Al-Dhahir and G. Saulnier, "A high-performance reduced-complexity GMSK demodulator," *IEEE Trans. Communications*, vol. 46, no. 11, pp. 1409–1412, 1998.
- [40] N. Al-Dhahir and G. Saulnier, "A high-performance reduced-complexity GMSK demodulator," Tech. Rep. 96CRD107, GE Global Research, Niskayuna, NY, USA, 1996.
- [41] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, NY, USA, 3rd edition, 1995.
- [42] J. Kurzweil, *An Introduction to Digital Communications*, John Wiley & Sons, New York, NY, USA, 2000.
- [43] J. M. H. Rábanos, *Comunicaciones Móviles GSM*, Fundación Airtel Móvil, Madrid, Spain, 1999.
- [44] J. D. Laster, *Robust GMSK Demodulation using demodulator diversity and BER estimation*, Ph.D. thesis, Virginia Polytechnic Institute and State University, Blacksburg, Va, USA, 1997.

research interests are in the field of statistical signal processing with emphasis on the topics of Bayesian analysis, sequential Monte Carlo methods, adaptive filtering, stochastic optimization, and their applications to multiuser communications, smart antenna systems, target tracking, and vehicle positioning and navigation.

Luis Castedo was born in Santiago de Compostela, Spain, in 1966. He received his Ingeniero de Telecomunicación (M.S.) and Doctor Ingeniero de Telecomunicación (Ph.D.) degrees, both from Universidad Politécnica de Madrid (UPM), Spain, in 1990 and 1993, respectively. From 1990 to 1994, he was with the Departamento de Señales, Sistemas y Radiocomunicación at the UPM, where he worked in array processing applied to digital communications. During the academic year 1991/92, he was a Visiting Scholar at the University of Southern California, USA. In 1994, he joined the Departamento de Electrónica y Sistemas at Universidad da Coruña, Spain, where he is currently a Professor and teaches courses in signal processing, digital communications, and linear control systems. His research interests include adaptive filtering and signal processing methods for space and code diversity exploitation in communication systems.



Miguel González-López was born in Santiago de Compostela, Spain, in 1977. He received his Ingeniero en Informática (M.S.) degree from Universidade da Coruña in 2000, where he is currently working to obtain his Ph.D. degree. His research interests include the application of the Turbo principle to channel estimation/equalization and coding on graphs, with special focus on their generalization to MIMO systems and their implementation issues.



Joaquín Míguez was born in Ferrol, Galicia, Spain, in 1974. He obtained his Licenciado en Informática (M.S.) and Doctor en Informática (Ph.D.) degrees from Universidade da Coruña, Spain, in 1997 and 2000, respectively. Late in 2000, he joined the Departamento de Electrónica y Sistemas, Universidade da Coruña, where he became an Associate Professor in July 2003. From April 2001 through December 2001, he was a Visiting Scholar in the Department of Electrical and Computer Engineering, the State University of New York at Stony Brook. His

