A Sub-Optimal Iterative Method for Maximum Likelihood Sequence Estimation in a Multipath Context *

Olivier Rousseaux, Geert Leus†, Marc Moonen
Katholieke Universiteit Leuven, ESAT-SISTA,
Kasteelpark Arenberg 10, B-3001 Heverlee
Tel:+32-16-321924, Fax:+32-16-321970
(olivier.rousseaux, geert.leus, marc.moonen)@esat.kuleuven.ac.be

June 20, 2002

Abstract

In this paper, we present an iterative semi-blind sub-optimal Maximum Likelihood Sequence Estimation (MLSE) method for single-carrier block transmission over stationary multipath channels. This sub-optimal ML detector is based on an Iterative Least Squares with Projection (ILSP) algorithm exploiting both the finite alphabet properties of the transmitted signal and its cyclic prefixed structure in order to approach ML detection in a cheap way. Since the initial channel estimate is crucial for the convergence speed of the ILSP algorithm, we propose a new low complexity stochastic method for providing an initial channel estimate. We therefore rely on some known symbols that are provided by a variant of Cyclic Prefix Only (CP-Only) transmission, known as the Known Symbol Padding Only (KSP-Only) technique. The resulting channel model is sufficiently accurate to be used as a starting point for the iterations. The final result is a direct symbol estimation method that is characterised by its low computational complexity and its promising results in terms of Bit Error Rate (BER).

*This research work was carried out at the ESAT laboratory of the K.U.Leuven, in the frame of the Belgian State, Prime Minister's Office - Federal Office for Scientific, Technical and Cultural Affairs - Interuniversity Poles of Attraction Programme - IUAP P4-02 (1997-2001): Modeling, Identification, Simulation and Control of Complex Systems, the Concerted Research Action GOA-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology)of the Flemish Government, Research Project FWO nr. G.0196.02 and was partially sponsored by IMEC (Flemish Interuniversity Microelectronics Center)

†Geert Leus is a Postdoctoral Fellow of the FWO

1 Introduction

The constantly increasing need for high data rate transmission systems has driven the research in broadband communications in the last years. Multipath effects, resulting in frequency-selective fading, are a major impediment of these broadband communication systems since they introduce Inter Symbol Interference (ISI), which needs to be tackled by appropriate techniques. However, these are most often computationally demanding. In this context, block transmission techniques (multi-carrier or single-carrier) based on the use of a Cyclic-Prefix (CP) have attracted a lot of attention in the last years for they allow an efficient and computationally cheap ISI cancellation procedure [1], [2]. ISI can be suppressed by a single-tap "frequency-domain" equalisation on blocks of data symbols using FFT and IFFT operations relying on the fact that the blocks of transmitted data are made cyclic by the use of a CP, whose length is at least equal to the channel order. More details on these techniques are presented in section 1.1.

In this paper we consider the case where the transmitter sends data over an unknown convolutive channel and where the receiver has no a priori information on that channel. The problem in this case consists of estimating the transmitted data without having an explicit channel model. The different approaches to solving this problem fall under two categories. The first category of methods aims at computing an equaliser that will filter the received data in order to cancel the effects of the convolutive channel. This can be done using blind or semi-blind techniques that either directly search an equaliser or first compute a channel model that is then used to compute an equaliser. The second category, which is the one that will be investigated here, aims at maximising the a-posteriori probability of the transmitted sequence exploiting the finite alphabet properties of the transmitted signals. These techniques are known as Maximum Likelihood (ML) techniques. Their main advantage is their performance; when optimal ML decoding is done, the lowest achievable error floor for the estimation of the transmitted sequence is obtained. Their major impediment though is their computational complexity since most of them use exhaustive or trellis searching techniques in order to estimate the transmitted sequence. Sub-optimal techniques aim at approaching the ML detector with a limited computational complexity. The problem of ML estimation of data transmitted over an unknown channel and some recently proposed techniques are briefly presented in section 1.2.

1.1 Block Transmission Techniques

Many different Block Transmission Techniques have been proposed in the last years, and we will briefly describe some of these here discussing their compared advantages and drawbacks. Multi-carrier block transmission techniques based on a CP (i.e., Discrete Multi-Tone (DMT) techniques, also known as OFDM (Orthogonal Frequency Division Modulation)) perform an IFFT at the transmitter after which the CP is added; the receiver performs an FFT followed by a one-tap frequency domain equalisation. These techniques can be used in combination with carrier loading, which allows the transmitter to optimize the Power Spectral Density (PSD) of the transmitted signal as a function of the channel [3], [1,

p 7]. If the channel is unknown to the transmitter, however, this loading is not applicable and the performance of the system becomes very sensitive to deep channel fades in the frequency-domain. A solution that is often used to reduce the sensitivity to channel fades in the frequency domain coonsists of encoding the data across the tones with the drawback of an increased complexity yielded by the encoding and decoding operations. Another drawback of classical DMT is the occurrence of large peaks in the transmitted signal, known as the Peak to Average Power Ratio (PAPR) problem [4]. In this paper, we consider the case where the channel is unknown to the transmitter, making carrier loading unapplicable. We therefore focus on single-carrier block transmission techniques based on a CP (i.e., CP-Only techniques) [5, pp 103-104], [2, p 36], where the transmitter simply adds a CP to every block of data symbols. This solves the PAPR problem since the transmitted data sequences now show finite alphabet properties, while keeping the advantage of computationally cheap ISI mitigation. CP-Only can be seen as a classical OFDM transmission scheme where the data are encoded across the tones by means of an FFT. The sensitivity to channel fades in the frequency-domain is therefore reduced by this technique but channel-irrespective symbol recovery is not guaranteed (if there is a zero on the FFT grid of the channel). A variant of CP-Only, which has recently been proposed, is known as Known Symbol Padding Only (KSP-Only) [6] [7] [8]. A sequence of known symbols is padded to every block of transmitted data symbols, which makes the data pseudo-cyclic and thus allows the same equalisation scheme as in the classical CP-Only context. The known symbols can be further exploited as a training sequence for channel estimation while the percentage of redundant symbols remains roughly the same. Zero-Padding (ZP) techniques [2], which were also proposed recently can be seen as a special case of the KSP-Only techniques where the known symbols are all set to zero. The advantage of these KSP-Only techniques is twofold: they allow new channel estimation techniques exploiting the knowledge of the padded symbols (see e.e. [7]) and they also allow channel-irrespective symbol recovery.

1.2 Blind Maximum Likelihood Sequence Estimation Techniques

The problem of blindly estimating a sequence that was transmitted over a convolutive channel $\mathbf{h}$ of order L is given as follows. Let $\mathbf{x}$ be the transmitted sequence and $\mathbf{y}$ the received sequence. The goal is to find among all possible combinations of $\mathbf{x}$ and $\mathbf{h}$ the pair $\hat{\mathbf{x}}$ and $\hat{\mathbf{h}}$ that yields the highest probability of observing the received sequence $\mathbf{y}$:

$$\max_{\hat{\mathbf{x}},\hat{\mathbf{h}}} p(\mathbf{y}|\hat{\mathbf{x}},\hat{\mathbf{h}}) \tag{1}$$

x This maximization problem has to be solved under the constraint that the elements of the transmitted sequence belong to a finite alphabet while the channel coefficients are unconstrained. When the noise at the receiver is AWGN (Additive White Gaussian Noise), which is a common assumption, this is equivalent to a joint Least Squares (LS) minimisation problem:

$$\min_{\hat{\mathbf{x}},\hat{\mathbf{h}}} \|\mathbf{y} - \hat{\mathbf{x}} * \hat{\mathbf{h}}\|^2 \tag{2}$$

where $*$ denotes the linear convolution operator.

Since this problem is separable in its two variables $\hat{\mathbf{x}}$ and $\hat{\mathbf{h}}$, it can be solved in two steps. In a first step we compute the LS channel estimate for every possible transmitted sequence $\mathbf{x}$. In a second step, the value of the cost function (2) is computed for all possible candidates. The joint ML sequence and channel estimator chooses the pair $\mathbf{x}$, $\mathbf{h}$ with the smallest cost function.

This exhaustive search procedure is optimal but computationally prohibitive, its cost being exponential in the length of the transmitted sequence. Cheaper optimal decoding techniques have been proposed. A first family of solutions proposes iterative algorithms that rely upon an initial channel estimate. They find the ML transmitted sequence under the hypothesis that the real channel is equal to the available channel estimate. The channel is then re-estimated based on the decoded sequence. These algorithms proceed iteratively until convergence. The different proposed solutions differ in the way they estimate the ML sequence. The K-means Algorithm [9] uses the Viterbi algorithm to decode the sequence, while the EM algorithm [10] computes the likelihoods of all possible data sequences conditioned on the available channel estimate.

In [11], non-iterative blind methods are proposed. A first proposed optimal technique performs decoding using Viterbi-related trellis search techniques on several hypothesized trellises. The computational cost of this optimal technique remains prohibitive enough to discount its use in practical systems. A sub-optimal blind trellis searched technique whose complexity is comparable to adaptive Viterbi decoding with exact channel knowledge was proposed in [11] as well. Here, instead of retaining a single survivor into each state as is done in classical Viterbi, the M best paths into a state are retained. The channel model is then updated for each of the retained paths as in adaptive Viterbi. Each survivor path is thus coupled to a channel estimate. The channel model associated to the path with the lowest metric converges to the real channel model whatever the initial channel estimate. Here again the computational cost, which is exponential in the channel order, remains problematic especially in broadband communications where the channels have long impulse responses.

When no inter-symbol interference (ISI) is present in the transmission system, the problem of blindly estimating the transmitted sequence becomes much less complex thanks to the absence of memory in the channel. In [12], iterative methods exploiting finite alphabet properties of the transmitted signals have been proposed in a MIMO context for such systems. One of these proposed methods, known as Iterative Least Squares with Projection (ILSP), approaches Maximum Likelihood (ML) detection of the transmitted sequence whilst maintaining the complexity at an acceptably low level.

1.3 Proposed Method

In this paper, we aim at performing blind MLSE (Maximum Likelihood Sequence Estimation) or sub-optimal blind sequence estimation in the context of multipath channels whilst maintaining the computational complexity at affordable levels. When block transmission techniques using a CP are used, the ISI is totally cancelled, creating flat-fading

channels in the frequency-domain. This property allows us to use simpler MLSE methods that are designed for flat-fading channels, but these methods have to be applied in the frequency-domain instead of the time-domain as it is usually done.

We use CP-Only transmission techniques rather than classical DMT in order to reduce the sensitivity of the system to the frequency-domain fades of the unknown channel. To approach MLSE, we propose a novel ILSP algorithm that exploits both the cyclic prefixed structure of the transmitted signals and their finite alphabet properties. The idea of combining block transmission techniques with MLSE techniques designed for flat fading-channels is not new. It has already been used in [13], where a modified ILSP procedure for OFDM systems is proposed. This method reconstructs the frequency-domain channel matrix from a time-domain estimate of the channel with a limited number of taps. In the algorithm proposed here, this tap constraint on the frequency-domain channel matrix is relaxed. Experimental results show that this difference is crucial for the performance of the algorithm, allowing our method to perform much better in terms of bit error rates (BER) than the existing one. The reasons for these improved perfomances are discussed in more details in section 6.1.

Experimental results show that the starting point of the iterations is critical for the convergence speed of the ILSP algorithm. In order to reduce the number of iterations before convergence, we need a sufficiently accurate initial channel estimate. We rely therefore on KSP-Only modulation, which is a special case of CP-Only. The knowledge of the padded symbols allows us to find an initial channel estimate in a cheap way using a new stochastic method that is described in this paper. This initial channel estimate is the only reason for using KSP-Only instead of CP-Only; the ILSP procedure can be used in a classical CP-Only framework provided a sufficiently accurate initial channel estimate is made available by other means.

We end up with a semi-blind iterative sub-optimal Maximum Likelihood Sequence Estimation (MLSE) method in the context of Cyclic Prefix Only (CP-Only) block transmission over unknown stationary multipath channels, which is characterised by both a low computational complexity and a high BER performance.

The rest of the paper is structured as follows. In section 2, we present the data model that will be used in this paper. For the sake of simplicity, we directly present the model for KSP-Only; the extension to CP-Only is straightforward. In section 3, we present the ML detector in the CP-Only context. In section 4, we propose an ILSP algorithm suited for this context whose performance approaches the ML detector. In section 5, we propose a stochastic method that exploits the known symbols and provides a sufficiently accurate channel estimate for initialising the ILSP algorithm. In section 6, we compare the proposed method with existing ones, namely [13] and [11]. In section 7, we present more detailed simulation results, and we finally draw some conclusions in section 8.

2 Data Model

Scalars are represented with small letters, vectors with boldface small letters and matrices with boldface capital letters. We are working in a block transmission context where successive blocks of data are transmitted after one another forming a long sequence referred to as a burst. A superscript indicates the block index within the burst; a subscript indicate the time index. Hence, $\mathbf{x}^i$ is the sequence constituting the ith transmitted block of the burst, whereas x^i_j is the jth element of the ith data block.

The data model we present describes a Known Symbol Padding system with ISI. Let us first define a transmission channel of order L:

$$\mathbf{h} = [h_0 \dots h_L]^T, \tag{3}$$

where $(.)^T$ denotes transpose, and define a training sequence $\mathbf{t}$ of length T:

$$\mathbf{t} = [t_1 \dots t_T], \tag{4}$$

with the condition $T \geqslant L$.

Let P be the transmitted block size. The sequence of transmitted data symbols is organised in blocks of length $B = P - T$. The nth transmitted data block is defined as:

$$\mathbf{s}^n = [s_{nB+1} \dots s_{(n+1)B}], \tag{5}$$

where s_i is the ith transmitted data symbol.

The transmitter transmits a burst of $\mathcal{V}$ blocks where every block is padded with the sequence $\mathbf{t}$. The total transmitted sequence is thus $\mathbf{x} = [\ \mathbf{t}\ \ \mathbf{s}^1\ \ \mathbf{t}\ \ \dots\ \ \mathbf{t}\ \ \mathbf{s}^{\mathcal{V}}\ \ \mathbf{t}\ ]$. We consider the channel to be constant across the transmission of an entire burst. Define the matrix of transmitted data, omitting the first transmitted training sequence as

$$\mathbf{X} = \begin{bmatrix} \mathbf{s}^{1T} & \cdots & \mathbf{s}^{\mathcal{V}T} \\ \mathbf{t}^T & \cdots & \mathbf{t}^T \end{bmatrix}. \tag{6}$$

The received sequence is the convolution of the transmitted sequence with the channel:

$$y_i = h_0 x_i + h_1 x_{i-1} + \dots + h_L x_{i-L} + n_i \tag{7}$$

where x_i is the ith element of $\mathbf{x}$ and n_i is the Additive White Gaussian Noise (AWGN) at the receiver. The received symbols are organised in $\mathcal{V}$ blocks of length P: $\mathbf{y}^n = [y_{T+(n-1)P+1} \ \cdots\ y_{T+nP}]$ is defined as the nth received block. These blocks are organised in a $(P \times \mathcal{V})$ matrix:

$$\mathbf{Y} = [\ \mathbf{y}^{1T}\ \ \cdots\ \ \mathbf{y}^{\mathcal{V}T}\ ]. \tag{8}$$

Define the $(P \times P)$ circulant channel matrix

$$\mathbf{H}_{circ} = \begin{bmatrix} h_0 & 0 & \dots & 0 & h_L & \dots & h_2 & h_1 \\ h_1 & h_0 & 0 & \dots & 0 & h_L & \dots & h_2 \\ & \ddots & \ddots & \ddots & & \ddots & \ddots & \\ h_L & \dots & h_0 & 0 & \dots & & 0 & 0 \\ 0 & h_L & \dots & h_0 & 0 & \dots & & 0 \\ 0 & 0 & \ddots & & \ddots & \ddots & & \\ \vdots & & \ddots & \ddots & & \ddots & & \vdots \\ 0 & \dots & & 0 & h_L & \dots & h_0 & 0 \\ 0 & & \dots & & 0 & h_L & \dots & h_0 \end{bmatrix}. \tag{9}$$

Exploiting the cyclic structure of the transmitted symbols, one obtains:

$$\mathbf{Y} = \mathbf{H}_{circ}\mathbf{X} + \mathbf{N}, \tag{10}$$

where $\mathbf{N}$ is the AWGN matrix. Define the frequency-domain equivalent of the channel:

$$\mathbf{h}_f = \mathcal{F}_P \begin{bmatrix} \mathbf{h} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{11}$$

where $\mathcal{F}_P$ is the P-point FFT matrix. The circulant channel matrix $\mathbf{H}_{circ}$ is diagonalised by means of an FFT and an IFFT. This allows us to describe the transmission scheme in a simplified way using FFT and IFFT operations [2]:

$$\mathbf{Y} = \mathcal{I}_P \mathbf{H}_f \mathcal{F}_P \mathbf{X} + \mathbf{N}, \tag{12}$$

where $\mathcal{I}_P$ is the P-point IFFT matrix and $\mathbf{H}_f = diag(\mathbf{h}_f)$ is a diagonal matrix with the frequency-domain description of the channel on the main diagonal.

Note that the ratio of training symbols over the total number of transmitted symbols for this KSP-Only context is equal to L/P when $T = L$. In a classical CP-Only context, the percentage of redundant symbols is $(L/P)(1 + L/P)^{-1}$ which tends to L/P as the number of sub-carrier increases. This shows that for a sufficiently large number of sub-carriers P, CP-Only and KSP-Only are equivalent as far as the percentage of redundant symbols is concerned. Hence we do note waste extra bandwidth using KSP-Only instead of CP-Only [8].

3 Maximum Likelihood Sequence Estimation

The data model shows us that the received signals can be modeled as deterministic sequences corrupted by AWGN. The log likelihood function of the transmitted data is

$$\mathcal{L} = -\alpha - \beta \ln(\sigma^2) - \frac{1}{\sigma^2} \sum_{n=1}^{\mathcal{V}} \|\mathbf{y}^{nT} - \mathcal{I}_P \mathbf{H}_f \mathcal{F}_P \mathbf{x}^{nT}\|^2, \tag{13}$$

where α and β are constants, σ^2 is the noise variance and $\mathbf{x}^n$ is the nth column of $\mathbf{X}$. The ML detector maximises $\mathcal{L}$ with respect to the unknown parameters $\mathbf{H}_f$ and $\mathbf{x}^n$, $n = 1 \cdots \mathcal{V}$ under the following constraint set:

Constraint Set 1 (CS1):

- $\mathbf{x}$ belongs to a finite alphabet,
- $\mathbf{H}_f = \mathcal{F}_P \mathbf{H}_{circ} \mathcal{I}_P$ where $\mathbf{H}_{circ}$ is a circulant matrix built from any unconstrained L^{th} order channel $\mathbf{h}$ in accordance with (9).

This is equivalent to the following minimisation problem:

Cost Function 1 (CF1):

$$\min_{\mathbf{H}_f, \mathbf{X}} \|\mathbf{Y} - \mathcal{I}_P \mathbf{H}_f \mathcal{F}_P \mathbf{X}\|^2, \tag{14}$$

under the same constraints. Multiplying both terms by $\mathcal{F}_P$ and using the notations $\mathbf{X}_f = \mathcal{F}_P \mathbf{X}$ and $\mathbf{Y}_f = \mathcal{F}_P \mathbf{Y}$ yields the following alternative formulation of the Maximum Likelihood problem:

Cost Function 2 (CF2):

$$\min_{\mathbf{H}_f, \mathbf{X}_f} \|\mathbf{Y}_f - \mathbf{H}_f \mathbf{X}_f\|^2, \tag{15}$$

with the same constraint on $\mathbf{H}_f$, and $\mathbf{X}_f$ being constrained by the finite alphabet structure of $\mathbf{X}$. This problem can be solved using exhaustive search procedures as described in section 1.2, the cost of this method being totally prohibitive.

We know that the constraint on $\mathbf{H}_f$ gives it the structure of a diagonal matrix with $L + 1$ degrees of freedom. Let us investigate what happens if we give $\mathbf{H}_f$ P degrees of freedom, that is if we allow the elements of its diagonal to vary freely[1]. The constraint set becomes:

Constraint Set 2 (CS2):

- $\mathbf{x}$ belongs to a finite alphabet,
- $\mathbf{H}_f$ is a diagonal matrix with unconstrained elements on the diagonal.

The MLSE problem in our context is thus equivalent to:

[1]This choice of allowing full freedom to the elements of the diagonal of $\mathbf{H}_f$ is further discussed in section 6.1.

- Solving a minimisation problem equivalently expressed by CF1 or CF2
- This minimisation is either performed under the constraint set CS1 or under the constraint set CS2

In the Latter, we will focus on the solution of this problem under the constraint set CS2. Since the problem is separable in its two variables, the optimisation can be carried out in two steps [12], using the two equivalent formulations (14) and (15). We first solve (15) with respect to $\mathbf{H}_f$ using (15). Exploiting the new condition on $\mathbf{H}_f$ (CS2), we force it to have the desired diagonal structure by splitting the problem into P independent sub-problems:

$$\min_{\mathbf{H}_f(j,j)} \|\mathbf{Y}_f(j,:) - \mathbf{H}_f(j,j)\mathbf{X}_f(j,:)\|^2, \quad j = 1 \cdots P, \tag{16}$$

the solution of which is

$$\hat{\mathbf{H}}_f(j,j) = \mathbf{Y}_f(j,:)\mathbf{X}_f(j,:)^H \left(\mathbf{X}_f(j,:)\mathbf{X}_f(j,:)^H\right)^{-1}, \tag{17}$$

where $(.)^H$ denotes complex conjugate transpose and $\mathbf{A}(j,:)$ is the jth row of a matrix $\mathbf{A}$. Note that this solution would not be achievable under the initial constraint on $\mathbf{H}_f$ (CS1), all the sub-problems being then linked to one another. We thus see that using CS2 allows us to reduce the complexity of the problem. Inserting these results that only depend on $\mathbf{X}_f$ and $\mathbf{Y}_f$ into (14) allows us to find the optimum sequence by enumerating over all the possible values for the matrix $\mathbf{X}$, computing the value of the cost function, and choosing the matrix that minimises this cost function. This includes the computation of $\hat{\mathbf{H}}_f$ (and thus $\mathbf{X}_f$) for every possible combination of the inputs. The computational cost of this method, which is exponential in P, $\mathcal{V}$, and the alphabet size, again severely limits its practical interest.

In the noiseless case, the solution to this modified problem (CS2) and the initial ML problem (CS1) are the same. When the received sequence is noisy, we cannot guarantee that the derived solution will be equal to the ML solution under CS1. However, we still find the solution of a ML problem, but with a modified constraint set (CS2). We therefore propose to classify this method as a modified MLSE method.

4 Iterative Least Squares with Projection

In the previous section, we derived an expression for a modified ML detector. The associated minimisation problem is separable in its continuous and discrete variables. In this section, we apply an Iterative Least Squares with Projection algorithm inspired by [12], that uses this separation property to approach the solution of the modified ML problem by iteratively minimising the cost functions (14) and(15) for one variable and then for the other.

Assume we have an initial channel estimate $\hat{\mathbf{H}}_f$. We first minimise the cost function (14) with respect to $\mathbf{X}$ with a fixed $\hat{\mathbf{H}}_f$:

$$\min_{\mathbf{X}} \|\mathbf{Y} - \mathcal{I}_P \hat{\mathbf{H}}_f \mathcal{F}_P \mathbf{X}\|^2. \tag{18}$$

This is equivalent to computing a soft estimate of the transmitted symbols, implicitly performing a classical frequency-domain equalisation on the received symbols:

$$\hat{\mathbf{X}} = \mathcal{I}_P \hat{\mathbf{H}}_f^{-1} \mathcal{F}_P \mathbf{Y}. \tag{19}$$

Note that $\hat{\mathbf{H}}_f^{-1}$ is a diagonal matrix with $\hat{\mathbf{H}}_f^{-1}(j,j) = \hat{\mathbf{H}}_f(j,j)^{-1}$. This step is followed by a finite alphabet projection of the soft estimates: $\hat{\mathbf{X}} = FAP(\hat{\mathbf{X}})$, where FAP denotes the finite alphabet projection operation.

Note that in the special case of KSP-Only, the last T symbols of every column of $\mathbf{X}$ are known in advance and do not need to be re-estimated at each step. The problem of estimating the transmitted sequence then becomes a true Least-Squares problem rather than P equations in P unknowns:

$$\hat{\mathbf{S}} = \mathcal{I}_{BP} \hat{\mathbf{H}}_f^{-1} \mathcal{F}_P \mathbf{Y} \tag{20}$$

where $\mathbf{S}$ is the $B \times \mathcal{V}$ matrix of unknown symbols and $\mathcal{I}_{BP}$ is the $B \times P$ partial IFFT matrix made of the first B rows of $\mathcal{I}_P$

We then use this estimate to minimise the cost function (15) with respect to $\mathbf{H}_f$, where $\hat{\mathbf{X}}_f = \mathcal{F}_{\mathcal{P}} \hat{\mathbf{X}}$ is derived from the previous step:

$$\min_{\mathbf{H}_f} \|\mathbf{Y}_f - \mathbf{H}_f \hat{\mathbf{X}}_f\|^2. \tag{21}$$

Again we split this into P parallel independent problems to force $\mathbf{H}_f$ to have a diagonal structure:

$$\min_{\mathbf{H}_f(j,j)} \|\mathbf{Y}_f(j,:) - \mathbf{H}_f(j,j) \hat{\mathbf{X}}_f(j,:)\|^2, \quad j = 1 \cdots P, \tag{22}$$

the solution of which is

$$\hat{\mathbf{H}}_f(j,j) = \mathbf{Y}_f(j,:) \hat{\mathbf{X}}_f(j,:)^H \left(\hat{\mathbf{X}}_f(j,:) \hat{\mathbf{X}}_f(j,:)^H \right)^{-1}. \tag{23}$$

Since $\mathbf{H}_f^{-1}$ is used in the iterations to estimate $\mathbf{X}$, we can alternatively estimate it directly by solving the following minimisation problem:

$$\min_{\mathbf{H}_f^{-1}(j,j)} \|\mathbf{H}_f^{-1}(j,j) \mathbf{Y}_f(j,:) - \hat{\mathbf{X}}_f(j,:)\|^2, \quad j = 1 \cdots P, \tag{24}$$

this is a simple least squares problem whose solution is:

$$\hat{\mathbf{H}}_f^{-1}(j,j) = \mathbf{X}_f(j,:) \hat{\mathbf{Y}}_f(j,:)^H \left(\hat{\mathbf{Y}}_f(j,:) \hat{\mathbf{Y}}_f(j,:)^H \right)^{-1}. \tag{25}$$

Note that this modification implies that we perform an implicit MMSE equaliser estimate rather than a direct channel estimate, which also has the benefit of avoiding noise enhancement when the next $\mathbf{X}$ estimate is performed. This $\mathbf{H}_f^{-1}$ is used to compute a new $\hat{\mathbf{X}}$ and we proceed iteratively. The iterations are stopped when two consecutive finite alphabet

estimates of the transmitted sequences are identical. The convergence point is a minimum of the cost function associated to the modified ML problem. Note that a fairly good initial channel estimate is needed to avoid a too large number of iterations and/or the convergence of the algorithm to local minima. In the following section, we therefore provide a cheap channel estimation technique for initialising the algorithm, which allows this iterative procedure to converge to minima that are close or equal to the global minimum containing the ML sequence.

An important advantage of this ILSP method is its low computational complexity. At every iteration, one symbol estimation step and one channel estimation step are performed. The complexity of the symbol estimation step is $\mathcal{O}(\mathcal{V}P\log(P))$. The complexity of the channel estimation step is $\mathcal{O}(\mathcal{V}P\log(P))$ including the computation of $\mathbf{Y}_f$ from $\mathbf{Y}$. The total complexity of the proposed method is thus $\mathcal{O}(\log(P))$ per symbol as well. Unlike existing methods, the complexity of the proposed one is independant from the channel order and grows only linearly with the length of the transmitted sequence.

The use of an Iterative Least Squares with Enumerations algorithm [12] is an alternative solution to ILSP that always converges. However, in our context, this ILSE procedure has a complexity $\mathcal{O}\left(2^P\right)$ which makes it unaffordable for most practical systems where the number of sub-carriers is large. For a practical system with 64 carriers using BPSK modulation, there are in total $1.84\ 10^{19}$ different combinations that have to be enumerated. This approach will therefore not be developed here.

5 Initial Channel Estimate

Define $\mathbf{y}_T^n$ as the $(T+L)$ -long vector of received samples capturing all the energy of the nth transmitted training sequence:

$$\mathbf{y}_T^n = \left[y_{(n-1)P+1} \cdots y(n-1)P+T+L\right], \tag{26}$$

where $n = 1 \cdots \mathcal{V}+1$. Note that there are $\mathcal{V}+1$ such vectors. Define $\mathbf{T}^n$ as the $(T+L) \times (L+1)$ matrix of transmitted symbols that contribute to $\mathbf{y}_T^n$. In the case where $L = T$, we have:

$$\mathbf{T}^n = \begin{bmatrix} t_1 & s_B^{n-1} & \cdots & \cdots & s_{B-L+1}^{n-1} \\ t_2 & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ t_T & \cdots & \cdots & t_1 & s_B^{n-1} \\ s_1^n & t_T & \cdots & \cdots & t_1 \\ \vdots & s_1^n & t_T & \cdots & t_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ s_L^n & s_{L-1}^n & \cdots & s_1^n & t_T \end{bmatrix}. \tag{27}$$

Using (7) we can express $\mathbf{y}_T^n$ as :

$$\mathbf{y}_T^{n\ T} = \mathbf{T}^n\mathbf{h} + \mathbf{n}^{n\ T}, \tag{28}$$

where $\mathbf{n}^n$ is the AWGN vector.

This expression contains both deterministic and random variables. If one looks at the expected value of ${\mathbf{y}_T^n}^T$, assuming $\mathbf{E}\{s_j^i\} = \mathbf{E}\{n_i\} = 0 \quad \forall i, j$, which is the case if the noise is zero-mean and if the data symbols are equiprobable, (28) reads.

$$\mathbf{E}\{{\mathbf{y}_T^n}^T\} = \mathcal{T}\mathbf{h}, \tag{29}$$

where

$$\mathcal{T} = \mathbf{E}\{\mathbf{T}^n\} = \begin{bmatrix} t_1 & 0 & \cdots & 0 & 0 \\ t_2 & t_1 & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ t_T & \cdots & \cdots & t_1 & 0 \\ 0 & t_T & \cdots & \cdots & t_1 \\ \vdots & \ddots & \ddots & & t_2 \\ \vdots & & \ddots & t_T & \vdots \\ 0 & 0 & \cdots & 0 & t_T \end{bmatrix}. \tag{30}$$

Define the average value of $\mathbf{y}_T^n$ as:

$$\overline{\mathbf{y}_T} = \frac{1}{\mathcal{V}+1} \sum_{i=1}^{\mathcal{V}+1} \mathbf{y}_T^i. \tag{31}$$

It is clear that $\lim_{\mathcal{V}\to\infty} \overline{\mathbf{y}_T} = \mathbf{E}\{\mathbf{y}_T^n\}$, and we will therefore use it as an approximation of $\mathbf{E}\{\mathbf{y}_T^n\}$ even with a finite $\mathcal{V}$. We can thus write

$$\overline{\mathbf{y}_T}^T \simeq \mathcal{T} \quad \mathbf{h}, \tag{32}$$

from which we derive a channel estimate based on the knowledge of the training symbols:

$$\hat{\mathbf{h}} = \left(\mathcal{T}^H\mathcal{T}\right)^{-1} \mathcal{T}^H \overline{\mathbf{y}_T}^T. \tag{33}$$

This channel estimate is presented here as a method for initialising $\mathbf{H}_f$ in the ILSP procedure. Experimental results show that this initial estimate yields good performance of the algorithm, dramatically reducing the number of iterations needed to reach convergence. While any other initial channel estimate can be used as a starting point for the iterations, the advantage of this one is its very low computational complexity.

6 Comparison with Existing Methods

6.1 ILSP for OFDM Systems

In [13], an ILSP procedure for joint blind estimation of channel and data symbols in OFDM systems is proposed. Like in our method, the data are organised in blocks of B symbols

padded with T known symbols with the condition $T \geqslant L$. The known symbols are set to zero in this case. An IFFT is then computed on the data blocks before transmission and no CP is added. This transmission scheme makes it possible to avoid Inter Block Interference by removing T samples at the receiver. The system can be described with a matrix formulation similar to (10) where $\mathbf{H}_{circ}$ is replaced by some appropriate matrix $\mathbf{A}$. This channel matrix $\mathbf{A}$ is directly derived from the time-domain model of the channel $\mathbf{h}$. The proposed detection algorithm starts with an arbitrary estimate of the channel matrix $\mathbf{A}$. Based on this channel estimate, a soft estimate of the transmitted sequence is computed, followed by a hard decision. Based on this sequence estimate, the channel matrix $\mathbf{A}$ is re-estimated. One could decide tostart the next iteration with this new channel matrix estimate. However, the fact that the channel matrix is derived from an Lth order channel model can be seen as a constraint on the structure of $\mathbf{A}$. To fulfill this constraint, [13] seeks the Lth order channel model that is the best LS fit to the estimated channel matrix. A new channel matrix is then computed based on this estimate of $\mathbf{h}$. The iterations proceed with this last channel matrix estimate.

Let us now compare this method with the one we propose. We first slightly modify our transmission scheme by setting the padded sequence to zeros: $\mathbf{t} = [0 \cdots 0]$. This modification makes our transmission scheme the exact CP-Only counterpart of the proposed OFDM scheme. This modification makes the initial channel estimate procedure described in section 5 unapplicable. As a consequence, our iterations are started with random channel estimates. If the random channel estimate is too far from the real channel, the method may converge to irrelevant local minima. When a bad local minimum is reached, we thus restart the iterations with another random channel estimate untill a good convergence is achieved. As a consequence, the total number of iterations needed to converge is largely increased.

As mentioned earlier, the most important difference between this OFDM method and the one we propose resides in the constraints that are put on the channel matrix. In the OFDM method, the $\mathbf{A}$ channel matrix is forced to be derived from an Lth order channel model, allowing only $L+1$ degrees of freedom to the channel matrix. If this constraint is relaxed, $\mathbf{A}$ has B^2 degrees of freedom and the convergence properties of the iterative algorithm become erratic because the problem is under-constrained. In the proposed method, we could have decided to put an equivalent constraint on $\mathbf{H}_f$ by forcing it to be the frequency-domain description of a general L^{th} order channel, allowing $L+1$ degrees of freedom to the channel matrix. However, in order to reduce the computational complexity of the algorithm, we decided to relax this constraint on $\mathbf{H}_f$, replacing it by the constraint of having a diagonal structure. This increases the number of degrees of freedom on $\mathbf{H}_f$ to P instead of the $L+1$ offered by the previous constraint. Note that this is still much less than the P^2 degrees of freedom we would have had if $\mathbf{H}_f$ had been totally unconstrained.

Experimental results that are presented further show that relaxing this L^{th} order time-domain equivalent channel constraint on the channel matrix yields spectacular improvements in terms of BER. We give next an intuitive explanation for these improved performances.

When blind MLSE is performed by exhaustive search (see section 1.2), the cost function (2), (14) or (15) is minimised for every possible input sequence. Iterative methods avoid this computationally prohibitive exhaustive search by only considering a few selected candidates among all the possible sequences. The general procedure is as follows: we start with an initial sequence candidate; the LS channel model associated with this sequence candidate is then computed (this LS channel model is also the ML channel model). The next step consists in selecting a new sequence candidate under the hypothesis that the provided channel model is exact. Finally, the best proposed sequence is selected (the one with the lowest cost function). The technique that is used to select new candidate sequences is crucial for the performance of the iterative methods. Only the candidates selected by this technique will be considered for estimating the transmitted sequence. For instance, the K-means Algorithm [9] performs ML detection using the Viterbi algorithm under the hypothesis that the provided channel model is exact in order to select new candidates. The selection procedure used in the existing ILSP method for OFDM systems implicitly performs a zero-forcing (ZF) equalisation of the estimated channel. The best possible candidate is obtained performing such a ZF equalisation with the real channel model. However, ZF equalisation procedures are known to have a poor performance for ill conditioned channels and low SNRs. This poor selection method severely limits the performance that could possibly be achieved by such methods. The proposed method behaves a bit differently. As we let the iterations run, $\mathbf{H}_f$ converges to the frequency-domain diagonal matrix that will best project the received sequence onto the known finite alphabet of the transmitted sequence. The $\mathbf{H}_f$ that is computed at each step is thus a sort of frequency-domain best fit between the hypothesised input and the received noisy channel output, instead of the ML channel model for the corresponding input sequence that is usually computed. The noise effects are implicitly included in the computed $\mathbf{H}_f$ since its estimate is based on noisy data observations. We end up with a frequency-domain MMSE (Minimum Mean Squared Error) estimator that generates much better candidate sequences than the existing ZF sequence estimator. Note that it would be impossible to find this MMSE estimator if $\mathbf{H}_f$ was constrained according to CS1. Using CS2 rather than CS1 thus not only allows to reduce the complexity of the method, it also allows to find better candidate sequences qnd thereby increases the performance of the algorithm

In order to illustrate this discussion, we performed simulations comparing the proposed method with the existing OFDM one. As mentioned earlier, the padded symbols were set to zero and the iterations were started with random initial channel estimates for our method, the initial channel estimate for the OFDM method being set to the identity matrix as recommended by the authors.

In a first experiment, we compare the OFDM method with a modified version of our algorithm that uses the initial constraint set CS1 rather than the modified one. We thus perform a LS estimation of the time-domain channel model $\mathbf{h}$ at each iteration. We then perform an implicit ZF equalisation on the received symbols in order to find the next candidate input sequence. This CS1 method performs slightly better in terms of BER than the existing ILSP method for OFDM. The only reason for this is that the proposed method is a CP-Only method, it is thus less sensitive to the frequency-selectivity of the

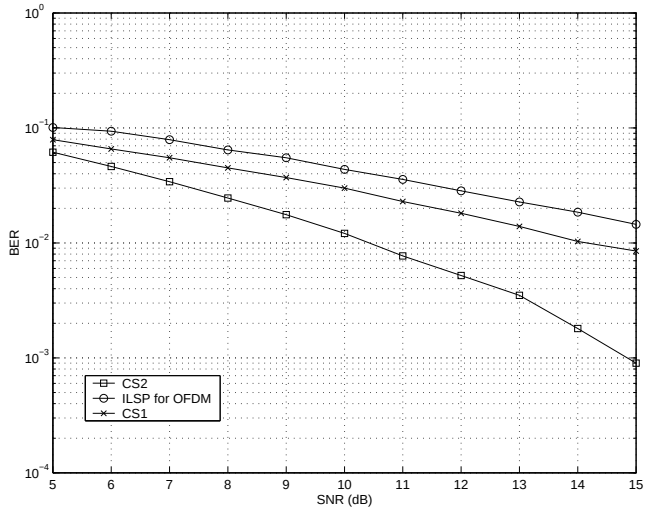

Figure 1: BER vs. SNR for the OFDM ILSP method (upper line), our method under the traditional constraint set (CS1, middle line) and our method with the modified constraint set (CS2, lower line).

transmission channel than the existing OFDM method (see discussion in section 1.1).

In a second experiment, we relax the constraint on $\mathbf{H}_f$, only forcing it to have a diagonal structure,applying the second set of constraint CS2 as we did in the rest of the paper. This results in much better BER performance than in the previous experiment. Since the problem is now under-constrained, the convergence properties of the algorithm get much worse than what they are in the previous case, requiring many more iterations and initial channel guesses to reach the optimum. Note that if Known Symbols were to be used like we initially proposed, the convergence would not be problematic anymore since the initial starting point would then be accurate enough to achieve fast and cheap convergence.

The results are presented in Fig. 1, where the achieved BERs are shown for different SNRs. The experiments were performed on a set of 200 different realisations of a first order Rayleigh fading channel using BPSK signaling. We ran the three proposed methods on the same data sequences and noise realisations. The experiments were performed with a block length $\mathcal{V} = 100$ and 64 sub-carriers. The upper line is the BER obtained with the existing OFDM method. The second curve is obtained with our modified method. The lower curve is obtained with our method when $\mathbf{H}_f$ is only constrained to have a diagonal matrix structure.

6.2 Trellis Searched Blind Equalisation

In the previous section, we have seen how relaxing the constraint on $\mathbf{H}_f$ allows our method to reach high performance. In this section, we investigate how the resulting iterative method compares with existing sub-optimal blind MLSE methods. We chose to compare

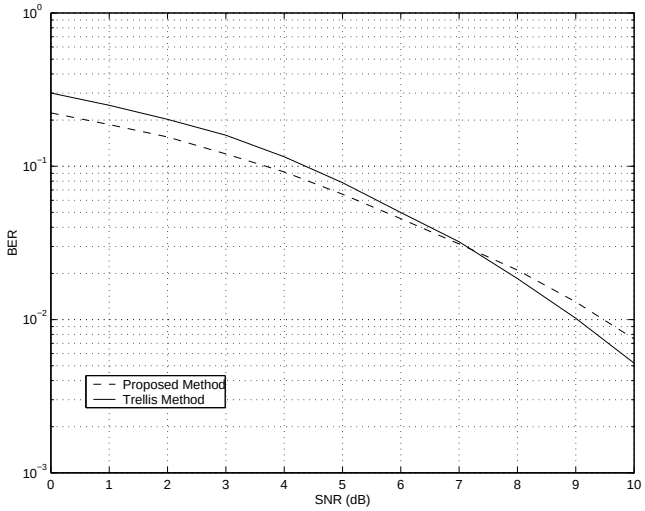

Figure 2: BER vs. SNR for the trellis searched method (full line) and our iterative method (dashed line).

our method with a trellis searched blind equalisation method that was first proposed in [11], which approaches blind ML decoding whilst keeping the computational complexity at a reasonable level. This method was already presented in section 1.2. The KSP context we are working in allows us to make a few changes, lowering both its computational complexity and implementation cost. In [11], the M best paths into a state are retained instead of one as done in the classical Viterbi algorithm. The main reason for this is the absence of a reliable initial channel estimate. In our context though, a quite accurate initial channel estimate is available exploiting the known symbols of the transmitted sequence (see section 5). This allows us to retain only one surviving path into each state. Another feature of interest which is offered by the known padded sequence is that it forces the trellis initialisation and termination. At the beginning and the end of a transmitted block, the channel memory is filled with the training sequence that is padded to the blocks of data symbol. This allows us to perform a block per block detection without having to retain the trellis metrics and surviving path between different blocks.

The results are presented in Fig. 2, where the BERs achieved for different SNRs are shown. The experiments were performed on a set of 200 different realisations of a third order Rayleigh fading channel using BPSK signaling. We ran the different methods on the same data sequences and noise realisations. The experiments were performed with a blocklength $\mathcal{V} = 100$ and 64 sub-carriers. The two methods show similar performance in terms of BERs. However, simulations showed that the proposed method requires much less computing power than the trellis searched one. As the channel order increases, the complexity of our method remains the same whilst the complexity of the trellis method grows exponentially.

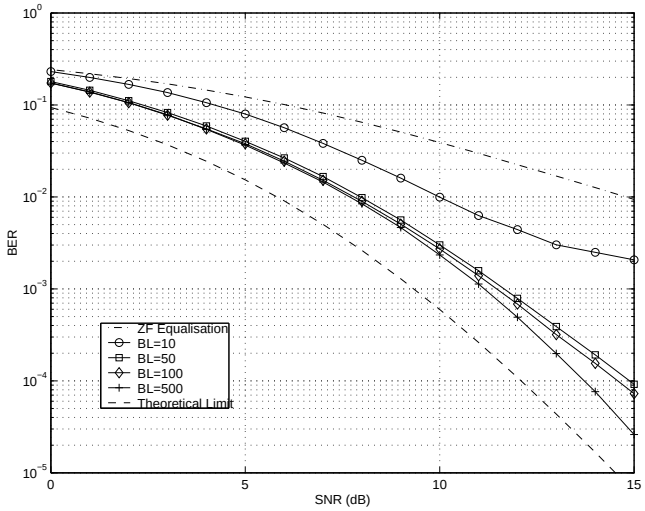

Figure 3: BER vs. SNR for different Burst Lengths (BLs).

7 Simulation Results

In this last section, we investigate the performance of the proposed method for higher order channels. We also study the impact of the burstlength $\mathcal{V}$ both in terms of performance and complexity. We present simulation results for Rayleigh-fading channels with $L = 5$, $T = 5$ and $P = 64$. The simulations were performed using BPSK modulation. We used a large number of randomly generated Rayleigh-fading channels, and approximately estimated 10^7 data symbols for each point of the graph.

Fig. 3 shows the BER for the proposed method as a function of the SNR. The different curves show the results for different values of the burstlength $\mathcal{V}$. The lower curve shows the theoretical BER floor for ML decoding of BPSK signals that are transmitted over an Lth order Rayleigh-fading SISO channels [14, p 955]. The upper curve shows the BER obtained with classical ZF equalisation in the frequency-domain assuming perfect channel knowledge.

The achieved BER approaches the MLSE limit as $\mathcal{V}$ increases. This is probably because the curve of the cost function gets smoother as the number of data symbols increases, which keeps the algorithm from getting stuck in local minima, or because the initial estimate becomes more accurate.

Fig. 4 shows the average number of iterations that were required for the algorithm to converge. This number is maintained at a quite low level thanks to the good initial channel estimator. This is especially true for high SNRs. It should be noted that large $\mathcal{V}$'s, which yield better BER performance, require more iterations, especially for low SNRs.

Fig. 5 shows the evolution of the BER performance as the number of iterations increase. The upper curve shows the performance after one iteration, the second curve shows the BER obtained after 2 iterations, and so on. The experiments are performed on 3rd order

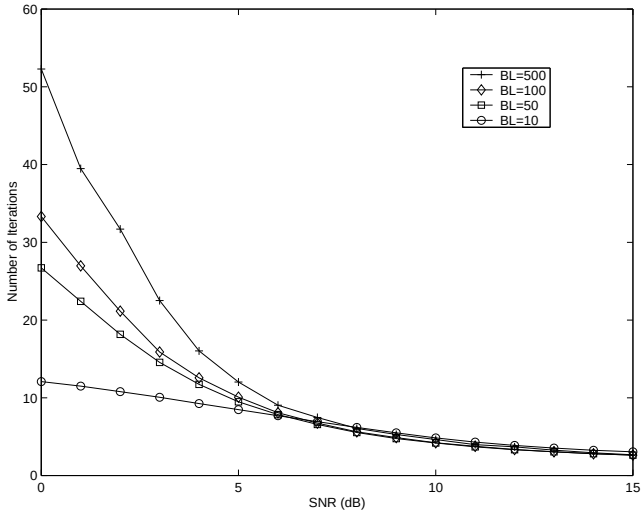

Figure 4: Num. of It. vs. SNR for different Burst Lengths (BLs).

channels with a block length $\mathcal{V}$ set to 100 We see from these experiments that most of the gain is achieved after 4 iteraions. Further iterations still improve the BER but only in a marginal way. This shows thus that a very small number of iterations is required in order to approach closely the optimal BER for this method.

8 Conclusions

The contribution of this paper is twofold. First, a new semi-blind iterative modified MLSE method for CP-Only transmission over stationary multipath channels has been proposed. This method achieves a very low BER, approaching the ML sequence estimate, while maintaining the computational complexity at very low levels.

Second, a new stochastic channel estimation method was proposed that can be used in a KSP-Only context. This method provides a sufficiently good channel estimate to initialise the iterative algorithm. Its main advantage is its low computational complexity.

References

[1] J.A.C. Bingham, "Multicarrier Modulation for Data Transmission: An Idea whose Time has Come," *IEEE Communications Magazine*, pp. 5–14, May 1990.

[2] Z. Wang and G.B. Giannakis, "Wireless Multicarrier Communications: Where Fourier meets Shannon," *IEEE Signal Processing Magazine*, pp. 29–48, May 2000.

[3] R.F.H. Fischer and J.B. Huber, "A New Loading Algorithm for Discrete Multitone Transmission," in *Proc. of Globecom*, London, England, Nov. 1996, pp. 724–728.

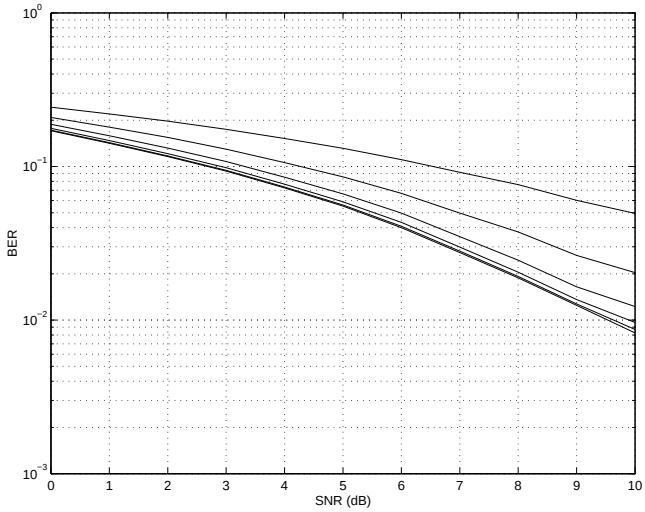

Figure 5: BER vs. SNR for increasing numbers of iterations.

[4] V. Tarokh and H. Jafarkhani, "On the Computation and Reduction of the Peak-to-Average Power Ration in Multicarrier Communications," *IEEE Transactions on Communications*, vol. 48, no. 1, pp. 37–44, Jan. 2000.

[5] H. Sari, G. Karam and I. Jeanclaude, "Transmission Techniques for Digital Terrestrial TV Broadcasting," *IEEE Communicatons Magazine*, pp. 100–109, Feb. 1995.

[6] R. Cendrillon, M. Moonen, "Efficient equalizers for single and multi-carrier environments with known symbol padding," in *Proc. of the Sixth International Symposium on Signal Processing and its Applications (ISSPA 2001)*, Kuala-Lumpur, Malaysia, Aug. 2001.

[7] G. Leus and M. Moonen, "Semi-Blind Channel Estimation for Block Transmission with Non-Zero Padding," in *Proc. of the Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, California, Nov. 4-7 2001.

[8] L. Deneire, B. Gyselinckx and M. Engels., "Training Sequence vs. Cyclic Prefix: A New Look on Single Carrier Communication," in *Proc. GLOBECOM*, San Fransisco, California, November/December 2000.

[9] B. H. Juang, L. R. Rabiner, "The Segmental K-Means Algorithm for Estimating Parameters of Hidden Markov Models," *IEEE transactions on Acoustics, Speech and Signal Processing*, vol. 38(9), pp. 1639–1641, Sept. 1990.

[10] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm," *Journal of Royal Statistical society B*, vol. 39, pp. 1–38, 1977.

[11] N. Seshadri, "Joint Data and Channel Estimation Using Blind Trellis Search Techniques," *IEEE Transaction on Communications*, vol. 42, no. 2/3/4, pp. 1000–1011, April 1994.

[12] S. Talwar, M. Viberg and A. Paulraj, "Blind Separation of Synchronous Co-Channel Digital Signals Using an Antenna Array - Part I: Algorithms," *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 1184–1197, May 1996.

[13] Y. Song, S. Roy an L. A. Akers, "Joint Blind Estimation of Channel and Data Symbols in OFDM," in *Proc. of the 51st Vehicular Technology Conference Proceedings*, Tokyo, Japan, 2000, vol. 1, pp. 46–50.

[14] John. G. Proakis, *Digital Communications*, Mc Graw-Hill International Edition, 4 edition, 2000.

List of Figures

1 BER vs. SNR for the OFDM ILSP method (upper line), our method under the traditional constraint set (CS1, middle line) and our method with the modified constraint set (CS2, lower line). 15
2 BER vs. SNR for the trellis searched method (full line) and our iterative method (dashed line). 16
3 BER vs. SNR for different Burst Lengths (BLs). 17
4 Num. of It. vs. SNR for different Burst Lengths (BLs). 18
5 BER vs. SNR for increasing numbers of iterations. 19