Generalized Training Based Channel Identification

Olivier Rousseaux[1], Geert Leus[1], Petre Stoica[2] and Marc Moonen[1]
[1] K.U.Leuven - ESAT
Kasteelpark Arenberg 10, 3001 Leuven, Belgium
Email: orousso@esat.kuleuven.ac.be
[2] Uppsala University - Department of Systems and Control
P.O. Box 337 SE-751 05, Uppsala Sweden
Email: ps@syscon.uu.se

***Abstract*— In this paper, we address the general problem of identifying convolutive channels when several training sequences are inserted in the transmitted data symbols stream. We analyze the general situation where the training sequences differ from each other. We consider quasi-static channels (i.e. the sampling period is several orders of magnitude below the coherence time of the channel). There are no requirements on the length of the training sequence and all the received symbols that contain contributions from the training symbols are used for the identification. We first propose an iterative method that quickly converges to the Maximum Likelihood (ML) channel estimate. We also derive a simple closed form expression that approximates the ML channel estimate.**

I. Introduction

A major impediment of broadband communication systems is that the sampling period can get smaller than the delay spread of the channel, especially in severe multipath scenarios. This results in ISI (Inter Symbol Interference), a phenomenon that needs to be combatted at the receiver in order to restore the transmitted information. This is usually done using serial or block equalization techniques. Channel State Information (CSI) is needed at the receiver in order to design the equalizer and combat the ISI in an efficient way.

The CSI is obtained through the use of channel identification algorithms. In this paper, we focus on the family of training-based channel identification algorithms. The problem of optimally identifying the channel at the receiver when training sequences are inserted in the transmitted signals has been widely discussed in the literature (see e. g. [1]). When the training sequences are long enough, some of the received symbols only contain contributions from the known training symbols (as opposed to the received symbols that contain contributions from both the unknown data symbols and the training symbols or from the data symbols only). The problem of performing Maximum Likelihood (ML) channel identification when only these received symbols are used is equivalent to a least squares problem. However, this classical approach is sub-optimal since not all the received symbols that contain contributions from the training symbols are used in the identification procedure.

In [2], we presented a ML channel identification method in a context where a fixed training sequence was repeatedly inserted between blocks of data symbols. The method we present here is a generalization of the existing one to the more general case where the training sequence is changed after each block of data. As will be discussed in the experimental results section, allowing the training sequences to change also yields improved accuracy of the channel estimate. The proposed method does not require a minimal length for the training sequences and performs ML channel estimation exploiting all the received symbols that contain contributions from the training sequences. We consider quasi-static channels (the channel stays constant during the transmission of several blocks of data). Note that there are several existing transmission schemes to which this new channel estimation method can be applied. Examples are for instance Known Symbol Padding (KSP) block transmission [3], [4] or Pilot Symbol Assisted Modulation (PSAM) [5].

The structure of the paper is as follows. In section II, we introduce a data model suited to our problem. In section III, we derive an expression for the likelihood function of the system and define the ML channel estimate. We propose an iterative solution to find this ML channel estimate in section IV and propose an approximate closed form expression of this ML channel estimate in section V. In section VI, we derive the Cramer Rao Bound in the presented context. We analyze the experimental performance of the proposed method in section VII and finally draw some conclusions in section VIII

Notation: We use upper (lower) case bold face letters to denote matrices (column vectors). $\mathbf{I}_N$ is the identity matrix of size $N \times N$ and $\mathbf{0}$ is the all-zero matrix. The operator $(.)^*$ denotes the complex conjugate, $(.)^T$ the transpose of a matrix and $(.)^H$ its complex conjugate transpose. Finally, tr$(.)$ denotes the trace of a matrix, and $|.|$ its determinant.

II. Data Model

We consider stationary Finite Impulse Response (FIR) convolutive channels of order L: $\mathbf{h} = [h[0], h[1], \cdots, h[L]]^T$. A sequence $x[n]$ of symbols is transmitted over the channel. The received sequence $y[n]$ is the linear convolution of the transmitted sequence with the channel impulse response:

$$y[n] = \sum_{i=0}^{L} h[i]x[n-i] + \eta[n],$$

where $\eta[n]$ is the Additive White Gaussian Noise (AWGN) at the receiver.

We consider a transmission scheme where training sequences are inserted in the transmitted symbols stream. The kth training sequence, $\mathbf{t}_k = [t_k[1], \ldots, t_k[n_t]]^T$, starts at position n_k, i.e. $[x[n_k], \ldots, x[n_k + n_t - 1]]^T = \mathbf{t}_k$. A total number of K such training sequences are observed to estimate the channel. The channel is assumed to stay constant during this time interval.

Define the vector $\mathbf{u}_k$ of received symbols that contain a contribution from the kth transmitted training sequence:

$$\mathbf{u}_k = [y[n_k], \ldots, y[n_k + n_t + L - 1]]^T .$$

The vector $\mathbf{u}_k$ contains a contribution from the training sequence $\mathbf{t}_k$ plus an additional term that collects the contributions from both the unknown surrounding data symbols and the noise. We can thus describe $\mathbf{u}_k$ as the sum of a deterministic and a stochastic term:

$$\mathbf{u}_k = \mathbf{T}_k \mathbf{h} + \boldsymbol{\epsilon}_k, \tag{1}$$

where $\mathbf{T}_k$ is an $(n_t + L) \times (L + 1)$ tall Toeplitz matrix with $[\mathbf{t}_k^T, 0, \ldots, 0]^T$ as first column and $[t_k[1], 0, \ldots, 0]$ as first row. $\mathbf{T}_k\mathbf{h}$ is the deterministic term; the stochastic term, $\boldsymbol{\epsilon}_k$, is described as follows:

$$\boldsymbol{\epsilon}_k = \underbrace{\left[\begin{array}{ccc|ccc} h_L & \cdots & h_1 & & & \mathbf{0} \\ & \ddots & \vdots & h_0 & & \\ & & h_L & \vdots & \ddots & \\ \mathbf{0} & & & h_{L-1} & \cdots & h_0 \end{array}\right]}_{\mathbf{H}_s (n_t + L) \times (2L)} \mathbf{s}_k + \boldsymbol{\eta}_k, \tag{2}$$

where $\mathbf{s}_k = [s_k[1], \cdots, s_k[2L]] = [x[n_k - L], \ldots, x[n_k - 1], x[n_k + n_t], \ldots, x[n_k + n_t + L]]^T$ is the vector of surrounding data symbols, and $\boldsymbol{\eta}_k$ is the corresponding AWGN term. Assuming that both the noise and the data are white and zero-mean ($E\{s_k[i]s_l[j]^*\} = E\{\eta[k]\eta[l]^*\} = 0, \quad \forall i, j, k, l$ when $k \neq l$ or $i \neq j$, and $E\{s_k[i]\} = E\{\eta[k]\} = 0$), we can say that $\boldsymbol{\epsilon}_k$ is zero-mean. Defining the signal and noise variances as $\lambda^2 = E\{s_k[i]s_k[i]^*\}$ and $\sigma^2 = E\{\eta[k]\eta[k]^*\}$ respectively, we can derive the covariance matrix of $\boldsymbol{\epsilon}_k$ from (2) as $E\{\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^H\} \triangleq \mathbf{Q} = \lambda^2\mathbf{H}_s\mathbf{H}_s^H + \sigma^2\mathbf{I}$. Defining n_s as the length of the shortest sequence of data symbols ($n_s = \min_k \{n_{k+1} - (n_k + n_t - 1)\}$), we assume $n_s \geqslant 2L$. This ensures that the $\boldsymbol{\epsilon}_k$'s are uncorrelated. The first and second order statistics of the stochastic term are thus as follows: $E\{\boldsymbol{\epsilon}_k\} = \mathbf{0}$ **(C1)**, $E\{\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k^H\} = \mathbf{Q}$ **(C2)** and $E\{\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_l^H\} = \mathbf{0}, \quad \forall k \neq l$ **(C3)**.

III. Maximum Likelihood Channel Estimation

The noise part $\boldsymbol{\eta}_k$ of $\boldsymbol{\epsilon}_k$ can be considered as circularly gaussian distributed (which corresponds to the classical AWGN approximation), but this is not the case for the data part. However, we can apply the gaussian likelihood function as a statistically sound fitting criterion. In other terms, we consider the $\boldsymbol{\epsilon}_k$'s to be circularly gaussian distributed. Relying on this approximation, we can express (up to a constant term) the negative log-likelihood function of the system as:

$$-\mathcal{L} = K \ \ln|\mathbf{Q}| + \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}_k\mathbf{h})^H \mathbf{Q}^{-1} (\mathbf{u}_k - \mathbf{T}_k\mathbf{h}). \tag{3}$$

Relying on the definition of $\mathbf{Q}$, the log-likelihood can be expressed as a direct function of the unknown parameters $\mathbf{h}$ and σ^2. The ML channel estimate minimizes this expression w.r.t. $\mathbf{h}$ and σ^2.This minimization problem boils down to a computationally demanding $(L + 1)$-dimensional nonlinear search.

To overcome this complexity problem, we propose to disregard the structure of $\mathbf{Q}$, and ignore the relation that binds it to the parameters $\mathbf{h}$ and σ^2. We thus assume that the covariance matrix $\mathbf{Q}$ of the stochastic term $\boldsymbol{\epsilon}_k$ can be any symmetric positive definite matrix, regardless of $\mathbf{h}$ and σ^2. These assumptions (unstructured $\mathbf{Q}$ and gaussian $\boldsymbol{\epsilon}_k$) transform the initial ML problem into an optimization problem which is separable in its two variables $\mathbf{Q}$ and $\mathbf{h}$. We exploit this separation property in the next paragraphs in order to solve the ML problem in a less complex way than the $(L + 1)$-dimensional nonlinear search.

IV. Iterative Solution

When a minimization problem is separable in its variables, a solution that is commonly used is an iterative one. Each iteration consists in analytically minimizing the cost function with respect to one variable whilst keeping the other(s) fixed. The variable with respect to which the cost function is minimized is changed in each iteration. This procedure converges to a minimum of the cost function. If the starting point is accurate enough or if the surface is smooth enough, the point of convergence is the global minimum of the cost function, which corresponds to the ML estimate of $\mathbf{Q}$ and $\mathbf{h}$ in this case.

Assume that at the ith iteration an estimate $\hat{\mathbf{Q}}^i$ of the covariance matrix $\mathbf{Q}$ is available. We first seek the channel estimate $\hat{\mathbf{h}}^i$ that minimizes the cost function (3) with respect to $\mathbf{h}$ under the hypothesis that the available $\hat{\mathbf{Q}}^i$ is the true $\mathbf{Q}$, i.e. we compute $\hat{\mathbf{h}}^i = \mathbf{h}_{ML}(\mathbf{Q}^i)$, where $\mathbf{h}_{ML}(\mathbf{Q}) = \arg\min_{\mathbf{h}} -\mathcal{L}$, whose solution can be computed as:

$$\mathbf{h}_{ML}(\mathbf{Q}) = \left(\sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{T}_k\right)^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{u}_k. \tag{4}$$

We then seek the covariance matrix $\hat{\mathbf{Q}}^{i+1}$ that minimizes (3) under the hypothesis that this new channel estimate $\hat{\mathbf{h}}^i$ is the true $\mathbf{h}$: $\hat{\mathbf{Q}}^{i+1} = \mathbf{Q}_{ML}(\hat{\mathbf{h}}^i)$, where $\mathbf{Q}_{ML}(\mathbf{h}) = \arg\min_{\mathbf{Q}} -\mathcal{L}$, whose solution can be computed as [6, pp. 200-202]:

$$\mathbf{Q}_{ML}(\mathbf{h}) = K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}_k\mathbf{h}) (\mathbf{u}_k - \mathbf{T}_k\mathbf{h})^H. \tag{5}$$

$\hat{\mathbf{Q}}^{i+1}$ is then used as a starting point for the next iteration. The procedure is stopped when there is no significant difference between the estimates produced by two consecutive iterations. We still have to find an acceptable starting point for the

iterations. In [7], we proposed an iterative method for channel identification in a similar context. It is straightforward to see that this method, which was presented as an iterative Weighted Least Squares one, is actually similar to the one we propose here and is thus an iterative ML method. We show in [7] that initializing the iterative procedure with a simple Least Squares channel estimate yields good convergence properties. It is easy to show that applying the method we propose here with an identity matrix as initial covariance matrix yields exactly the same path through the iterations. We thus propose to initialize the iterative ML method with $\hat{\mathbf{Q}}^0 = \mathbf{I}$.

V. Closed Form Solution

An alternative strategy to the iterative procedure described above consists in directly finding an analytical expression for the global minimum of the likelihood function. The separation property of the cost function can be exploited again in order to find this global minimum. First observe that the likelihood function (3) can be expressed as:

$$-\mathcal{L} = K\ln|\mathbf{Q}| + \text{tr}\left(\mathbf{Q}^{-1}\sum_{k=1}^{K}(\mathbf{u}_k - \mathbf{T}_k\mathbf{h})(\mathbf{u}_k - \mathbf{T}_k\mathbf{h})^H\right). \tag{6}$$

We first minimize this cost function with respect to $\mathbf{Q}$ leading to $\mathbf{Q}_{ML}(\mathbf{h})$ as given by (5). Replacing $\mathbf{Q}$ by $\mathbf{Q}_{ML}(\mathbf{h})$ in (6) leaves us with an expression of the cost function that only depends on $\mathbf{h}$:

$$-\mathcal{L} = K\ln\left|K^{-1}\sum_{k=1}^{K}(\mathbf{u}_k - \mathbf{T}_k\mathbf{h})(\mathbf{u}_k - \mathbf{T}_k\mathbf{h})^H\right| + K\text{tr}(\mathbf{I})$$

Because the problem is separable in its two variables, minimizing this new expression of the cost function w.r.t. $\mathbf{h}$ yields the global minimum. The ML channel estimate $\mathbf{h}_{ML}$ is thus computed as:

$$\mathbf{h}_{ML} = \arg\min_{\mathbf{h}}\left|\sum_{k=1}^{K}(\mathbf{u}_k - \mathbf{T}_k\mathbf{h})(\mathbf{u}_k - \mathbf{T}_k\mathbf{h})^H\right|. \tag{7}$$

Define the following:

$$\begin{aligned}
&\hat{\mathbf{h}}_{LS} \triangleq \left(\sum_{k=1}^{K}\mathbf{T}_k^H\mathbf{T}_k\right)^{-1}\sum_{k=1}^{K}\mathbf{T}_k^H\mathbf{u}_k,\\
&\mathbf{g} \triangleq \mathbf{h} - \hat{\mathbf{h}}_{LS},\\
&\mathbf{g}_{ML} \triangleq \mathbf{h}_{ML} - \hat{\mathbf{h}}_{LS},\\
&\mathbf{e}_k \triangleq \mathbf{u}_k - \mathbf{T}_k\hat{\mathbf{h}}_{LS},\\
&\hat{\mathbf{Q}} \triangleq K^{-1}\sum_{k=1}^{K}\mathbf{e}_k\mathbf{e}_k^H.
\end{aligned} \tag{8}$$

where $\hat{\mathbf{Q}}$ is assumed to be positive definite[1]. Using these notations, the minimization problem (7) can be rephrased as:

$$\mathbf{g}_{ML} = \arg\min_{\mathbf{g}}\left|\sum_{k=1}^{K}(\mathbf{e}_k - \mathbf{T}_k\mathbf{g})(\mathbf{e}_k - \mathbf{T}_k\mathbf{g})^H\right|. \tag{9}$$

[1]We see from (8) that a necessary condition therefore is $K \geqslant n_t + L$. When this condition is fulfilled, the randomness of the noise and the data ensures that $\hat{\mathbf{Q}}$ is a positive definite matrix with probability 1.

The determinant in this last expression can be expressed as (up to a positive factor):

$$\left|\mathbf{I} + \hat{\mathbf{Q}}^{-1}K^{-1}\sum_{k=1}^{K}\mathbf{T}_k\mathbf{g}\mathbf{g}^H\mathbf{T}_k^H - \mathbf{T}_k\mathbf{g}\mathbf{e}_k^H - \mathbf{e}_k\mathbf{g}^H\mathbf{T}_k^H\right|. \tag{10}$$

When K is large, both $\hat{\mathbf{h}}_{LS}$ and $\mathbf{h}_{ML}$ are close to the true $\mathbf{h}$. We can thus assume that $\mathbf{g}$, and consequently the second term in (10), are small in the vicinity of the solution. It is well known that, for $||\mathbf{\Delta}|| \ll 1$, $|\mathbf{I} + \mathbf{\Delta}| \approx 1 + tr(\mathbf{\Delta})$. Hence, for $K \gg 1$, the minimization problem in (9) can be approximated by:

$$\arg\min_{\mathbf{g}} tr\left(\hat{\mathbf{Q}}^{-1}\sum_{k=1}^{K}\mathbf{T}_k\mathbf{g}\mathbf{g}^H{\mathbf{T}_k}^H - \mathbf{T}_k\mathbf{g}\mathbf{e}_k^H - \mathbf{e}_k\mathbf{g}^H\mathbf{T}_k^H\right).$$

The solution $\hat{\mathbf{g}}_{ML}$ of this problem, which is an approximate of the true $\mathbf{g}_{ML}$, can be computed as:

$$\hat{\mathbf{g}}_{ML} = \left(\sum_{k=1}^{K}{\mathbf{T}_k}^H\hat{\mathbf{Q}}^{-1}\mathbf{T}_k\right)^{-1}\sum_{k=1}^{K}\mathbf{T}_k^H\hat{\mathbf{Q}}^{-1}\mathbf{e}_k. \tag{11}$$

Rephrasing (11) provides us with an approximate $\hat{\mathbf{h}}_{ML}$ of the true ML channel estimate $\mathbf{h}_{ML}$:

$$\hat{\mathbf{h}}_{ML} = \left(\sum_{k=1}^{K}{\mathbf{T}_k}^H\hat{\mathbf{Q}}^{-1}\mathbf{T}_k\right)^{-1}\sum_{k=1}^{K}\mathbf{T}_k^H\hat{\mathbf{Q}}^{-1}\mathbf{u}_k. \tag{12}$$

Based on the first and second order statistics of $\boldsymbol{\epsilon}_k$, it is easy to check that $\lim_{K\to\infty}\hat{\mathbf{Q}} = \mathbf{Q}$. Exploiting this result, it is possible to prove that the proposed channel estimate is consistent:

$$\lim_{K\to\infty}\hat{\mathbf{h}}_{ML} = \mathbf{h}.$$

Using this approximation of the ML channel estimate, we can derive the corresponding ML covariance matrix estimate replacing $\mathbf{h}$ by $\hat{\mathbf{h}}_{ML}$ in (5):

$$\hat{\mathbf{Q}}_{ML} = K^{-1}\sum_{k=1}^{K}\left(\mathbf{u}_k - \mathbf{T}_k\hat{\mathbf{h}}_{ML}\right)\left(\mathbf{u}_k - \mathbf{T}_k\hat{\mathbf{h}}_{ML}\right)^H,$$

which differs from $\hat{\mathbf{Q}}$.

In [7], it appears that the iterative method seems to almost converge after a single iteration. The above discussion on the approximate closed form ML channel estimate explains this fast convergence. It is straightforward to see that initializing the iterative method with $\hat{\mathbf{Q}}^0 = \mathbf{I}$ yields $\hat{\mathbf{h}}^0 = \hat{\mathbf{h}}_{LS}$. It follows that $\hat{\mathbf{Q}}^1 = \hat{\mathbf{Q}}$ and, observing the similarity between (4) and (12), $\hat{\mathbf{h}}^1 = \hat{\mathbf{h}}_{ML}$. Hence the iterative procedure yields the approximate closed form ML channel estimate after one iteration. Since $\hat{\mathbf{h}}_{ML}$ is close to the convergence point $\mathbf{h}_{ML}$, the iterative method approximately converges in one iteration.

VI. Cramer-Rao Bound

The Cramer-Rao Bound (CRB) is a theoretical lower bound on the covariance matrix of an unbiased channel estimate. We seek here an expression for this bound for the identification problem under consideration in this paper. In order to find

that expression, we use an alternative formulation of the data model. Define $\mathbf{T}$ as the collection of all the $\mathbf{T}_k$ matrices: $\mathbf{T} = \left[\mathbf{T}_1^H, \cdots, \mathbf{T}_K^H\right]^H$, $\mathbf{u}$ as the collection of all $\mathbf{u}_k$ vectors: $\mathbf{u} = \left[\mathbf{u}_1^H, \cdots, \mathbf{u}_K^H\right]^H$, and $\boldsymbol{\epsilon}$ as the collection of all $\boldsymbol{\epsilon}_k$ vectors: $\boldsymbol{\epsilon} = \left[\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_K\right]^H$. Collecting the K equations from (1) allows us to express the data model as:

$$\mathbf{u} = \mathbf{T}\mathbf{h} + \boldsymbol{\epsilon}. \tag{13}$$

It is clear from the statistics of $\boldsymbol{\epsilon}_k$ that $\boldsymbol{\epsilon}$ is zero-mean ($E\{\boldsymbol{\epsilon}\} = \mathbf{0}$) and that its covariance $\mathbf{Q}_{tot} = E\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^H\}$ is a block-diagonal matrix with $\mathbf{Q}$ repeated along the main diagonal. In this context and under the assumptions that allowed us to derive the iterative ML solution, the CRB, $\boldsymbol{\mathcal{J}}^{-1}$, is well known to be [6, pp. 564]: $\boldsymbol{\mathcal{J}}^{-1} = (\mathbf{T}^H\mathbf{Q}_{tot}^{-1}\mathbf{T})^{-1}$. Exploiting the block-diagonal structure of $\mathbf{Q}_{tot}$, this result can be written as:

$$\boldsymbol{\mathcal{J}}^{-1} = \left(\sum_{k=1}^{K} \mathbf{T}_k^H\mathbf{Q}^{-1}\mathbf{T}_k\right)^{-1}.$$

Even though our estimator is not proven to be unbiased, the CRB remains a good indicator of the theoretically achievable performance of a channel estimator. We thus use it as a benchmark in the experimental results section. Note that this bound depends both on the channel realization (through the covariance matrix $\mathbf{Q}$) and on the chosen training sequences.

VII. Experimental Results

The performance metric that is used throughout this section is the expected value of the Normalized Mean Square Error (NMSE) of the proposed channel estimate: NMSE $= \mathrm{E}\left\{\frac{||\hat{\mathbf{h}}-\mathbf{h}||^2}{||\mathbf{h}||^2}\right\}$. We use the CRB as a benchmark in the experiments. The CRB curves displayed on the graphs represent the expected value of the NMSE of an estimator that achieves the CRB, which is $\frac{\mathrm{tr}(\boldsymbol{\mathcal{J}}^{-1})}{||\mathbf{h}||^2}$. The experiments are performed on convolutive Rayleigh Fading channels of varying order L. The different channel taps are assumed to be independently identically distributed. The training sequences are random constant modulus sequences. The presented results are obtained after averaging over a set of 100 channel realizations. For each of these channel realizations, the results are averaged over 100 different sets of training sequences. The Signal to Noise Ratio (SNR) is defined as $SNR = \frac{||\mathbf{h}||^2\lambda^2}{\sigma^2}$.

A. Performance of the proposed method

In this section, we analyze the performance of the proposed method and see how it compares with the CRB.

In Fig. 1, we compare the NMSE of our method with the CRB as a function of the SNR. This comparison is done for different values of the channel order and different values of the number of data blocks K. We see that the simulated results match the CRB quite well, especially for large values of K and for low SNRs. When the channel order is small, the CRB decreases with a constant slope. The simulated results follow this behavior if K is sufficiently large but saturate at high SNRs if K is too small. When the channel order gets larger, the NMSE appears to saturate at higher SNRs both for the CRB and the simulated results. This floor appears when there is no exact solution to the channel identification problem in the noiseless case, this is when $n_t \geqslant L+1$ is not fulfilled.

We mentioned in the introduction that using changing training sequences rather than repeating the same training sequences improved the accuracy of the channel estimate. The compared experimental performance of these schemes are not presented here but can be found in [7]. When a constant training sequence is repeatedly inserted in the data symbols stream, the condition for the existence of an exact solution to the channel identification problem in the noiseless case is more stringent, i.e. the condition is now $n_t \geqslant 2L+1$. When this condition is not fulfilled, there is a floor in the NMSE of the channel estimate at high SNRs. Hence, when n_t is fixed, for intermediate channel orders, i.e. channel orders that fulfill $\frac{n_t-1}{2} \leqslant L \leqslant n_t - 1$, using changing training sequences will yield a constant slope in the channel estimate NMSE for increasing SNRs whereas a floor will appear at high SNRs if a constant training sequence is repeated. When the channel order is outside this interval, both schemes yield the same slope in the channel estimate NMSE but experimental results show that there is still an advantage in using changing training sequences.

In Fig. 2, we evaluate the impact of the number of data blocks K on the NMSE. The CRB decreases with a constant slope as K increases. We observe a good match between this bound and the experimental results for large values of the channel order L. For smaller channel orders, there is a larger difference between the CRB and the simulated results. The mismatch decreases as K increases and SNR decreases.

B. Comparison with Traditional ML Methods

Classical training-based ML channel estimation techniques discard the received samples that are corrupted by contributions of the unknown data symbols. They rely on the received symbols that only contain contributions from the known training symbols. Such received symbols can be observed at the receiver only when $n_t \geqslant L+1$. In this case, the data model is changed into $\mathbf{u}'_k = \mathbf{T}'_k\mathbf{h} + \boldsymbol{\epsilon}'_k$, where $\mathbf{u}'_k = \mathbf{u}_k(L+1:n_t)$, $\mathbf{T}'_k = \mathbf{T}_k(L+1:nt,:)$ and $\boldsymbol{\epsilon}'_k$ is the AWGN term. Note that $\mathbf{T}'_k$ is an $(n_t - L) \times (L+1)$ Toeplitz matrix with $[t_k[L+1], \ldots, t_k[n_t]]^T$ as first row and $[t_k[L+1], \ldots, t_k[1]]$ as first column. Similarly to what we did in (13), we can collect these equations in a larger data model: $\mathbf{u}' = \mathbf{T}'\mathbf{h} + \boldsymbol{\epsilon}'$. When the noise term is not colored, which is the case here, the solution of the ML channel identification problem is well known to be a simple least squares (LS) fit of $\mathbf{T}'\mathbf{h}$ to $\mathbf{u}'_k$. The classical ML estimate can be expressed as: $\mathbf{h}_{ML} = \left(\sum_{k=1}^{K} \mathbf{T}'^H_k\mathbf{T}'_k\right)^{-1}\sum_{k=1}^{K}\mathbf{T}'_k\mathbf{u}'_k$. Note that when the condition on L and n_t is fulfilled and the traditional method has a solution, the new method has a constant slope. When the channel order increases and this condition is not fulfilled, all the received symbols contain contributions from the unknown data symbols and the traditional method cannot be applied,

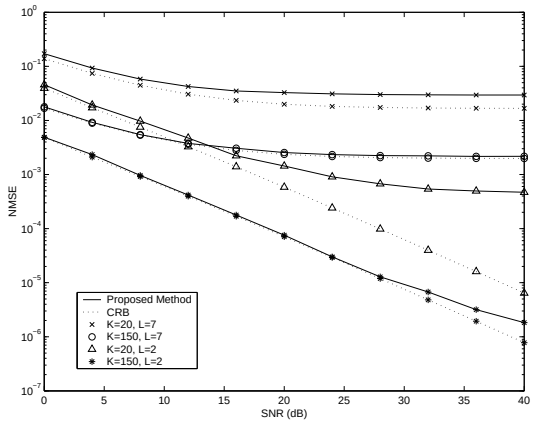

Fig. 1. Comparison of the simulated and theoretical NMSE vs. SNR for different channel orders when $n_T = 5$. The results are plotted for two different values of K, namely 20 and 150.

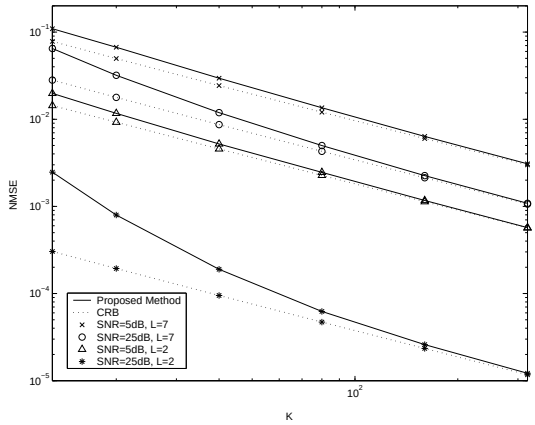

Fig. 2. Comparison of the simulated and theoretical NMSE vs. K for different channel orders when $n_T = 5$. The results are plotted for two different values of the SNR, namely 5 and 25dB.

whereas the proposed ML method still provides us with accurate estimates of the channel.

In Fig. 3, we compare the results for different channel orders when the length of the training sequence n_t is fixed. We see that the new method and the traditional one yield an equivalent performance when the channel order is small. When the channel order increases, the new method clearly outperforms the classical one, especially at low SNRs. When the channel order keeps growing and $n_t \geqslant 2L+1$ is not fulfilled anymore, the new method still provides reliable channel estimates, whilst the traditional method cannot be applied anymore.

VIII. Conclusions

In this paper, we presented a new training based ML channel identification method. We first proposed an iterative ML method and then derived an approximate closed form expression for the ML channel estimate. This new ML method outperforms classical training based ML estimation methods. The reason for this is that all the energy that is received from the known training symbols is exploited in order to

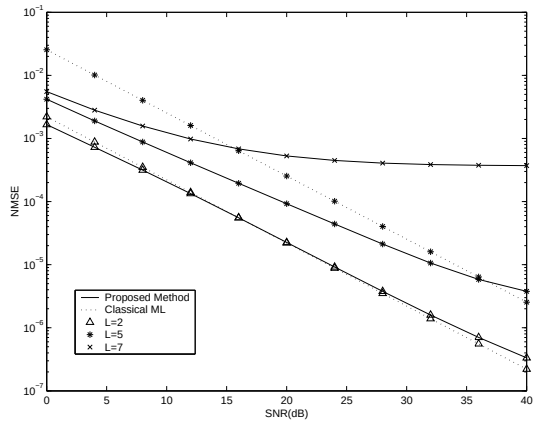

Fig. 3. Simulated NMSE vs. SNR for the proposed and traditional ML channel estimation methods for different channel orders when $n_T = 6$ and $K = 400$.

estimate the channel, which is not the case for traditional methods. Furthermore, this new method approaches the CRB and, as opposed to traditional ML methods, is able to provide us with accurate channel estimates, even when the channel order increases to values that make it impossible to use the classical ML method.

Acknowledgments:

This work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of the Belgian State's IUAP Programme (2002-2007): IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modeling') and IUAP P5/11 ('Mobile multimedia communication systems and networks') and the Concerted Research Action GOA-MEFISTO-666 of the Flemish Government: Research Project FWO nr.G.0196.02 ('Design of efficient communication techniques for wireless time-dispersive multi-user MIMO systems'). It was partially sponsored by the Swedish Science Council. The scientific responsibility is assumed by its authors.

References

[1] H. Vikalo, B. Hassibi, B. Hochwald, T. Kailath, "Optimal Training for Frequency-Selective Fading Channels," in *Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Salt Lake City, Utah, May 2001.

[2] O. Rousseaux, G. Leus, P. Stoica and M. Moonen, "Training Based Maximum Likelihood Channel Identification: Constant Training Sequences," in *Signal Processing Advances in Wireless Communications (SPAWC 2003)*, Rome, Italy, June 2003, accepted for publication.

[3] 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.

[4] L. Deneire, B. Gyselinckx and M. Engels., "Training Sequence vs. Cyclic Prefix: A New Look on Single Carrier Communication," *IEEE Communication Letters*, vol. 5, no. 7, pp. 292–294, 2001.

[5] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels (mobile radio)," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 4, pp. 686–693, November 1991.

[6] T. Söderström, P. Stoica, *System Identification*, International Series in Systems and Control Engineering. Prentice Hall, 1989.

[7] O. Rousseaux, G. Leus, P. Stoica and M. Moonen, "A Stochastic Method for Training Based Channel Identification," in *Seventh International Symposium on Signal Processing and its Applications (ISSPA 2003)*, Paris, France, July 2003, submitted.