Gaussian Maximum Likelihood Channel Estimation with Short Training Sequences*

Olivier Rousseaux[1], Geert Leus[2], Petre Stoica[3] and Marc Moonen[1]

[1] K.U.Leuven - ESAT

Kasteelpark Arenberg 10, 3001 Leuven, Belgium

Email: olivier.rousseaux@esat.kuleuven.ac.be

[2] T. U. Delft - Department of Electrical Engineering

Mekelweg 4, 2628CD Delft, The Netherlands

Email: leus@cas.et.tudelft.nl

[3] Uppsala University - Department of Systems and Control

P.O. Box 337 SE-751 05, Uppsala Sweden

Email: ps@syscon.uu.se

* This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of the Belgian State, Prime Minister's Office - Federal Office for Scientific, Technical and Cultural Affairs - Inter-university Poles of Attraction Programme (2002-2007) - IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modeling') and P5/11 ('Mobile multimedia communication systems and networks'), 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 ('Design of efficient communication techniques for wireless time-dispersive multi-user MIMO systems') and was partially sponsored by the Swedish Science Council

Abstract

In this paper, we address the problem of identifying convolutive channels using a Gaussian Maximum Likelihood (ML) approach when short training sequences (possibly shorter than the channel impulse response length) are periodically inserted in the transmitted signal. We consider the case where the channel is quasi-static (i.e. the sampling period is several orders of magnitude smaller than the coherence time of the channel). Several training sequences can thus be used in order to produce the channel estimate. We derive an expression of the Gaussian likelihood function for this system and introduce a simplifying hypothesis that is needed in order to derive the proposed ML channel estimation methods. We check the validity of this hypothesis through an analysis of the Cramer Rao Bound. We propose an *iterative method* that rapidly converges to the Gaussian ML channel estimate and then derive an approximate low complexity *closed-form expression* of this Gaussian ML channel estimate. We show that the proposed closed form channel estimates are consistent and asymptotically efficient. The key feature of the proposed method is that all channel output samples containing contributions from the training sequences (including those containing contributions from the unknown surrounding data symbols) are exploited towards the estimation of the channel, whilst most existing training-based methods only rely on the channel output samples containing contributions from the training sequences only. We present some simulation results and compare the performance of the proposed methods with the Cramer-Rao Bound.

I. INTRODUCTION

A major impediment of broadband communication systems is that the sampling period can become smaller than the delay spread of the channel, especially in 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. These can be divided in two families which are termed “blind” and “training-based”. Blind algorithms estimate the channel based on properties of the transmitted signals (finite alphabet properties, higher order statistics, cyclo-stationarity, ... see e.g. [1], [2] or [3], and references therein). Training-based techniques assume that known symbols (training or pilot symbols) are inserted in the transmitted

signals. It is then possible to identify the channel at the receiver exploiting the knowledge of these known symbols. Semi-blind techniques have been proposed recently, which allow to improve the performance of blind techniques by exploiting both the knowledge of known symbols and properties of the transmitted signals.

In this paper, we focus on semi-blind and 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. [4], [5] or [6]). Most existing methods require the training sequences to be significantly longer than the channel impulse response. In this case, some of the received symbols contain contributions only 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). In Additive White Gaussian Noise (AWGN) conditions, the problem of performing Maximum Likelihood (ML) channel identification when only these data-free received symbols are used is equivalent to a simple 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.

Most transmission schemes include some known symbols that are used for synchronization or as guard intervals, and several existing transmission schemes insert short sequences of known symbols in the stream of unknown data symbols (e.g. Known Symbol Padding (KSP) transmission [7] or Pilot Symbol Assisted Modulation (PSAM) [8]), possibly allowing an accurate channel estimation solely relying on the knowledge of these short sequences of known symbols (and thus allowing to avoid the bandwidth-consuming insertion of long training sequences traditionally required by training-based channel identification algorithms). Neither classical training-based channel identification algorithms nor the purely blind ones are well suited to optimally identify the channel in this context. Training-based methods will discard most (if not all) of the channel output samples containing contributions from the known symbols, providing largely sub-optimal

channel estimates, whilst purely blind techniques do not offer the possibility of exploiting the knowledge of the inserted pilot symbols.

Several semi-blind techniques have been proposed recently that allow accurate channel estimation in this context. A first technique, which was proposed in the framework of KSP transmission [9], is deterministic and exploits all the energy that is received from the training sequences. This method only works when a constant training sequence, which must be at least as long as the channel order, is periodically inserted in the transmitted sequence. Several semi-blind ML channel estimation techniques have been proposed as well. Depending on the hypothesis upon which the expression of the likelihood function is built, one can distinguish between two families of ML methods: Deterministic ML in which the data symbols are considered as deterministic, and Gaussian ML in which the data symbols are assumed to be Gaussian distributed. Some deterministic ML methods are presented in [10] and [11] for instance. However, in [11], a theoretical comparison of the CRB indicates that Gaussian ML methods outperform deterministic ML methods. In [12], a first Gaussian ML method is proposed in the specific context of colored noise and co-channel interference. A method based on hidden Markov models was presented in [13], but performs significantly worse than the achievable CRB. Finally, several Gaussian ML methods are proposed in [14]: a Pseudo-Quadratic ML (PQML) method that is developed only for the AWGN channel case and an approximate semi-blind ML method that requires Single Input Multiple Output (SIMO) channels. None of these methods achieves the Gaussian CRB.

We present here a new Gaussian ML method for Single Input Single Output (SISO) channel identification that is able to cope with arbitrarily short training sequences (possibly shorter than the channel impulse response length) and performs channel estimation exploiting all the received symbols that contain contributions from the training sequences. The proposed method asymptotically achieves the CRB and has a small computational complexity. We consider a transmission scheme where training sequences are inserted between blocks of data symbols. For the sake of simplicity, we consider all the training sequences to have the same length, but it is

straightforward to adapt the method to the more general case of training sequences of variable length. We consider quasi-static channels (the channel stays constant during the transmission of several blocks of data). We investigate both the situation where the same training sequence is repeated after each block of data and the situation where the training sequence is changed after each block of data.

The structure of the paper is as follows. In section II, we present our data model. In section III, we derive an expression for the gaussian likelihood function of a channel estimate and introduce some approximations on which we have to rely in order to derive low-complexity ML channel estimates. In section IV, we show that the proposed approximation has a negligible impact on the achievable performance of ML channel estimation through a CRB analysis. We then propose an iterative algorithm that converges to the ML channel estimate (section V). We next derive an approximate closed form expression of the ML channel estimate, both for a constant (section VI-A) and for a changing training sequence (section VI-B). We experimentally test the proposed methods and compare them with classical ML methods in section VIII, and finally draw some conclusions in section IX.

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 $\mathbf{0}_{M\times N}$ is the all-zero matrix of size $M \times N$. The operator $(.)^*$ denotes the complex conjugate, Re(.) the real part and Im(.) the imaginary part of a complex number. The superscript $(.)^T$ denotes the transpose of a matrix and $(.)^H$ the complex conjugate transpose. Finally, tr$(.)$ denotes the trace of a matrix, $|.|$ its determinant and $\mathbf{A}(i,j)$ denotes the ith element of the jth column of $\mathbf{A}$.

II. Data Model

We consider a stationary Finite Impulse Response (FIR) convolutive channel 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:

$$y[n] = \sum_{i=0}^{L} h[i]x[n-i] + \eta[n], \tag{1}$$

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

As mentioned in the introduction, we consider a transmission scheme where constant length training sequences are inserted between blocks of data symbols. There are two different possibilities: either the same training sequence is repeated after each block of data, or the training sequence is changed after each block. We refer to these two alternative schemes as the constant training sequence case and the changing training sequence case. As we will see later, these two alternative schemes yield different channel identification procedures. We describe the more general situation of a changing training sequence as often as possible, and only analyze the case of a constant training sequence when explicitly needed.

A total number of K training sequences is inserted in the stream of unknown data symbols. The kth training sequence, $\mathbf{t}_k = [t_k[1], \ldots, t_k[n_t]]^T$, starts at position n_k: $[x[n_k], \ldots, x[n_k+n_t-1]]^T = \mathbf{t}_k$. 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:

$$\mathbf{u}_k = \mathbf{T}_k\mathbf{h} + \boldsymbol{\epsilon}_k, \tag{2}$$

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

$$\boldsymbol{\epsilon}_k = \underbrace{\left[\begin{array}{ccc|ccc} h_L & \cdots & h_1 & & & \mathbf{0} \\ & \ddots & \vdots & h_0 & & \\ & & h_L & \vdots & \ddots & \\ \mathbf{0} & & & h_{L-1} & \cdots & h_0 \end{array}\right]}_{\mathbf{H}_s (n_t + L) \times (2L)} \mathbf{s}_k + \boldsymbol{\eta}_k, \tag{3}$$

where $\mathbf{s}_k = [s_k[1], \cdots, s_k[2L]] = [x[n_k - L], \ldots, x[n_k - 1], x[n_k + n_t], \ldots, x[n_k + n_t + L - 1]]^T$ is the vector of surrounding data symbols, and $\boldsymbol{\eta}_k = [\eta[n_k], \cdots, \eta[n_k + n_t + L - 1]]$ is the AWGN term. Assuming that both the noise and the data are white and zero-mean ($E\{s_k[i]s_k[j]^*\} = E\{\eta[i]\eta[j]^*\} = 0, \quad \forall i, j, k : i \neq j$, and $E\{s_k[i]\} = E\{\eta[k]\} = 0$), we can say that $\boldsymbol{\epsilon}_k$ is zero-mean. 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 $\mathbf{s}_k$'s are uncorrelated, i.e. $E\left\{\mathbf{s}_k \mathbf{s}_l^H\right\} = \mathbf{0} \quad \forall k, l : k \neq l$. 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 first and second order statistics of $\boldsymbol{\epsilon}_k$:

$$\begin{aligned} \mathrm{E}\{\boldsymbol{\epsilon}_k\} &= \mathbf{0}_{(n_t+L)\times 1}, \\ \mathrm{E}\left\{\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^H\right\} \triangleq \mathbf{Q} &= \lambda^2 \mathbf{H}_s \mathbf{H}_s^H + \sigma^2 \mathbf{I}, \\ \mathrm{E}\left\{\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_l^H\right\} &= \mathbf{0}_{(n_t+L)\times(n_t+L)} \quad \forall k, l : k \neq l. \end{aligned} \tag{4}$$

III. Maximum Likelihood Approach for Channel Identification

As discussed in the introduction, there are two alternative ways of expressing the Likelihood function of our system. The expression of the deterministic likelihood function is obtained considering the unknown data symbols as deterministic disturbances. The Gaussian likelihood function is established making the hypothesis that these symbols are Gaussian variables, hence $\mathbf{u}_k = \mathcal{N}(\mathbf{T}_k\mathbf{h}, \mathbf{Q})$. It has been shown in [11], that the Gaussian hypothesis yields more accurate channel estimates. Adopting this hypothesis, we can express (up to a constant term) the negative Gaussian log-likelihood function of the system as:

$$-\mathcal{L} = K \ \ln|\mathbf{Q}| + \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}_k\mathbf{h})^H \mathbf{Q}^{-1} (\mathbf{u}_k - \mathbf{T}_k\mathbf{h}). \tag{5}$$

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 corresponding ML channel estimate minimizes this expression w.r.t. $\mathbf{h}$ and σ^2. This minimization problem boils down to a computationally demanding $(L+2)$-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. This hypothesis turns the initial ML problem into a new one. We call the initial problem the *parametric ML problem*; the problem resulting from the proposed approximations will be called the *non-parametric ML problem*. The non-parametric ML channel estimate thus maximizes the likelihood function w.r.t $\mathbf{h}$ and $\mathbf{Q}$ (instead of $\mathbf{h}$ and σ^2). These assumptions transform the parametric ML problem in $\mathbf{h}$ and σ^2into a new optimization problem which is separable in its two variables $\mathbf{h}$ and $\mathbf{Q}$. We exploit this separability property in the next sections in order to solve the optimization problem in a less complex way than the $(L+2)$-dimensional nonlinear search of the parametric ML problem. The solution of the non-parametric ML problem differs from the solution of the parametric ML. Hence, it is worthwhile to first check the impact of the proposed hypothesis on the accuracy of the resulting ML channel estimates. This is what we do in the next section through an analysis of the respective Cramer-Rao Bounds.

IV. Cramer Rao Bounds

We show later in the text (see section VII) that the channel estimates derived from the non-parametric ML problem are consistent and thus asymptotically unbiased. The Cramer-Rao Bound (CRB) is a theoretical lower bound on the covariance matrix of an unbiased estimate. In this section, we analyze the impact of the non-parametric hypothesis on the accuracy of the derived channel estimate through this theoretical bound. It can be shown (see e.g. [15, pp. 562]) that for any unbiased estimate $\hat{\mathbf{h}}$ of the parameter vector $\mathbf{h}$, the following inequality holds:

$$\text{cov}(\hat{\mathbf{h}}) \geqslant \boldsymbol{\mathcal{J}}(\mathbf{h})^{-\mathbf{1}},$$

where the covariance matrix of the channel estimate is defined as $\text{cov}(\hat{\mathbf{h}}) = E\left\{(\hat{\mathbf{h}}-\mathbf{h})(\hat{\mathbf{h}}-\mathbf{h})^H\right\}$, and the Fisher Information Matrix (FIM) is defined as $\boldsymbol{\mathcal{J}}(\mathbf{h}) = E\left\{\left(\frac{\partial \mathcal{L}}{\partial \mathbf{h}}\right)^T\left(\frac{\partial \mathcal{L}}{\partial \mathbf{h}}\right)\right\}$, $\mathcal{L}$ being (up to a constant term) the Gaussian log-likelihood function of the system. Adapting the results

presented in [11], the real FIM of the *parametric ML problem* can be formulated as:

$$\boldsymbol{\mathcal{J}}(\mathbf{h}) = 2 \begin{bmatrix} \text{Re}(\mathbf{J}_1) & \text{-Im}(\mathbf{J}_1) \\ \text{Im}(\mathbf{J}_1) & \text{Re}(\mathbf{J}_1) \end{bmatrix} + 2 \begin{bmatrix} \text{Re}(\mathbf{J}_2) & \text{-Im}(\mathbf{J}_2) \\ \text{Im}(\mathbf{J}_2) & \text{Re}(\mathbf{J}_2) \end{bmatrix}, \tag{6}$$

where

$$\mathbf{J}_1(i,j) = \left(\sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{T}_k \right)(i,j) + \text{tr} \left\{ \mathbf{Q}^{-1} \frac{\partial \mathbf{Q}}{\partial h[i-1]^*} \mathbf{Q}^{-1} \frac{\partial \mathbf{Q}}{\partial h[j-1]^*} \right\}, \tag{7}$$

$$\mathbf{J}_2(i,j) = \text{tr} \left\{ \mathbf{Q}^{-1} \frac{\partial \mathbf{Q}}{\partial h[i-1]^*} \mathbf{Q}^{-1} \frac{\partial \mathbf{Q}}{\partial h[j-1]^*} \right\}, \tag{8}$$

and

$$\frac{\partial \mathbf{Q}}{\partial h[i]^*} = \lambda^2 \mathbf{H}_s \left(\frac{\partial \mathbf{H}_s}{\partial h[i]} \right). \tag{9}$$

The approximation inserted in the *non-parametric ML problem* simplifies the expression of the FIM since the $\frac{\partial \mathbf{Q}}{\partial h[i]^*}$ terms are equal to zero. The complex FIM can be used and is expressed as

$$\boldsymbol{\mathcal{J}}(\mathbf{h}) = \left(\sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{T}_k \right). \tag{10}$$

Since the traces in (7) and (8) are always positive, the *CRB of the parametric ML problem will always be tighter than the CRB of its non-parametric counterpart*. However, numerical evaluations of the CRB in realistic situations show that the impact of these trace terms is negligible: the relative difference between the two CRBs is less than 10^{-4} for experimental setups similar to the ones that are used in section VIII. This shows that the impact of the hypothesis of an unstructured $\mathbf{Q}$ is very limited. The difference between the solutions of the parametric and non-parametric ML problems will hardly be noticeable. We can thus safely work under the proposed hypothesis and consider the solutions we will obtain in this framework as the true ML channel estimates. Further in the text, the CRB will be used as a benchmark for the proposed methods. The CRB can simply be evaluated in the non-parametric framework as:

$$\boxed{\boldsymbol{\mathcal{J}}(\mathbf{h})^{-1} = \left(\sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{T}_k \right)^{-1}} \tag{11}$$

Note that this bound depends both on the channel realization (through the covariance matrix $\mathbf{Q}$) and on the chosen training sequences (through the training sequence matrices $\mathbf{T}_k$).

V. Iterative Procedure

When a minimization problem is separable in its variables, a common approach to find the solution is an iterative one. One 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 (see e.g. [16] where this approach is used to jointly estimate the transmitted data symbols and the channel). This procedure converges to a minimum of the cost function. If the starting point is accurate enough, the point of convergence is the global minimum of the cost function. In the sequel, we apply this approach to the likelihood function of the system, which leads to the ML estimate of $\mathbf{Q}$ and $\mathbf{h}$.

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 (5) with respect to $\mathbf{h}$ for a fixed $\mathbf{Q} = \hat{\mathbf{Q}}^i$, i.e. we compute $\hat{\mathbf{h}}^i = \mathbf{h}_{ML}(\hat{\mathbf{Q}}^i)$, where $\mathbf{h}_{ML}(\mathbf{Q}) = \arg\min_{\mathbf{h}} -\mathcal{L}$. The solution to this optimization problem can be computed as:

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

We then seek the covariance matrix $\hat{\mathbf{Q}}^{i+1}$ that minimizes (5) with respect to $\mathbf{Q}$ for a fixed $\mathbf{h} = \hat{\mathbf{h}}^i$: $\hat{\mathbf{Q}}^{i+1} = \mathbf{Q}_{ML}(\hat{\mathbf{h}}^i)$, where $\mathbf{Q}_{ML}(\mathbf{h}) = \arg\min_{\mathbf{Q}} -\mathcal{L}$, the solution to this optimization problem can be computed as (see Appendix A):

$$\mathbf{Q}_{ML}(\mathbf{h}) = K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}_k \mathbf{h}) (\mathbf{u}_k - \mathbf{T}_k \mathbf{h})^H. \tag{13}$$

$\hat{\mathbf{Q}}^{i+1}$ 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 show in [17] that initializing the iterative procedure with a simple Least Squares (LS) channel estimate yields good convergence properties. We thus propose to initialize the iterative ML method with $\hat{\mathbf{Q}}^0 = \mathbf{I}$.

VI. Closed Form Solution

An alternative strategy to the iterative procedure described above consists of directly finding an analytical expression for the global minimum of the likelihood function (5). The separability

property of the cost function can be exploited again in order to find this global minimum. The idea is to analytically minimize the cost function with respect to one variable. This minimum is a function of the other variable. The first variable can then be eliminated in the original cost function, which then becomes a single variable expression. When the problem is separable in its two variables, minimizing this new expression of the cost function with respect to the only variable left yields the global minimum (see e.g. [18]). At this point, we need to distinguish between the constant training sequence case and the changing training sequence case.

A. Constant Training Sequence

In order to indicate that the training sequence after each block is the same, we simply omit the block index (subscript k) for the vector $\mathbf{t}$ and the matrix $\mathbf{T}$. We first minimize the cost function with respect to $\mathbf{Q}$, the solution of which is given by (13) with a constant $\mathbf{T}$ matrix:

$$\mathbf{Q}_{ML}(\mathbf{h}) = K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}\mathbf{h}) (\mathbf{u}_k - \mathbf{T}\mathbf{h})^H. \tag{14}$$

Replacing $\mathbf{Q}$ by $\mathbf{Q}_{ML}(\mathbf{h})$ in the likelihood function as expressed in (5) yields the following:

$$-\mathcal{L} = K \mathrm{tr}\,(\mathbf{I}) + K \ln \left| K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}\mathbf{h}) (\mathbf{u}_k - \mathbf{T}\mathbf{h})^H \right| + \mathrm{cst}. \tag{15}$$

The ML channel estimate $\mathbf{h}_{ML}$ minimizes (15) with respect to $\mathbf{h}$:

$$\mathbf{h}_{ML} = \arg \min_{\mathbf{h}} \left| K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}\mathbf{h}) (\mathbf{u}_k - \mathbf{T}\mathbf{h})^H \right|. \tag{16}$$

Define

$$\hat{\mathbf{R}} \triangleq K^{-1} \sum_{k=1}^{K} \mathbf{u}_k \mathbf{u}_k^H, \quad \bar{\mathbf{u}} \triangleq K^{-1} \sum_{k=1}^{K} \mathbf{u}_k, \quad \hat{\mathbf{Q}} \triangleq \hat{\mathbf{R}} - \bar{\mathbf{u}}\bar{\mathbf{u}}^H, \tag{17}$$

where $\hat{\mathbf{Q}}$ is assumed to be positive definite (a necessary condition for this to hold is $K \geqslant n_t + L$). Using these definitions, the matrix in the minimization problem (16) can be re-expressed as:

$$K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}\mathbf{h}) (\mathbf{u}_k - \mathbf{T}\mathbf{h})^H = \hat{\mathbf{Q}} \left(\mathbf{I} + \hat{\mathbf{Q}}^{-1} (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}) (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}})^H \right).$$

Keeping in mind that $\hat{\mathbf{Q}}$ is positive definite, our minimization problem (16) is thus equivalent to:

$$\mathbf{h}_{ML} = \arg\min_{\mathbf{h}} \left| \mathbf{I} + \hat{\mathbf{Q}}^{-1} (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}) (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}})^H \right|. \tag{18}$$

It can be shown that (see Appendix B):

$$\left| \mathbf{I} + \hat{\mathbf{Q}}^{-1} (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}) (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}})^H \right| = 1 + (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}})^H \hat{\mathbf{Q}}^{-1} (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}) .$$

Hence, the minimization problem (18) is equivalent to:

$$\mathbf{h}_{ML} = \arg\min_{\mathbf{h}} (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}})^H \hat{\mathbf{Q}}^{-1} (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}}).$$

The solution is obtained by nulling the partial derivative of this expression with respect to $\mathbf{h}^H$, which yields:

$$\boxed{\mathbf{h}_{ML} = \left(\mathbf{T}^H \hat{\mathbf{Q}}^{-1} \mathbf{T} \right)^{-1} \left(\mathbf{T}^H \hat{\mathbf{Q}}^{-1} \bar{\mathbf{u}} \right).} \tag{19}$$

This ML channel estimate is easy to compute and also intuitively quite appealing for it shows that the ML channel estimate is simply a fit of $\mathbf{T}\mathbf{h}$ to $\bar{\mathbf{u}}$ in a weighted least squares sense.

B. Changing Training Sequence

We proceed in the same way as for the constant training sequence case. First observe that the likelihood function (5) can be expressed as:

$$-\mathcal{L} = K \ln|\mathbf{Q}| + \text{tr} \left(\mathbf{Q}^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}_k \mathbf{h})(\mathbf{u}_k - \mathbf{T}_k \mathbf{h})^H \right). \tag{20}$$

We first minimize this cost function with respect to $\mathbf{Q}$ leading to $\mathbf{Q}_{ML}(\mathbf{h})$ as given by (13). Replacing $\mathbf{Q}$ by $\mathbf{Q}_{ML}(\mathbf{h})$ in (20) leaves us with an expression of the cost function that only depends on $\mathbf{h}$:

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

The ML channel estimate is thus computed as:

$$\mathbf{h}_{ML} = \arg\min_{\mathbf{h}} \left| \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}_k \mathbf{h}) (\mathbf{u}_k - \mathbf{T}_k \mathbf{h})^H \right|. \tag{21}$$

Although this problem seems similar to (16), the varying $\mathbf{T}_k$ forces us to adopt a different approach which will only lead us to an approximate solution. Let us first introduce the following notations:

$$\mathbf{h}_{LS} \triangleq \left(\sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{T}_k\right)^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{u}_k, \mathbf{g} \triangleq \mathbf{h} - \mathbf{h}_{LS}, \mathbf{g}_{ML} \triangleq \mathbf{h}_{ML} - \mathbf{h}_{LS}, \mathbf{e}_k \triangleq \mathbf{u}_k - \mathbf{T}_k \mathbf{h}_{LS},$$
$$\hat{\mathbf{Q}}' \triangleq K^{-1} \sum_{k=1}^{K} \mathbf{e}_k \mathbf{e}_k^H. \tag{22}$$

where $\hat{\mathbf{Q}}'$ is assumed to be positive definite (a necessary condition therefore is $K \geqslant n_t + L$). Using these notations, the minimization problem (21) can be rephrased as:

$$\mathbf{g}_{ML} = \arg\min_{\mathbf{g}} \left| \sum_{k=1}^{K} (\mathbf{e}_k - \mathbf{T}_k \mathbf{g}) (\mathbf{e}_k - \mathbf{T}_k \mathbf{g})^H \right|. \tag{23}$$

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

$$\left| \mathbf{I} + \hat{\mathbf{Q}}'^{-1} K^{-1} \sum_{k=1}^{K} \left[\mathbf{T}_k \mathbf{g}\mathbf{g}^H \mathbf{T}_k^H - \mathbf{T}_k \mathbf{g}\mathbf{e}_k^H - \mathbf{e}_k \mathbf{g}^H \mathbf{T}_k^H \right] \right|. \tag{24}$$

When K is large, both $\mathbf{h}_{LS}$ and $\mathbf{h}_{ML}$ are close to the true $\mathbf{h}$ (this hypothesis is confirmed by the experimental results presented in section VIII). We can thus assume that $\mathbf{g}$, and consequently the second term in (24), is small in the vicinity of the solution. It is well known that, for $||\mathbf{\Delta}|| \ll 1$, $|\mathbf{I} + \mathbf{\Delta}| \approx 1 + \mathrm{tr}(\mathbf{\Delta})$. Hence, for $K \gg 1$, the minimization problem in (23) can be approximated by:

$$\hat{\mathbf{g}}_{ML} = \arg\min_{\mathbf{g}} \mathrm{tr} \left(\hat{\mathbf{Q}}'^{-1} \sum_{k=1}^{K} \left[\mathbf{T}_k \mathbf{g}\mathbf{g}^H \mathbf{T}_k^H - \mathbf{T}_k \mathbf{g}\mathbf{e}_k^H - \mathbf{e}_k \mathbf{g}^H \mathbf{T}_k^H \right] \right),$$

where $\hat{\mathbf{g}}_{ML}$ is an approximation of $\mathbf{g}_{ML}$. Exploiting the permutation property of the trace of a product, this problem can be rephrased as:

$$\hat{\mathbf{g}}_{ML} = \arg\min_{\mathbf{g}} \sum_{k=1}^{K} \left[\mathbf{g}^H \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \mathbf{g} - \mathbf{e}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \mathbf{g} - \mathbf{g}^H \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{e}_k \right]. \tag{25}$$

The solution to this minimization problem is obtained by nulling the partial derivate of this expression with respect to $\mathbf{g}^H$, and is given as:

$$\hat{\mathbf{g}}_{ML} = \left(\sum_{k=1}^{K} \mathbf{T}_k{}^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \right)^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{e}_k$$

We know from (22) that $\mathbf{h} = \mathbf{g} + \mathbf{h}_{LS}$. If we additionally replace $\mathbf{e}_k$ by $\mathbf{u}_k - \mathbf{T}_k\mathbf{h}_{LS}$, we obtain the following approximation $\hat{\mathbf{h}}_{ML}$ of the channel estimate $\mathbf{h}_{ML}$:

$$\hat{\mathbf{h}}_{ML} = \mathbf{h}_{LS} + \left(\sum_{k=1}^{K} {\mathbf{T}_k}^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k\right)^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{u}_k - \left(\sum_{k=1}^{K} {\mathbf{T}_k}^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k\right)^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \mathbf{h}_{LS},$$

which effectively simplifies to:

$$\boxed{\hat{\mathbf{h}}_{ML} = \left(\sum_{k=1}^{K} {\mathbf{T}_k}^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k\right)^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{u}_k.} \tag{26}$$

VII. Asymptotic Properties of the Closed Form Channel Estimates

In this section, we study the asymptotic properties of the proposed closed form (approximate) ML channel estimates, that is their properties when the number of transmitted data blocks, K, is large.

A. Constant training

Let us first note that $\hat{\mathbf{Q}}$ can be rewritten as:

$$\begin{aligned} \hat{\mathbf{Q}} &= \hat{\mathbf{R}} - \bar{\mathbf{u}}\bar{\mathbf{u}}^H \\ &= K^{-1} \sum_{k=1}^{K} \mathbf{u}_k \mathbf{u}_k^H - K^{-2} \sum_{k=1}^{K} \mathbf{u}_k \sum_{k=1}^{K} \mathbf{u}_k^H. \end{aligned}$$

Keeping in mind that $\mathbf{u}_k = \mathbf{T}\mathbf{h} + \boldsymbol{\epsilon}_k$, we have:

$$\hat{\mathbf{Q}} = K^{-1} \sum_{k=1}^{K} \boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^H - K^{-2} \sum_{i,j=1}^{K} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_j^H. \tag{27}$$

Using the central limit theorem, the above time averages can be replaced by their expected values when K tends to infinity:

$$\begin{aligned} \lim_{K\to\infty} K^{-1} \sum_{k=1}^{K} \boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^H &= \mathrm{E}\left\{\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^H\right\} = \mathbf{Q}, \\ \lim_{K\to\infty} K^{-2} \sum_{i,j=1}^{K} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_j^H &= \mathbf{0}. \end{aligned}$$

Therefore,

$$\lim_{K\to\infty} \hat{\mathbf{Q}} = \mathbf{Q}. \tag{28}$$

Since $\mathbf{Q}$ has a finite non-zero energy, $\hat{\mathbf{Q}}$ can be replaced by its limit $\mathbf{Q}$ whenever a limit including $\hat{\mathbf{Q}}$ is computed. It follows that the ML channel estimate (19) is *consistent*:

$$\begin{aligned}\lim_{K\to\infty} \mathbf{h}_{ML} &= \lim_{K\to\infty} K^{-1} \sum_{k=1}^{K} \left(\mathbf{T}^H \hat{\mathbf{Q}}^{-1} \mathbf{T}\right)^{-1} \mathbf{T}^H \hat{\mathbf{Q}}^{-1} \mathbf{u}_k \\ &= \mathbf{h} + \left(\mathbf{T}^H \mathbf{Q}^{-1} \mathbf{T}\right)^{-1} \mathbf{T}^H \mathbf{Q}^{-1} \mathrm{E}\left\{\boldsymbol{\epsilon}_k\right\} \\ &= \mathbf{h}. \end{aligned} \tag{29}$$

Relying on these results, we can prove the *asymptotic efficiency* of the proposed channel estimate:

$$\begin{aligned}&\lim_{K\to\infty} \mathrm{E}\left\{K(\mathbf{h}_{ML}-\mathbf{h})(\mathbf{h}_{ML}-\mathbf{h})^H\right\} \\ &\quad= \mathrm{E}\left\{\lim_{K\to\infty} \left(\mathbf{T}^H \hat{\mathbf{Q}}^{-1} \mathbf{T}\right)^{-1} \mathbf{T}^H \hat{\mathbf{Q}}^{-1} K^{-1} \sum_{i=1}^{K} \boldsymbol{\epsilon}_i \sum_{j=1}^{K} \boldsymbol{\epsilon}_j^H \hat{\mathbf{Q}}^{-1} \mathbf{T} \left(\mathbf{T}^H \hat{\mathbf{Q}}^{-1} \mathbf{T}\right)^{-1}\right\} \\ &\quad= \left(\mathbf{T}^H \mathbf{Q}^{-1} \mathbf{T}\right)^{-1} \mathbf{T}^H \mathbf{Q}^{-1} \lim_{K\to\infty} \mathrm{E}\left\{K^{-1} \sum_{i,j=1}^{K} \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_j^H\right\} \mathbf{Q}^{-1} \mathbf{T} \left(\mathbf{T}^H \mathbf{Q}^{-1} \mathbf{T}\right)^{-1} \\ &\quad= \left(\mathbf{T}^H \mathbf{Q}^{-1} \mathbf{T}\right)^{-1} \mathbf{T}^H \mathbf{Q}^{-1} \sum_{j=1}^{K} \mathrm{E}\left\{\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_j^H\right\} \mathbf{Q}^{-1} \mathbf{T} \left(\mathbf{T}^H \mathbf{Q}^{-1} \mathbf{T}\right)^{-1} \\ &\quad= \left(\mathbf{T}^H \mathbf{Q}^{-1} \mathbf{T}\right)^{-1}. \end{aligned} \tag{30}$$

This last expression is equal to the normalized CRB (see (11)).

B. Changing training

First note that it is possible to show (the derivation is detailled in appendix C):

$$\lim_{K\to\infty} \hat{\mathbf{Q}}' = \mathbf{Q} \tag{31}$$

Replacing $\mathbf{u}_k$ by its equivalent $\mathbf{T}_k \mathbf{h} + \boldsymbol{\epsilon}_k$ in (26) yields:

$$\hat{\mathbf{h}}_{ML} = \mathbf{h} + \left(\sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k\right)^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \boldsymbol{\epsilon}_k. \tag{32}$$

If the training sequences have a constant non-zero energy, it is clear that

$$\lim_{K\to\infty} K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k = \lim_{K\to\infty} K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{T}_k \neq \mathbf{0}.$$

Furthermore,

$$\lim_{K\to\infty} K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}^{-1} \boldsymbol{\epsilon}_k = \mathrm{E}\left\{ \mathbf{T}_k^H \mathbf{Q}^{-1} \boldsymbol{\epsilon}_k \right\}.$$

Since $\mathbf{T}_k$ and $\boldsymbol{\epsilon}_k$ are independent, the expected value of the product is the product of the expected values. Since $\mathrm{E}\{\mathbf{T}_k\}$ is bounded and $\mathrm{E}\{\boldsymbol{\epsilon}_k\} = \mathbf{0}$, the limit is equal to zero. This shows that the approximate ML channel estimate is *consistent*:

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

Here also, it is possible to derive an expression for the expected value of the normalized covariance matrix as K tends to infinity, showing the *asymptotic efficiency* of the proposed channel estimate:

$$\begin{aligned}
&\lim_{K\to\infty} \mathrm{E}\left\{ K \left(\hat{\mathbf{h}}_{ML} - \mathbf{h} \right) \left(\hat{\mathbf{h}}_{ML} - \mathbf{h} \right)^H \right\} \\
&= \mathrm{E}\left\{ \lim_{K\to\infty} \left(K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \right)^{-1} K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \boldsymbol{\epsilon}_k \sum_{k=1}^{K} \boldsymbol{\epsilon}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \left(K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \right)^{-1} \right\} \\
&= \lim_{K\to\infty} \left(K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{T}_k \right)^{-1} \mathrm{E}\left\{ K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \boldsymbol{\epsilon}_k \sum_{k=1}^{K} \boldsymbol{\epsilon}_k^H \mathbf{Q}^{-1} \mathbf{T}_k \right\} \left(K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{T}_k \right)^{-1} \\
&= \lim_{K\to\infty} \left(K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{T}_k \right)^{-1} \mathrm{E}\left\{ \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{T}_k \right\} \left(K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{T}_k \right)^{-1} \\
&= \lim_{K\to\infty} \left(K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{T}_k \right)^{-1} K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{T}_k \left(K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{T}_k \right)^{-1} \\
&= \lim_{K\to\infty} \left(K^{-1} \sum_{k=1}^{K} \mathbf{T}_k^H \mathbf{Q}^{-1} \mathbf{T}_k \right)^{-1}
\end{aligned}$$

The above expression is equal to the normalized CRB (see (11)).

VIII. Experimental Results

The performance metric that is used throughout this section is the Normalized Mean Square Error (NMSE) of the proposed channel estimate:

$$\mathrm{NMSE} = E\left\{ \frac{||\hat{\mathbf{h}} - \mathbf{h}||^2}{||\mathbf{h}||^2} \right\}.$$

The results that are presented are obtained with the closed form channel estimates (19) and (26). When the iterative method results are investigated, we explicitly state it in the text. We use the CRB as a benchmark in the experiments. The CRB curves displayed on the graphs represent the NMSE of an estimator that achieves the CRB, which is $E\left\{\frac{\mathrm{tr}(\boldsymbol{\mathcal{J}}^{-1})}{||\mathbf{h}||^2}\right\}$. The experiments are performed on convolutive Rayleigh fading channels of varying order L. The different channel taps are independent identically distributed Gaussian random variables. The training and data sequences are white QPSK sequences. The energy of the transmitted symbols (both data and training) is set to λ^2. 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 in the changing training sequence case and over 100 different training sequences in the constant training sequence case. Note that this averaging is also done for the CRB results since the CRB depends both on the channel realization and the training sequences. The Signal to Noise Ratio (SNR) is defined as $SNR = E\left\{||\mathbf{h}||^2\right\}\frac{\lambda^2}{\sigma^2}$.

A. Performance of the Proposed Method

In this section, we analyze and compare the algorithms proposed for the two situations that have been considered throughout this article: the constant and changing training sequence cases.

1) Comparison of the Cramer-Rao Bounds : To have a first insight in how these compare, we check the CRB performance for these two configurations. We consider a transmission scheme where the length of the training sequences is set to $n_t = 10$ and the number of observed training sequences is set to $K = 100$. The CRB for different channel orders in that context is presented in Fig. 1. We see that the use of changing training sequences systematically results in a reduced CRB for all channel orders. When the channel order is small, the CRB keeps decreasing with a constant slope as the SNR increases in both contexts, but the use of changing training sequences always yields improved performance. For large channel orders, the CRB saturates at high SNR in both contexts and the changing training curve again shows a better performance. For intermediate channel orders, we observe a saturation in the CRB curve when constant training is used whereas

the use of changing training yields a constant slope in the CRB curve. It is possible to show that the CRB decreases with a constant slope as the SNR increases when there is an exact solution to the channel identification problem in the noiseless case. The saturation effect in the CRB appears when there is no exact solution to the channel identification problem in the noiseless case. Observe that the number of received symbols that do not contain contributions from the unknown data symbols is equal to $n_t - L$. When constant training sequences are used, the channel identification problem in the noiseless case has an exact solution if there are at least $L+1$ such received symbols. That is when $n_t \geqslant 2L+1$. When changing training sequences are used, an exact solution exists in the noiseless case as soon as there is one such received symbol per transmitted training sequence, that is when $n_t \geqslant L+1$. When n_t is fixed and the channel order is in the interval $\frac{n_t-1}{2} \leqslant L \leqslant n_t - 1$, using changing training sequences will yield a constant slope in the CRB for increasing SNR whereas a floor will appear at high SNR if a constant training sequence is used. For channel orders outside this interval, both methods show similar behaviors (constant slope for small channel orders and saturation for large channel orders), but there is still an advantage in using changing training sequences.

2) Changing Training Sequences: After this discussion on the CRB, we check how the proposed closed form channel estimates match this theoretical bound. We first check it for the closed-form ML channel estimate proposed in the context of changing training sequences. In Fig. 2, we compare the simulated performance of our method with the corresponding CRB as a function of the SNR. We perform this comparison for two different channel orders: one for which the CRB has a constant slope, the other being large enough to have the CRB saturating at high SNR. We repeat these experiments for two different values of K: a relatively small one and a larger one. We observe a relatively good match between the CRB and the experimental curves when the channel order is large and there is a floor in the CRB. The match is better for a larger K. When the channel order is small and there is no floor in the CRB, the theoretical and experimental curves match quite well at low SNR but we see the emergence of a floor on

the experimental NMSE for higher SNR. The value of this floor decreases as the number of data blocks, K increases. In Fig. 3, we evaluate the impact of the number of data blocks K on the channel estimate NMSE. Here again, the simulations are done for two different channel orders. We further test the impact of the factor K for two different values of the SNR. Here again, we see that when the channel order is large, there is a relatively good match between the theoretical and experimental curves. When the channel order is smaller, there is a big difference between the theoretical and experimental values for small values of K when the SNR is large, which corresponds to the zone where the saturation occurs in Fig. 2. However, as K increases, the experimental NMSE and the CRB tend to merge. This difference between the CRB and the experimental results originates from the approximations we made in order to derive the approximate closed form ML channel estimate. These approximations do not hold when the SNR is large, K is small and the channel order is small However, when the iterative method is used, the channel estimate converges to $\mathbf{h}_{ML}$. The experiments presented in Fig. 4 and Fig. 5 show that the gap between the CRB and the closed form estimate is closed after a few iterations. Hence, performing a few iterations allows us to avoid the saturation effect in the NMSE when the SNR is large and there are only a few training sequences available to perform channel estimation (small K).

3) Constant Training Sequences: In Fig. 6 and Fig. 7, we perform a similar analysis for the closed form ML channel estimate in the context of a constant training sequence. The figures show us that there is no significant difference between the CRB and the experimental results, except for very small values of K. This improved match between the CRB and the experimental results originates in the fact that we did not need to make any approximation when deriving the expression of the closed-form ML channel estimate in this case. Note that there is no point in using the iterative method in this context, since the closed form channel estimate corresponds to its convergence point, which is confirmed by experimental results (not shown in the figures).

B. Comparison with Traditional ML Methods

Classical training-based ML channel estimation techniques solely rely on the part of the received symbols that only contain contributions from the known training symbols. They simply discard the received samples that are corrupted by contributions from the unknown data symbols. Such symbols can be observed at the receiver only when $n_t \geqslant L+1$. In that case, based on the data model we derived in section II, we can derive a data model that focuses on the received symbols that do not contain any contributions from the unknown data symbols. Define $\mathbf{u}'_k = \mathbf{u}_k(L+1:n_t)$ and $\mathbf{T}'_k = \mathbf{T}_k(L+1:nt,:)$. We then have: $\mathbf{u}'_k = \mathbf{T}'_k\mathbf{h} + \boldsymbol{\eta}'_k$, where $\boldsymbol{\eta}'_k$ 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 column and $[t_k[L+1], \ldots, t_k[1]]$ as first row. When the noise term is not colored, the solution of the ML channel identification problem is well known to be a simple least squares fit of $\mathbf{T}'_k\mathbf{h}$ to $\mathbf{u}'_k$. In the changing training sequences case, the classical ML channel estimate is thus given by $\mathbf{h}'_{ML} = \left(\sum_{k=1}^{K} \mathbf{T}'^H_k \mathbf{T}'_k\right)^{-1} \sum_{k=1}^{K} \mathbf{T}'_k \mathbf{u}'_k$ Relying on the randomness of the training sequences, the inverse of the summation in this expression will exist with probability one as soon as $K \geqslant L+1$. When a constant training sequence is used the condition on the channel order is more stringent. If we want to have a solution to our LS problem, we need to have as many equations as there are coefficients to identify in the channel model. This means that we need $\mathbf{u}'_k$ to have at least dimension $L+1$. This only happens when $n_t \geqslant 2L+1$. In that case, the constant $\mathbf{T}'$ matrix is tall and we can find a LS solution to derive our ML channel estimate: $\mathbf{h}'_{ML} = K^{-1}\left(\mathbf{T}'^H\mathbf{T}'\right)^{-1} \sum_{k=1}^{K} \mathbf{T}'\mathbf{u}'_k$. Note that the conditions that relate the training sequences length n_t to the channel order L are quite stringent: it is simply impossible to identify the channel when these are not fulfilled. We can now compare the results obtained with these classical ML channel estimates with the proposed ML estimates. In Fig. 8, we consider the constant training sequence case and we analyze the changing training sequence case in Fig. 9. We compare the results for different channel orders when the length of the training sequences n_t is fixed. In both situations, we see that the new method and the traditional one yield equivalent performance

for small channel orders. When the channel order increases, the new method outperforms the classical one, especially at low SNR. The only situation where this is not the case is for large SNR values when changing training sequences are used. We know however that increasing the number of observed blocks K or performing a few iterations would solve this problem. When the channel order keeps growing, the new method still provides reliable channel estimates whilst traditional methods cannot be applied anymore.

IX. Conclusions

In this paper, we presented a new training based ML channel identification method where the training sequences can be shorter than the channel impulse response length. We analyzed two situations: the situation where the same training sequence is repeated at the end of each data block (constant training sequence case) and the situation where this training sequence is changed at the end of each data block (changing training sequence case). We first proposed an iterative ML method and then derived approximate closed form expressions for the ML channel estimates. The proposed closed-form expressions have a low complexity and effectively achieve the CRB in most practical situations. In the few situations where the CRB is not achieved (i.e. low channel order, large SNR and small number of training sequences), the iterative method can be used and will achieve the CRB in a few iterations. The proposed method can be used in white as well as in colored noise conditions. The proposed method clearly outperforms classical ML training-based methods. It outperforms as well existing semi-blind methods (none of them achieves the CRB).

References

[1] Lang Tong ans Sylvie Perreau, "Multichannel Blind Identification: from Subspace to Maximum Likelihood Methods," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.

[2] S. Talwar and A. Paulraj, "Blind Estimation of Multiple Co-Channel Digital Signals Received at an Antenna Array," in *proc. of the fifth Annual IEEE Dual-Use Technologies and Applications Conference*, May 1995.

[3] G. Leus, P. Vandaele and M. Moonen, "Deterministic Blind Modulation-Induced Source Separation for Digital Wireless Communications," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 219–227, Jan. 2001.

[4] 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.

[5] C. Fragouli, N. Al-Dhahir and W. Turin, "Training-Based Channel Estimation for Multiple-Antenna Broadband Transmissions," *IEEE Transactions on Wireless Communications*, vol. 2, no. 2, pp. 384–391, March 2003.

[6] Jonathan. H. Manton, "On Optimal Channel Identification by Use of Training Sequences and Pilot Tones," in *Proc. of the sixth International Symposium on Signal Processing and its Applications (ISSPA 2001)*, Kuala-Lumpur, Malaysia, 2001.

[7] 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.

[8] 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.

[9] 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.

[10] J. Ayadi, E. de Carvalho and D. T. M. Slock, "Blind and Semi-Blind Maximum Likelihood Methods for FIR Multichannel Identification," in *Proc. of the International Conference on Acoustics, Speech, and Signal Processing (ICCASP 98)*, Seattle, May 1998, pp. 3185–3188.

[11] E. de Carvalho and D. T. M. Slock, "Cramer-Rao Bounds for Semi-Blind, Blind and Training Sequence Based Channel Estimation," in *Proc. of the first IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC 97)*, Paris, France, April 1997, pp. 129–132.

[12] H. Trigui and D. T. M. Slock, "Optimal and Suboptimal Approaches for Training Sequence Based Spatio-Temporal Channel Identification in Coloured Noise," .

[13] H. A. Cipran and M. K. Tsatsanis, "Stochastic Maximum Likelihood Methods for Semi-Blind Channel Estimation," *IEEE Signal Processing Letters*, vol. 5, no. 1, pp. 21–24, January 1998.

[14] E. de Carvalho and D. T. M. Slock, "Maximum Likelihood Blind FIR Multi-Channel Estimation with Gaussian Prior for the Symbols," in *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP 97)*, Munich, Germany, April 1997, pp. 3593–3596.

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

[16] O. Rousseaux, G. Leus, M. Moonen, "A suboptimal Iterative Method for Maximum Likelihood Sequence Estimation in a Multipath Context," *EURASIP Journal on Applied Signal Processing (JASP)*, vol. 2002, no. 12, pp. 1437–1447, December 2002.

[17] 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.

[18] 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. 5, pp. 1184–1197, May 1996.

APPENDIX A

In this appendix, we show that

$$\mathbf{Q}_{ML}(\mathbf{h}) = K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}_k \mathbf{h}) (\mathbf{u}_k - \mathbf{T}_k \mathbf{h})^H. \tag{33}$$

The proof is adapted from [15, pp.201-202]. Let us first define the sample covariance matrix:

$$\mathbf{R}_K(\mathbf{h}) \triangleq K^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}_k \mathbf{h}) (\mathbf{u}_k - \mathbf{T}_k \mathbf{h})^H.$$

Using this definition, the log-likelihood function (5) can be re-expressed as

$$-\mathcal{L} = \frac{K}{2} \left(\text{tr} \left(\mathbf{R}_K(\mathbf{h}) \mathbf{Q}^{-1} \right) + \ln|\mathbf{Q}| \right) + \text{cst} \tag{34}$$

The proposed ML estimate defined in (33) is equivalent to $\mathbf{Q}_{ML}(\mathbf{h}) = \mathbf{R}_K(\mathbf{h})$. Claiming that (34) is minimized with respect to $\mathbf{Q}$ for $\mathbf{Q} = \mathbf{R}_K(\mathbf{h})$ is equivalent to claiming that

$$\text{tr} \left(\mathbf{R}\mathbf{Q}^{-1} \right) + \ln|\mathbf{Q}| \geqslant \text{tr} \left(\mathbf{R}\mathbf{R}^{-1} \right) + \ln|\mathbf{R}|, \qquad \forall \;\; \mathbf{Q} = \mathbf{Q}^H > 0.$$

where $\mathbf{R}$ is a shorthand notation for $\mathbf{R}_K(\mathbf{h})$. The following equivalences are easily derived:

$$\begin{aligned} & \text{tr} \left(\mathbf{R}\mathbf{Q}^{-1} \right) + \ln|\mathbf{Q}| \geqslant \text{tr} \left(\mathbf{I}_{n_t+L} \right) + \ln|\mathbf{R}| \\ \Leftrightarrow \quad & \text{tr} \left(\mathbf{R}\mathbf{Q}^{-1} \right) + \ln \left(|\mathbf{Q}|/|\mathbf{R}| \right) \geqslant n_t + L \\ \Leftrightarrow \quad & \text{tr} \left(\mathbf{R}\mathbf{Q}^{-1} \right) - \ln|\mathbf{R}\mathbf{Q}^{-1}| \geqslant n_t + L. \end{aligned} \tag{35}$$

It is clear from its definition that $\mathbf{R}$ can be factorized in a full rank square matrix and its complex conjugate transpose: $\mathbf{R} = \mathbf{G}\mathbf{G}^H$. Define next the matrix $\mathbf{F} \triangleq \mathbf{G}^H \mathbf{Q}^{-1} \mathbf{G}$. This matrix $\mathbf{F}$ is symmetric and positive definite. Its eigenvalues $\lambda_1 \ldots \lambda_{n_t+L}$ clearly satisfy $\lambda_i > 0$. Using these new definitions, we can proceed and rephrase (35) as:

$$
\begin{aligned}
& \text{tr}\left(\mathbf{G}\mathbf{G}^H\mathbf{Q}^{-1}\right) - \ln|\mathbf{G}\mathbf{G}^H\mathbf{Q}^{-1}| \geqslant n_t + L \\
\Leftrightarrow \quad & \text{tr}\left(\mathbf{G}^H\mathbf{Q}^{-1}\mathbf{G}\right) - \ln|\mathbf{G}^H\mathbf{Q}^{-1}\mathbf{G}| \geqslant n_t + L \\
\Leftrightarrow \quad & \text{tr}(\mathbf{F}) - \ln|\mathbf{F}| \geqslant n_t + L \\
\Leftrightarrow \quad & \sum_{i=1}^{n_t+L} \lambda_i - \ln \prod_{i=1}^{n_t+L} \lambda_i \geqslant n_t + L \\
\Leftrightarrow \quad & \sum_{i=1}^{n_t+L} \lambda_i - \ln\lambda_i - 1 \geqslant 0
\end{aligned}
$$

Since $\forall\lambda > 0, \quad ln\lambda \leqslant \lambda - 1$, the last inequality is obviously true. This concludes the proof.

APPENDIX B

In this appendix, we show that

$$\left|\mathbf{I}+\hat{\mathbf{Q}}^{-1}\left(\mathbf{Th}-\bar{\mathbf{u}}\right)\left(\mathbf{Th}-\bar{\mathbf{u}}\right)^{H}\right| = 1+\left(\mathbf{Th}-\bar{\mathbf{u}}\right)^{H}\hat{\mathbf{Q}}^{-1}\left(\mathbf{Th}-\bar{\mathbf{u}}\right).$$

Let $\mathbf{x}$ be a $n\times 1$ column vector and $\mathbf{y}$ a row vector of the same length and define the rank-1 matrix $\mathbf{A}=\mathbf{xy}$ of size $n\times n$.

Since $\mathbf{A}$ has rank 1, it has a single non-zero eigenvalue, λ_1. Moreover, its zero eigenspace has dimension $n-1$, and the characteristic polynomial is divisible by t^{n-1}. Since λ_1 is a root of the characteristic polynomial, we can express it as $t^n-\lambda_1 t^{n-1}$.

Observing that $\mathbf{Ax}=(\mathbf{xy})\mathbf{x}=\mathbf{x}(\mathbf{yx})$, shows that $\mathbf{x}$ is an eigenvector for $\mathbf{A}$ with eigenvalue $\lambda_1=\mathbf{yx}$. The characteristic polynomial can thus be developed into:

$$|t\mathbf{I}-\mathbf{A}| = t^n-\mathbf{yx}t^{n-1}.$$

Taking $t=-1$ yields

$$\begin{aligned} & |-\mathbf{I}-\mathbf{xy}| = (-1)^n(1+\mathbf{yx}) \\ \Leftrightarrow \quad & (-1)^n|\mathbf{I}+\mathbf{xy}| = (-1)^n(1+\mathbf{yx}) \\ \Leftrightarrow \quad & |\mathbf{I}+\mathbf{xy}| = 1+\mathbf{yx}. \end{aligned}$$

The sought property results from the following choices for $\mathbf{x}$ and $\mathbf{y}$: $\mathbf{x}=\hat{\mathbf{Q}}^{-1}\left(\mathbf{Th}-\bar{\mathbf{u}}\right)$ and $\mathbf{y}=\left(\mathbf{Th}-\bar{\mathbf{u}}\right)^{H}$.

APPENDIX C

In this appendix, we prove that

$$\lim_{K\to\infty} \hat{\mathbf{Q}}' = \mathbf{Q}$$

Let us first note that $\mathbf{e}_k$ can be rewritten as

$$\begin{aligned}
\mathbf{e}_k &= \mathbf{u}_k - \mathbf{T}_k \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{u}_l \\
&= \mathbf{T}_k \mathbf{h} + \boldsymbol{\epsilon}_k - \mathbf{T}_k \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \mathbf{h} + \boldsymbol{\epsilon}_l \right) \\
&= \boldsymbol{\epsilon}_k - \mathbf{T}_k \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \sum_{l=1}^{K} \mathbf{T}_l^H \boldsymbol{\epsilon}_l
\end{aligned}$$

Based on this observation, it is possible to check that the shorthand notation $\hat{\mathbf{Q}}'$ defined in (22) converges to the true $\mathbf{Q}$ when K tends to infinity:

$$\begin{aligned}
\lim_{K\to\infty} \hat{\mathbf{Q}}' &= \lim_{K\to\infty} K^{-1} \sum_{k=1}^{K} \mathbf{e}_k \mathbf{e}_k^H \\
&= \lim_{K\to\infty} K^{-1} \sum_{k=1}^{K} \boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^H - K^{-1} \sum_{k=1}^{K} \boldsymbol{\epsilon}_k \left(\sum_{l=1}^{K} \boldsymbol{\epsilon}_l^H \mathbf{T}_l \right) \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \mathbf{T}_k^H \\
&\quad - K^{-1} \sum_{k=1}^{K} \mathbf{T}_k \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \left(\sum_{l=1}^{K} \mathbf{T}_l^H \boldsymbol{\epsilon}_l \right) \boldsymbol{\epsilon}_k^H \\
&\quad + K^{-1} \sum_{k=1}^{K} \mathbf{T}_k \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \left(\sum_{l=1}^{K} \mathbf{T}_l^H \boldsymbol{\epsilon}_l \right) \left(\sum_{l=1}^{K} \boldsymbol{\epsilon}_l^H \mathbf{T}_l \right) \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \mathbf{T}_k^H .
\end{aligned}$$

Using the central limit theorem, we can replace the limit of $K^{-1} \sum_k$ by the expected value over k, which yields:

$$\begin{aligned}
\lim_{K\to\infty} \hat{\mathbf{Q}}' = \mathbf{Q} - \mathbf{Q}\, \mathrm{E} \left\{ \mathbf{T}_k \lim_{K\to\infty} \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \mathbf{T}_k^H \right\} - \mathrm{E} \left\{ \mathbf{T}_k \lim_{K\to\infty} \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \mathbf{T}_k^H \right\} \mathbf{Q}^H \\
+ \mathrm{E} \left\{ \mathbf{T}_k \lim_{K\to\infty} \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \left(\sum_{l=1}^{K} \mathbf{T}_l \mathbf{Q} \mathbf{T}_l^H \right) \left(\sum_{l=1}^{K} \mathbf{T}_l^H \mathbf{T}_l \right)^{-1} \mathbf{T}_k^H \right\} .
\end{aligned}$$

Since the training sequences have a non-zero energy, the limits present in this last expression are equal to zero. Hence, given the finite norm of $\mathbf{Q}$, all the terms containing these limits are equal to zero, which concludes our proof.

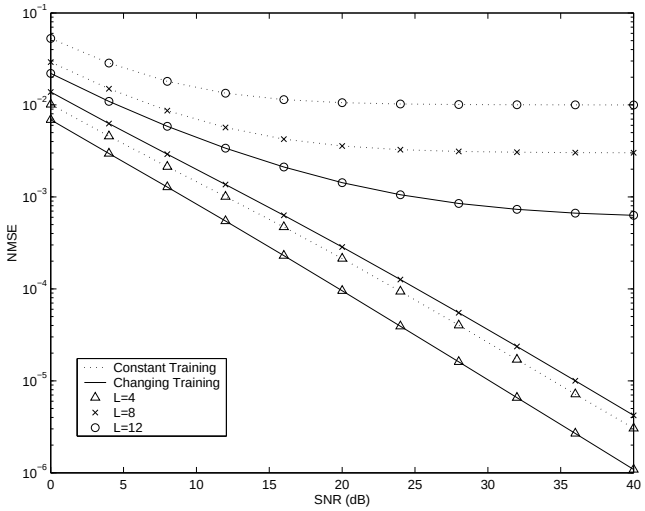

Fig. 1. Comparison of the CRB of the constant and changing training sequence cases vs. the SNR for different channel orders when $n_t = 10$ and $K = 100$.

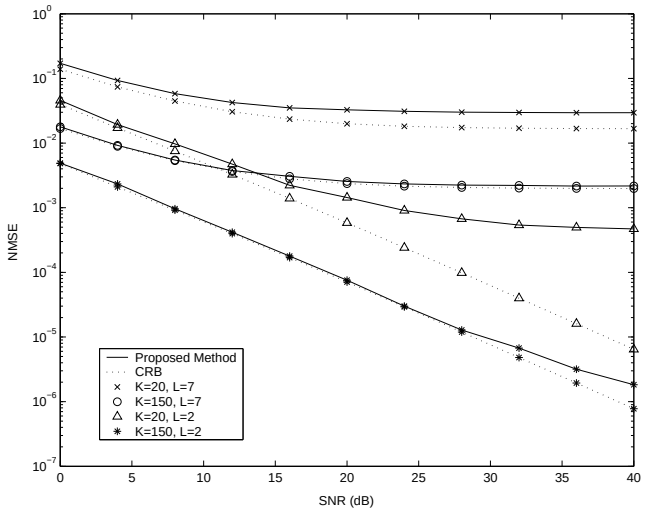

Fig. 2. Comparison of the simulated NMSE and the CRB vs. SNR for different channel orders when $n_t = 5$ and changing training sequences are used. The results are plotted for two different values of K, namely 20 and 150

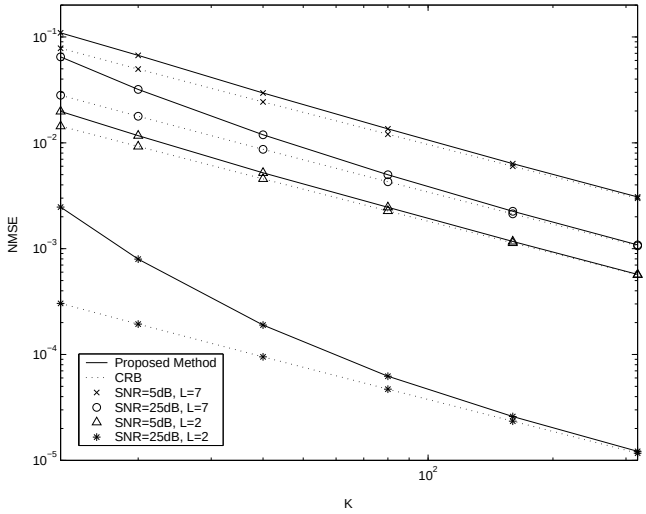

Fig. 3. Comparison of the simulated NMSE and the CRB vs. K for different channel orders when $n_t = 5$ and changing training sequences are used. The results are plotted for two different values of the SNR, namely 5 and $25dB$.

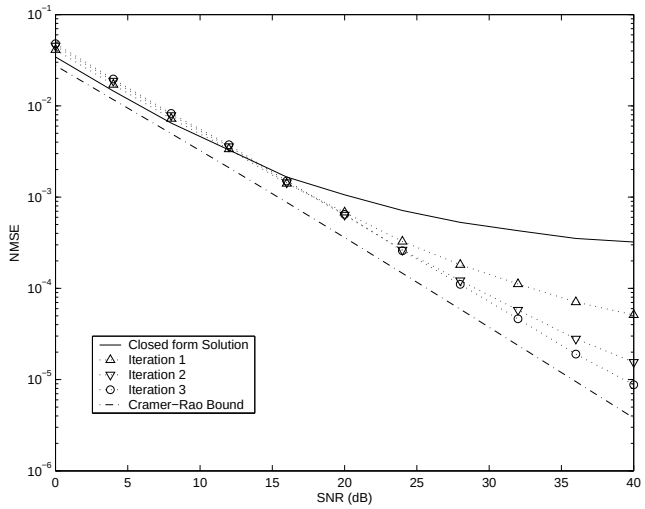

Fig. 4. Comparison of the simulated NMSE and the CRB vs. SNR for $L = 2$, $n_t = 5$ and $K = 20$ using changing training sequences. The NMSE converges to the CRB after a few iterations when the iterative method is used.

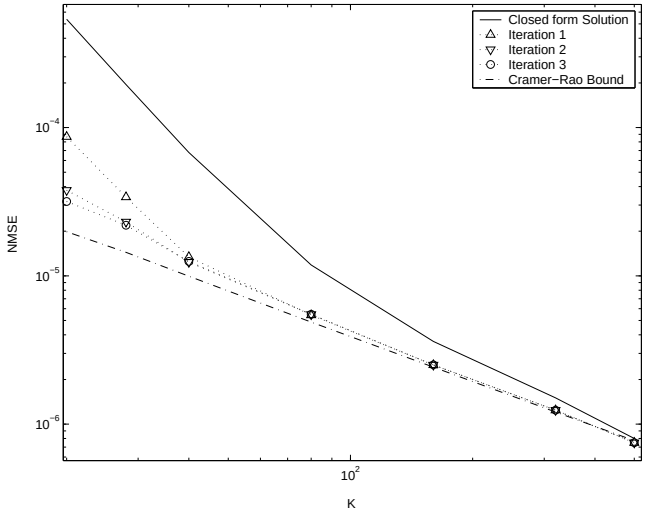

Fig. 5. Comparison of the simulated NMSE and the CRB vs. K for $L = 2$, $n_t = 5$ and SNR=35dB using changing training sequences. The NMSE converges to the CRB after a few iterations when the iterative method is used.

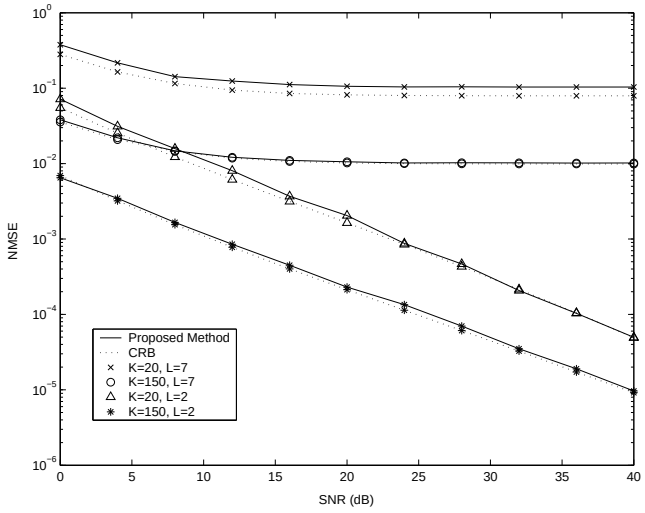

Fig. 6. Comparison of the simulated NMSE and the CRB vs. SNR for different channel orders when $n_t = 5$ and constant training sequences are used. The results are plotted for two different values of K, namely 20 and 150

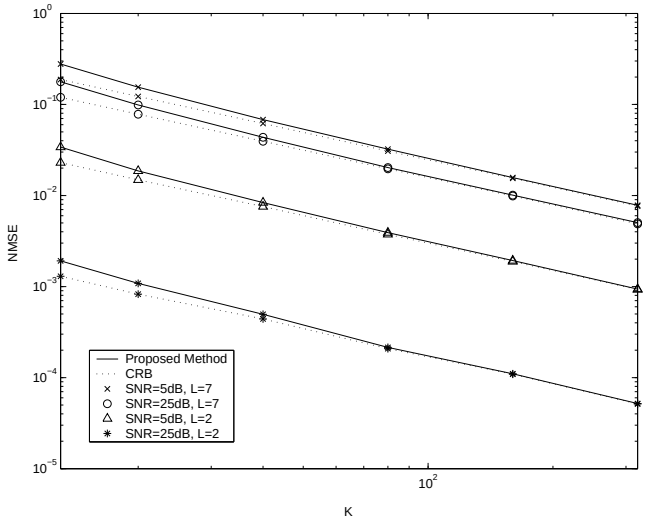

Fig. 7. Comparison of the simulated NMSE and the CRB vs. K for different channel orders when $n_t = 5$ and constant training sequences are used. The results are plotted for two different values of the SNR, namely 5 and $25dB$.

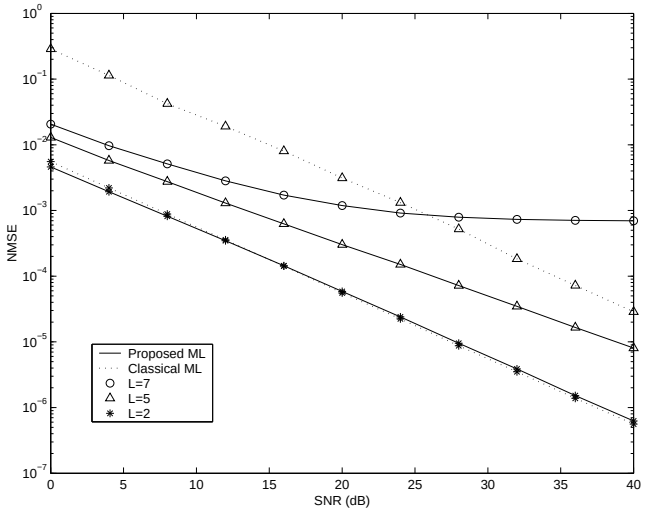

Fig. 8. Simulated NMSE vs. SNR for the proposed ML method and traditional ML channel estimation for different channel orders when constant training sequences of length $n_t = 11$ are used and the number of observed blocks is set to $K = 100$.

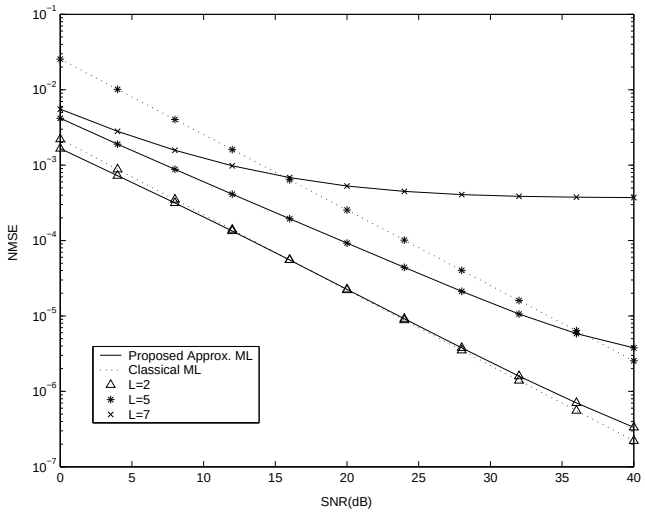

Fig. 9. Simulated NMSE vs. SNR for the proposed ML method and traditional ML channel estimation for different channel orders when changing training sequences of length $n_t = 6$ are used and the number of observed blocks is set to $K = 100$.