Gaussian Maximum-Likelihood Channel Estimation With Short Training Sequences

Olivier Rousseaux, Geert Leus, *Senior Member, IEEE*, Petre Stoica, *Fellow, IEEE*, and Marc Moonen, *Member, IEEE*

***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. The proposed method can be classified as semiblind and exploits all channel-output samples containing contributions from the training sequences (including those containing contributions from the unknown surrounding data symbols). Experimental results show that the proposed method closely approaches the Cramer–Rao bound and outperforms existing training-based methods (which solely exploit the channel-output samples containing contributions from the training sequences only). Existing semiblind ML methods are tested as well and appear to be outperformed by the proposed method in the considered context. A major advantage of the proposed approach is its computational complexity, which is significantly lower than that of existing semiblind methods.**

***Index Terms*—Block transmission, maximum-likelihood (ML) estimation, stationary multipath channel, training sequence.**

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 intersymbol interference (ISI), a phenomenon that needs to be combated 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 three families that are termed "blind," "semiblind," and "training-based."

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 (see, e.g., [1], [2], or [3]). Most existing methods require the training sequences to be significantly longer than the channel impulse response and only exploit the channel-output samples that are not corrupted by unknown data symbols. 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 then a simple least-squares problem.

Blind algorithms, on the other hand, estimate the channel based on the properties of the transmitted signals (finite alphabet properties, higher order statistics, cyclostationarity, see, e.g., [4] or [5], and references therein) and do not exploit the knowledge of possibly inserted pilots or training sequences in the stream of transmitted symbols.

Semiblind techniques have been proposed recently, which show improved performance as they simultaneously exploit the knowledge of the pilots or training symbols and blind properties of the transmitted signals towards the determination of the channel model. These techniques are most useful when the transmitted signals contain a mix of training symbols and unknown data symbols, in which case neither blind nor training-based techniques are optimal. A method that explicitly combines a blind and a training-based cost function was proposed in [6] and seems to be the first suboptimal semiblind method proposed in the literature. Later, this idea has also been adopted in [7] and [8]. A deterministic method exploiting all the energy that is received from the training sequences was presented in [9] and is suited to the case where a constant training sequence, which must be at least as long as the channel order, is periodically inserted in the transmitted sequence. Several semiblind 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 three families of ML methods: deterministic ML, in which the data symbols are considered as deterministic disturbances, Gaussian ML, in which the data symbols are

Manuscript received January 8, 2004; revised November 7, 2004; accepted November 7, 2004. The editor coordinating the review of this paper and approving it for publication is J. Cavers. This research paper was carried out at the ESAT Laboratory of the Katholieke Universiteit Leuven, in the frame of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office 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 G.0196.02 ("Design of efficient communication techniques for wireless time-dispersive multiuser MIMO systems") and was supported in part by the Swedish Science Council.

O. Rousseaux was with Signals, Identification, System Theory and Automation (SISTA), Katholieke Universiteit (K.U.) Leuven, Leuven, Belgium. He is now with the Inter Universiatair Micro Elektronica Centrum (IMEC), 3001 Leuven, Belgium (e-mail: olivier.rousseaux@imec.be).

G. Leus is with the Department of Electrical Engineering, Delft University of Technology, 2628CD Delft, The Netherlands (e-mail: leus@cas.et.tudelft.nl).

P. Stoica is with the Department of Systems and Control, Uppsala University, SE-751 05 Uppsala, Sweden (e-mail: ps@syscon.uu.se).

M. Moonen is with the Department of Electrical Engineering—ESAT, Katholieke Universiteit Leuven, Leuven, Belgium (e-mail: marc.moonen@esat.kuleuven.ac.be).

Digital Object Identifier 10.1109/TWC.2005.858353

assumed to be Gaussian distributed, and stochastic ML, where the true distribution of the data symbols is exploited. Some deterministic ML methods are presented, for instance, in [10]. In [11], a theoretical comparison of the Cramer–Rao bounds (CRBs) indicates that Gaussian ML methods outperform deterministic ML methods. In [12]–[14], Gaussian ML methods are proposed. The method presented in [14] is based on the expectation-maximization (EM) algorithm and will be used as a benchmark in this paper. Stochastic ML methods theoretically perform better as they assume a more realistic distribution for the data symbols, but suffer from a higher complexity. A method based on hidden Markov models and the EM algorithm was presented in [15], but performs significantly worse than the achievable CRB.

When facing a given transmission scheme, the choice between these three families of channel-estimation techniques depends on the way the transmitted data are formatted. When a long preamble of training symbols is inserted at the beginning of a packet of data, the training-based methods are the most attractive, as they generally allow the accurate identification of the channel at a very low computational cost. In this situation, the use of semiblind methods only marginally improves the quality of the channel estimate with a significantly higher computational cost. When there are not much training symbols or no training symbols at all, the channel cannot be estimated accurately relying on the knowledge of these training symbols and one has to rely on blind methods in order to obtain the CSI. When the number of training symbols is significant but not large enough to enable a sufficiently accurate channel estimation, or when the training symbols are not grouped in a long preamble and the classical training-based methods cannot be used efficiently, semiblind methods outperform others as they simultaneously exploit the knowledge of the training symbols and the useful properties of the transmitted signals.

In this paper, we consider the situation where short training sequences (possibly shorter than the channel impulse response length) are repeatedly inserted between blocks of data symbols, which corresponds to known-symbol-padding (KSP) transmission [16] or pilot-symbol-assisted modulation (PSAM) [17]. We consider multipath single-input single-output (SISO) channels with a relatively large coherence time (the channel stays constant during the transmission of several blocks of data). In the presented work, we aim at finding channel-estimation strategies that are suited to this specific context. Given the placement of the training symbols in the considered framework, semiblind channel-estimation strategies seem most appropriate to obtain accurate channel models. We propose a new Gaussian ML method for semiblind channel identification that is well suited to this situation. The proposed method 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, has a small computational complexity, and seems to outperform existing semiblind methods in a similar context. 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), which allows us to use several training sequences to construct the channel estimate. 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 which we have to rely on in order to derive low-complexity ML channel estimates. In Section IV, we show that the proposed approximations have 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 then discuss the identifiability conditions of the proposed methods (Section VI-C). We experimentally test the proposed methods and compare them with existing ML methods in Section VIII, and finally draw a conclusion in Section IX.

Notation: We use upper (lower) case boldface 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 $(\cdot)^*$ denotes the complex conjugate, and $\mathrm{Re}(\cdot)$, the real part, and $\mathrm{Im}(\cdot)$, the imaginary part, of a complex number. The superscript $(\cdot)^{\mathrm{T}}$ denotes the transpose of a matrix and $(\cdot)^{\mathrm{H}}$, the complex conjugate transpose. Finally, $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, $|\cdot|$ its determinant, and $\mathbf{A}(i,j)$ denotes the ith element of the jth column of $\mathbf{A}$.

II. Data Model

We consider a finite impulse response (FIR) convolutive channel of order $L : \mathbf{h} = [h[0] \ \cdots \ h[L]]^{\mathrm{T}}$. A burst $x[n]$, $n = 1,\ldots,N$, of symbols is transmitted over the channel. Considering that the coherence time of the channel is larger than the duration of the transmitted burst, 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 AWGN at the receiver.

A total number of K training sequences is inserted in the burst. The kth training sequence, $\mathbf{t}_k = [t_k[1] \ \cdots \ t_k[n_{\mathrm{t}}]]^{\mathrm{T}}$, starts at position $n_k : [x[n_k] \ \cdots \ x[n_k + n_{\mathrm{t}} - 1]]^{\mathrm{T}} = \mathbf{t}_k$. Two possibilities are considered in the text: either the same training sequence is repeated after each block of data (constant-training-sequence case), or the training sequence is changed after each block (changing-training-sequence case). Define the vector $\mathbf{u}_k$ of received symbols that contain a contribution from the kth transmitted training sequence: $\mathbf{u}_k = [y[n_k] \ \cdots \ y[n_k + n_\mathrm{t} + L - 1]]^\mathrm{T}$. It is the sum of a deterministic and a stochastic term

$$\mathbf{u}_k = \mathbf{T}_k \mathbf{h} + \epsilon_k \tag{2}$$

where $\mathbf{T}_k$ is an $(n_\mathrm{t} + L) \times (L + 1)$ tall Toeplitz matrix with $[\mathbf{t}_k^\mathrm{T} \ 0 \ \cdots \ 0]^\mathrm{T}$ as its first column and $[t_k[1] \ 0 \ \cdots \ 0]$ as its first row. The stochastic term ϵ_k is described as

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

where $\mathbf{s}_k = [s_k[1] \ \cdots \ s_k[2L]]^\mathrm{T} = [x[n_k - L] \ \cdots \ x[n_k - 1] \ x[n_k + n_\mathrm{t}] \ \cdots \ x[n_k + n_\mathrm{t} + L - 1]]^\mathrm{T}$ is the vector of surrounding data symbols, and $\boldsymbol{\eta}_k = [\eta[n_k] \ \ldots \ \eta[n_k + n_\mathrm{t} + L - 1]]^\mathrm{T}$ is the AWGN term. Assuming that both the noise and the data are white and zero mean ($\mathrm{E}\{s_k[i]s_k[j]^*\} = \mathrm{E}\{\eta[i]\eta[j]^*\} = 0, \ \forall i,j,k : i \neq j$, and $\mathrm{E}\{s_k[i]\} = \mathrm{E}\{\eta[k]\} = 0$), we can say that ϵ_k is zero mean. Defining n_s as the length of the shortest sequence of data symbols ($n_s = \min_k\{n_{k+1} - (n_k + n_\mathrm{t} - 1)\}$), we assume $n_s \geqslant 2L$. This ensures that the $\mathbf{s}_k$'s are uncorrelated, i.e., $\mathrm{E}\{\mathbf{s}_k\mathbf{s}_l^\mathrm{H}\} = \mathbf{0} \ \forall k,l : k \neq l$. Defining the signal and noise variances as $\lambda^2 = \mathrm{E}\{s_k[i]s_k[i]^*\}$ and $\sigma^2 = \mathrm{E}\{\eta[k]\eta[k]^*\}$, respectively, we can derive the first- and second-order statistics of ϵ_k

$$\begin{aligned} \mathrm{E}\{\epsilon_k\} &= \mathbf{0}_{(n_\mathrm{t}+L)\times 1} \\ \mathrm{E}\left\{\epsilon_k\epsilon_k^\mathrm{H}\right\} &\stackrel{\triangle}{=} \mathbf{Q} = \lambda^2 \mathbf{H}_s \mathbf{H}_s^\mathrm{H} + \sigma^2 \mathbf{I} \\ \mathrm{E}\{\epsilon_k\epsilon_l^\mathrm{H}\} &= \mathbf{0}_{(n_\mathrm{t}+L)\times(n_\mathrm{t}+L)} \ \forall k,l : k \neq l. \end{aligned} \tag{4}$$

III. ML Approach for Channel Identification

The Gaussian likelihood function is established making the hypothesis that unknown data 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 approach yields more accurate channel estimates than the deterministic approach where the unknown data symbols are considered as unknown deterministic disturbances. Adopting the Gaussian hypothesis, we can express (up to a constant term) the negative log likelihood function of the system as

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

Relying on the definition of $\mathbf{Q}$, the loglikelihood can be expressed as a direct function of the unknown parameters $\mathbf{h}$ and σ^2. The corresponding ML channel estimate minimizes this expression with respect to $\mathbf{h}$ and σ^2. This minimization problem boils down to a computationally demanding $(L+2)$-dimensional nonlinear search. To overcome this complexity problem, we propose that the structure of $\mathbf{Q}$ be disregarded, 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 ϵ_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 nonparametric ML problem. The nonparametric ML channel estimate thus maximizes the likelihood function with respect to $\mathbf{h}$ and $\mathbf{Q}$ (instead of $\mathbf{h}$ and σ^2). These assumptions transform the parametric ML problem in $\mathbf{h}$ and σ^2 into a new optimization problem that 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 nonparametric ML problem differs from the solution of the parametric ML problem. 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 CRBs.

IV. CRBs

We show later (see Section VII) that the channel estimates derived from the nonparametric ML problem are consistent and thus asymptotically unbiased. The CRB is a theoretical lower bound on the covariance matrix of an unbiased estimate (see, e.g., [18, p. 562]). In this section, we analyze the impact of the nonparametric hypothesis on the accuracy of the derived channel estimate through this theoretical bound. Adapting the results presented in [11], the real Fisher information matrix (FIM) of the parametric ML problem can be formulated as

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

where

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

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

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 nonparametric ML problem simplifies the expression of the FIM since the $\partial\mathbf{Q}/\partial h[i]^*$ terms are equal to zero. The complex FIM can then be used and is expressed as

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

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 nonparametric 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 (i.e., $(\text{parametric CRB} - \text{nonparametric CRB})/\text{parametric CRB}$) is less than 10^{-4} for experimental setups similar to the ones that are used in Section VIII. 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. The nonparametric CRB can be evaluated as

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

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 of analytically minimizing the cost function with respect to one variable while keeping the other(s) fixed. The variable with respect to which the cost function is minimized is changed in each iteration (see, e.g., [19], 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 cost function is convex, this convergence point is the global minimum. If the cost function is not convex, the convergence point is the global minimum only if the starting point is sufficiently close to the global minimum.[1] In the sequel, we apply this approach to the likelihood function of the system, which leads to the ML estimate of $\mathbf{Q}$ and $\mathbf{h}$. The convergence properties in this context are discussed at the end of this section.

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 $\hat{\mathbf{h}}^i = \mathbf{h}_{\mathrm{ML}}(\hat{\mathbf{Q}}^i)$, where $\mathbf{h}_{\mathrm{ML}}(\mathbf{Q}) = \arg\min_{\mathbf{h}} -\mathcal{L}$. The solution to this optimization problem can be computed as

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

We then seek the covariance matrix $\hat{\mathbf{Q}}^{i+1}$ that minimizes (5) with respect to $\mathbf{Q}$ for a fixed $\mathbf{h} = \hat{\mathbf{h}}^i : \hat{\mathbf{Q}}^{i+1} = \mathbf{Q}_{\mathrm{ML}}(\hat{\mathbf{h}}^i)$, where $\mathbf{Q}_{\mathrm{ML}}(\mathbf{h}) = \arg\min_{\mathbf{Q}} -\mathcal{L}$, the solution to this optimization problem can be computed as (see, e.g., [20])

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

$\hat{\mathbf{Q}}^{i+1}$ 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.

Assessing the convergence properties of this iterative procedure is not straightforward in this context as the likelihood function has not been proven as convex. Moreover, it is not straightforward either to assess the possible local minima of the likelihood function. Hence, it is hardly possible to derive theoretical conditions under which a given starting point would yield a guaranteed convergence of the iterative procedure to the global minimum.

Experimental results presented in [21] show however that when the iterative procedure is initialized with a simple least-squares (LS) channel estimate, the convergence point is systematically equal to the global minimum. Initializing the iterative procedure with the LS channel estimate is equivalent to choosing $\hat{\mathbf{Q}}^0 = \mathbf{I}$. In that case, it is possible to show analytically that the channel estimate $\hat{\mathbf{h}}^1$ obtained after two iterations is a good approximation of the closed-form ML channel estimates that are derived in the next section. Hence, after two iterations, the iterative scheme proceeds from a channel estimate that lies in the immediate vicinity of the global minimum, which ensures the convergence of the iterative procedure to the true ML channel estimate when $\hat{\mathbf{Q}}^0 = \mathbf{I}$ is chosen as starting point.

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., [22]). This approach is often used in ML problems and is known as the process of compressing or concentrating the likelihood function onto one variable (see, e.g., [20] or [23]). In this section, we concentrate the likelihood function on the variable $\mathbf{h}$ in order to derive the closed-form channel estimate.

First observe that the likelihood function (5) can be expressed as

$$-\mathcal{L} = K \ln |\mathbf{Q}| + \mathrm{tr}\left(\mathbf{Q}^{-1} \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}_k \mathbf{h})(\mathbf{u}_k - \mathbf{T}_k \mathbf{h})^{\mathrm{H}} \right). \tag{14}$$

[1]The convergence to the global minimum is guaranteed iff the cost function evaluated at the starting point of the iterations is lower than any local minimum, in which case the trajectory of the iterative procedure is limited to the convex region around the global minimum.

We first minimize this cost function with respect to $\mathbf{Q}$ leading to $\mathbf{Q}_{\mathrm{ML}}(\mathbf{h})$ as given by (13). Replacing $\mathbf{Q}$ by $\mathbf{Q}_{\mathrm{ML}}(\mathbf{h})$ in (14) leaves us with the following expression of the concentrated likelihood function

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

The ML channel estimate is thus computed as

$$\mathbf{h}_{\mathrm{ML}} = \arg\min_{\mathbf{h}} \left| \sum_{k=1}^{K} (\mathbf{u}_k - \mathbf{T}_k \mathbf{h})(\mathbf{u}_k - \mathbf{T}_k \mathbf{h})^{\mathrm{H}} \right|. \tag{15}$$

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 $\mathbf{T}$.

Define

$$\hat{\mathbf{R}} \triangleq K^{-1} \sum_{k=1}^{K} \mathbf{u}_k \mathbf{u}_k^{\mathrm{H}} \quad \bar{\mathbf{u}} \triangleq K^{-1} \sum_{k=1}^{K} \mathbf{u}_k \quad \hat{\mathbf{Q}} \triangleq \hat{\mathbf{R}} - \bar{\mathbf{u}}\bar{\mathbf{u}}^{\mathrm{H}} \tag{16}$$

where $\hat{\mathbf{Q}}$ is assumed to be positive definite (a necessary condition for this to hold is $K \geqslant n_{\mathrm{t}} + L$). Using these definitions, the matrix in the minimization problem (15) can be reexpressed as

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

Keeping in mind that $\hat{\mathbf{Q}}$ is positive definite, our minimization problem (15) is thus equivalent to

$$\mathbf{h}_{\mathrm{ML}} = \arg\min_{\mathbf{h}} \left| \mathbf{I} + \hat{\mathbf{Q}}^{-1} (\mathbf{T}\mathbf{h} - \bar{\mathbf{u}})(\mathbf{T}\mathbf{h} - \bar{\mathbf{u}})^{\mathrm{H}} \right|. \tag{17}$$

It can be shown that

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

Hence, the minimization problem (17) is equivalent to

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

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

$$\mathbf{h}_{\mathrm{ML}} = (\mathbf{T}^{\mathrm{H}} \hat{\mathbf{Q}}^{-1} \mathbf{T})^{-1} (\mathbf{T}^{\mathrm{H}} \hat{\mathbf{Q}}^{-1} \bar{\mathbf{u}}). \tag{18}$$

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

Let us first introduce the following notations

$$\begin{aligned}
\mathbf{h}_{\mathrm{LS}} &\triangleq \left(\sum_{k=1}^{K} \mathbf{T}_k^{\mathrm{H}} \mathbf{T}_k \right)^{-1} \sum_{k=1}^{K} \mathbf{T}_k^{\mathrm{H}} \mathbf{u}_k \\
\mathbf{g} &\triangleq \mathbf{h} - \mathbf{h}_{\mathrm{LS}}, \\
\mathbf{g}_{\mathrm{ML}} &\triangleq \mathbf{h}_{\mathrm{ML}} - \mathbf{h}_{\mathrm{LS}} \\
\mathbf{e}_k &\triangleq \mathbf{u}_k - \mathbf{T}_k \mathbf{h}_{\mathrm{LS}} \\
\hat{\mathbf{Q}}' &\triangleq K^{-1} \sum_{k=1}^{K} \mathbf{e}_k \mathbf{e}_k^{\mathrm{H}}
\end{aligned} \tag{19}$$

where $\hat{\mathbf{Q}}'$ is assumed to be positive definite (a necessary condition therefore is $K \geqslant n_{\mathrm{t}} + L$). Using these notations, the minimization problem (15) can be rephrased as

$$\mathbf{g}_{\mathrm{ML}} = \arg\min_{\mathbf{g}} \left| \mathbf{I} + \hat{\mathbf{Q}}'^{-1} K^{-1} \sum_{k=1}^{K} \left[\mathbf{T}_k \mathbf{g}\mathbf{g}^{\mathrm{H}} \mathbf{T}_k^{\mathrm{H}} - \mathbf{T}_k \mathbf{g} \mathbf{e}_k^{\mathrm{H}} - \mathbf{e}_k \mathbf{g}^{\mathrm{H}} \mathbf{T}_k^{\mathrm{H}} \right] \right|. \tag{20}$$

When K is large, both $\mathbf{h}_{\mathrm{LS}}$ and $\mathbf{h}_{\mathrm{ML}}$ 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 (20), is small in the vicinity of the solution. It is well known that, for $\|\mathbf{\Delta}\| \ll 1$, $|\mathbf{I} + \mathbf{\Delta}| \approx 1 + \mathrm{tr}(\mathbf{\Delta})$. Exploiting this result together with the permutation property of the trace of a product, the ML problem can be approximated by

$$\hat{\mathbf{g}}_{\mathrm{ML}} = \arg\min_{\mathbf{g}} \sum_{k=1}^{K} \left[\mathbf{g}^{\mathrm{H}} \mathbf{T}_k^{\mathrm{H}} \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \mathbf{g} - \mathbf{e}_k^{\mathrm{H}} \hat{\mathbf{Q}}'^{-1} \mathbf{T}_k \mathbf{g} - \mathbf{g}^{\mathrm{H}} \mathbf{T}_k^{\mathrm{H}} \hat{\mathbf{Q}}'^{-1} \mathbf{e}_k \right]. \tag{21}$$

The solution to this minimization problem is given as

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

Replacing $\mathbf{h}$ by $\mathbf{g} + \mathbf{h}_{\mathrm{LS}}$ and $\mathbf{e}_k$ by $\mathbf{u}_k - \mathbf{T}_k \mathbf{h}_{\mathrm{LS}}$, and rearranging the resulting expression, we obtain the following approximation $\hat{\mathbf{h}}_{\mathrm{ML}}$ of the true ML channel estimate $\mathbf{h}_{\mathrm{ML}}$:

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

Note that this solution is obtained after two steps of the previously proposed iterative procedure.

C. Identifiability Conditions

In theory (see, e.g., [24] for a broader discussion on channel identifiability), the channel is identifiable in the considered Gaussian problem formulation when two conditions are fulfilled: the number of channel-output samples considered for channel identification is larger than the channel order L and there is at least one nonzero training symbol not located at the edges of the considered burst.

The first condition is always fulfilled in the considered data model and the second is fulfilled as soon as there is a nonzero element in each training sequence.

However, as we need to invert the estimate of the matrix $\mathbf{Q}$ in the proposed methods (both the iterative method and the closed-form solutions), this estimate has to have a full rank. Relying on the randomness of the noise and the unknown data symbols, this happens with probability 1 as soon as $K \geqslant n_\mathrm{t} + L$. Hence, the identifiability conditions of the proposed method are summarized as follows.

1) $n_t \geqslant 1$.
2) $K \geqslant n_\mathrm{t} + L$.

VII. Asymptotic Properties of the Closed-Form Channel Estimates

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

A. Constant Training

Keeping in mind that $\mathbf{u}_k = \mathbf{T}\mathbf{h} + \epsilon_k$, $\hat{\mathbf{Q}}$ can be rewritten as

$$\hat{\mathbf{Q}} = K^{-1} \sum_{k=1}^{K} \epsilon_k \epsilon_k^\mathrm{H} - K^{-2} \sum_{i,j=1}^{K} \epsilon_i \epsilon_j^\mathrm{H}. \tag{23}$$

This expression shows that $\hat{\mathbf{Q}}$ is clearly a consistent estimate of $\mathbf{Q}$

$$\lim_{K\to\infty} \hat{\mathbf{Q}} = \mathbf{Q}. \tag{24}$$

It follows that the ML channel estimate (18) is consistent as well:

$$\lim_{K\to\infty} \mathbf{h}_\mathrm{ML} = \lim_{K\to\infty} K^{-1} \sum_{k=1}^{K} (\mathbf{T}^\mathrm{H}\hat{\mathbf{Q}}^{-1}\mathbf{T})^{-1}\mathbf{T}^\mathrm{H}\hat{\mathbf{Q}}^{-1}\mathbf{u}_k = \mathbf{h}. \tag{25}$$

Relying on these results, we can derive the asymptotic efficiency of the proposed channel estimate

$$\begin{aligned}
&\lim_{K\to\infty} \mathrm{E}\left\{K(\mathbf{h}_\mathrm{ML} - \mathbf{h})(\mathbf{h}_\mathrm{ML} - \mathbf{h})^\mathrm{H}\right\} \\
&= \mathrm{E}\Bigg\{ \lim_{K\to\infty} (\mathbf{T}^\mathrm{H}\hat{\mathbf{Q}}^{-1}\mathbf{T})^{-1}\mathbf{T}^\mathrm{H}\hat{\mathbf{Q}}^{-1}K^{-1} \\
&\qquad \times \sum_{i=1}^{K} \epsilon_i \sum_{j=1}^{K} \epsilon_j^\mathrm{H} \hat{\mathbf{Q}}^{-1}\mathbf{T}(\mathbf{T}^\mathrm{H}\hat{\mathbf{Q}}^{-1}\mathbf{T})^{-1} \Bigg\} \\
&= (\mathbf{T}^\mathrm{H}\mathbf{Q}^{-1}\mathbf{T})^{-1}.
\end{aligned} \tag{26}$$

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 detailed in the Appendix)

$$\lim_{K\to\infty} \hat{\mathbf{Q}}' = \mathbf{Q}. \tag{27}$$

Replacing $\mathbf{u}_k$ by its equivalent $\mathbf{T}_k\mathbf{h} + \epsilon_k$ in (22) yields

$$\hat{\mathbf{h}}_\mathrm{ML} = \mathbf{h} + \left(\sum_{k=1}^{K} \mathbf{T}_k^\mathrm{H}\hat{\mathbf{Q}}'^{-1}\mathbf{T}_k\right)^{-1} \sum_{k=1}^{K} \mathbf{T}_k^\mathrm{H}\hat{\mathbf{Q}}'^{-1}\epsilon_k. \tag{28}$$

Using this expression of $\hat{\mathbf{h}}_\mathrm{ML}$ highlights the fact that the proposed approximate ML channel estimate is consistent: $\lim_{K\to\infty} \hat{\mathbf{h}}_\mathrm{ML} = \mathbf{h}$.

Here also, it is possible to derive the asymptotic efficiency of the proposed channel estimate

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

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

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

The results that are presented are obtained with the closed-form channel estimates (18) and (22). 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

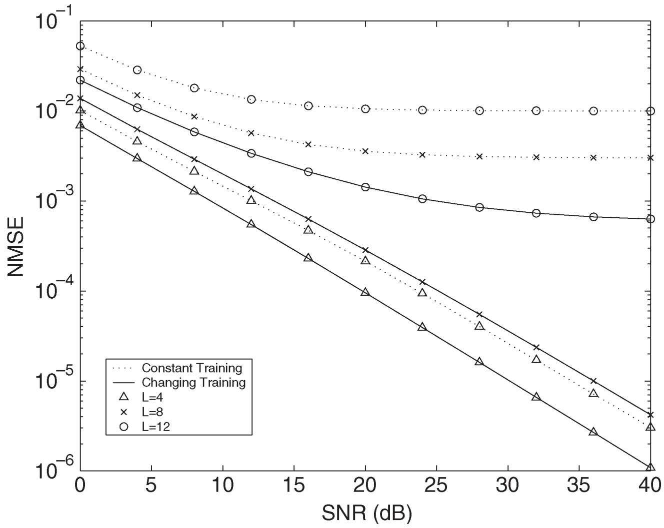

Fig. 1. Comparison of the CRB of the constant-and-changing-training-sequence cases versus the SNR for different channel orders when $n_t = 10$ and $K = 100$.

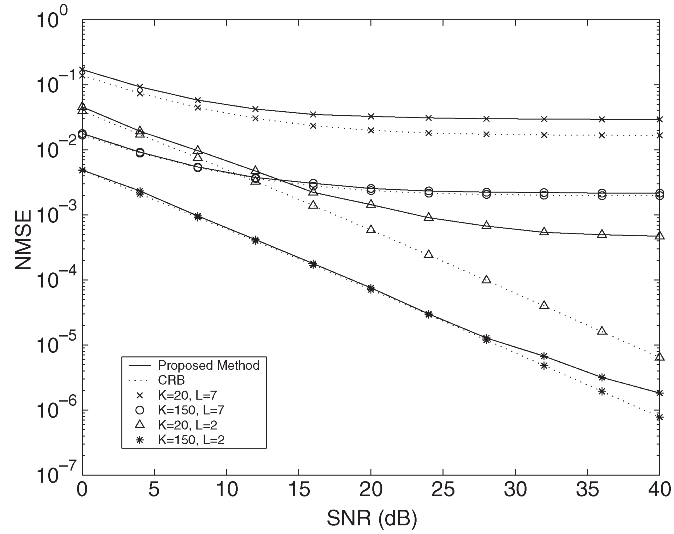

Fig. 2. Comparison of the simulated NMSE and the CRB versus SNR for different channel orders when $n_t = 5$ and changing training sequences are used. The results are plotted for two different values of K, namely 20 and 150.

that achieves the CRB, which is $\mathrm{E}\{\mathrm{tr}(\mathcal{J}^{-1})/\|\mathbf{h}\|^2\}$. 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 randomly picked white QPSK sequences. The energy of the transmitted symbols (both data and training) is set to $\lambda^2 = 1$. 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 $\mathrm{SNR} = \mathrm{E}\{\|\mathbf{h}\|^2\}(\lambda^2/\sigma^2)$.

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 CRBs: 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 as a function of the SNR in Fig. 1. We see that the use of changing training sequences systematically results in a reduced CRB for all channel orders. In both situations, 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, i.e., when $n_t \geqslant 2L+1$ when constant training sequences are used and when $n_t \geqslant L+1$ when changing training sequences are used. The CRB saturates at high SNR, if there is no exact solution to the channel-identification problem in the noiseless case, i.e., when $n_t < 2L+1$ for constant training sequences and $n_t < L+1$ for changing training sequences. When n_t is fixed and the channel order is in the interval $((n_t-1)/2) \leqslant L \leqslant n_t - 1$, 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 a similar behavior (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. When the channel order is large (saturation of the CRB at high SNRs), we observe a relatively good match between the CRB and the experimental curves. The match is tighter for a larger value of K. When the channel order is small (constant slope in the CRB curves), the theoretical and experimental curves match quite well when K is large, but we see the emergence of a floor on the experimental NMSE for higher SNR when K is small. In Fig. 3, we evaluate the impact of the number of data blocks K on the channel estimate NMSE. The simulations are done for two different channel orders and two different values of the SNR. The presented results confirm the asymptotical efficiency of the closed-form channel estimate as the experimental performance systematically achieves the CRB for large values of K. When K gets small, the match between the CRB and the experimental results remains acceptable, except

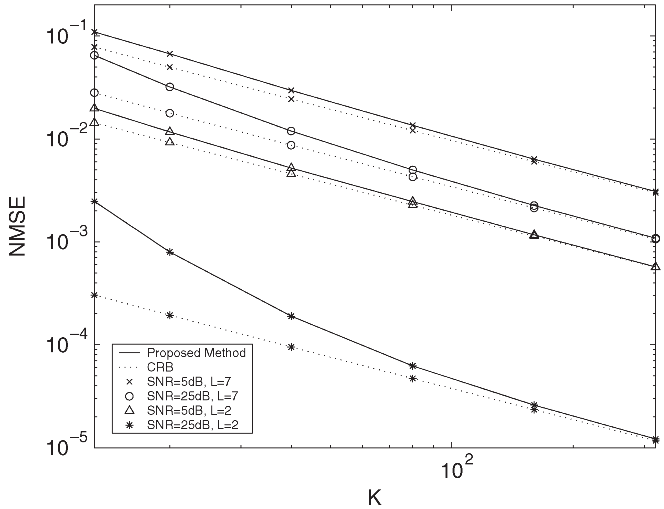

Fig. 3. Comparison of the simulated NMSE and the CRB versus K for different channel orders when $n_{\mathrm{t}} = 5$ and changing training sequences are used. The results are plotted for two different values of the SNR, namely 5 and 25 dB.

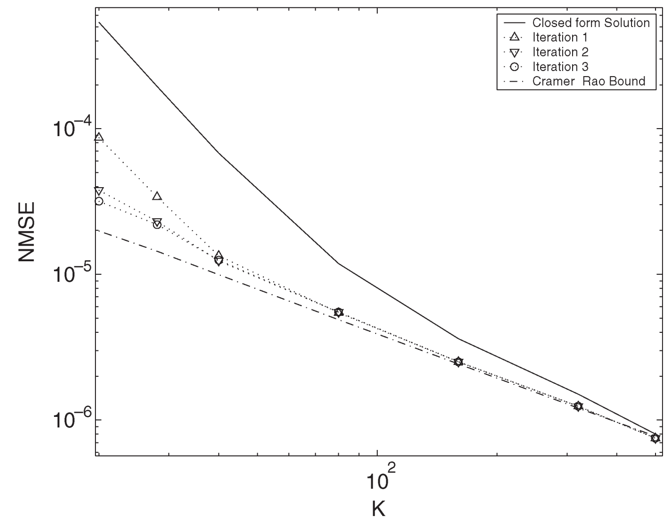

Fig. 5. Comparison of the simulated NMSE and the CRB versus K for $L = 2$, $n_{\mathrm{t}} = 5$, and $\mathrm{SNR} = 35$ dB using changing training sequences. The NMSE converges to the CRB after a few iterations when the iterative method is used.

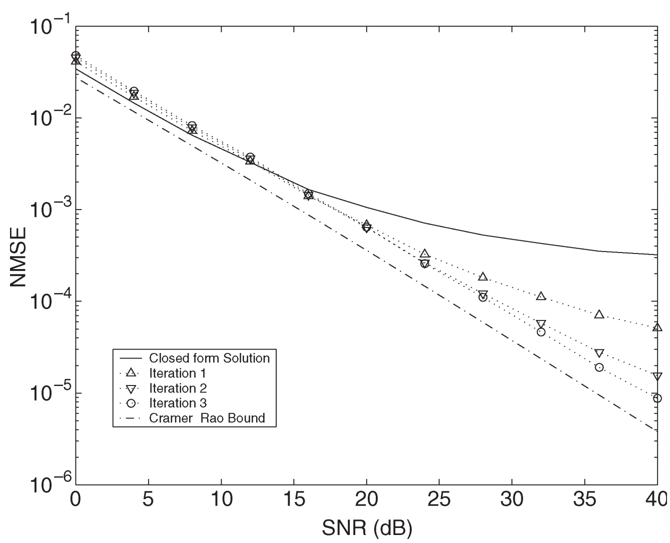

Fig. 4. Comparison of the simulated NMSE and the CRB versus SNR for $L = 2$, $n_{\mathrm{t}} = 5$, and $K = 20$ using changing training sequences. The NMSE converges to the CRB after a few iterations when the iterative method is used.

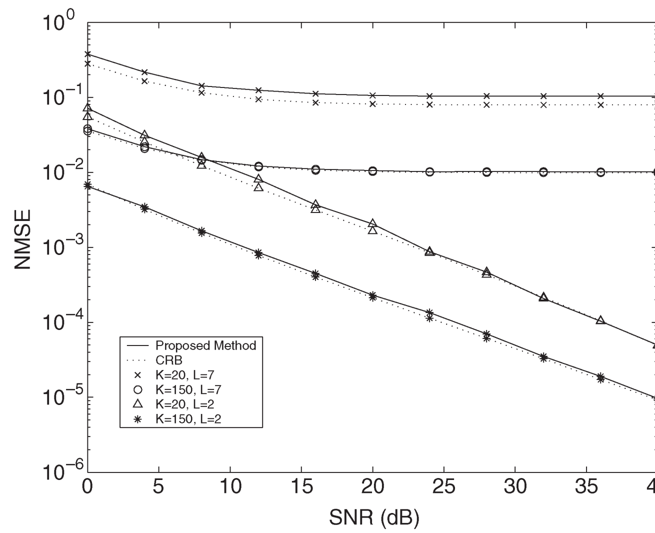

Fig. 6. Comparison of the simulated NMSE and the CRB versus SNR for different channel orders when $n_{\mathrm{t}} = 5$ and constant training sequences are used. The results are plotted for two different values of K, namely 20 and 150.

in the situation where the channel order is small and the SNR is high. In that case, the closed-form estimate is at a significant distance from the CRB. This difference between the CRB and the experimental results originates from the approximations that were needed 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 $\mathbf{h}_{\mathrm{ML}}$. The experiments presented in Figs. 4 and 5 show that the gap between the CRB and the closed-form estimate is closed after a few iterations. Hence, for small K and high SNR, performing a few iterations allows us to effectively achieve the CRB.

3) Constant Training Sequences: In Figs. 6 and 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 Existing Methods

We compare the proposed method with two other methods: a classical training-based ML approach and a more advanced semiblind method based on the EM algorithm presented in [14].

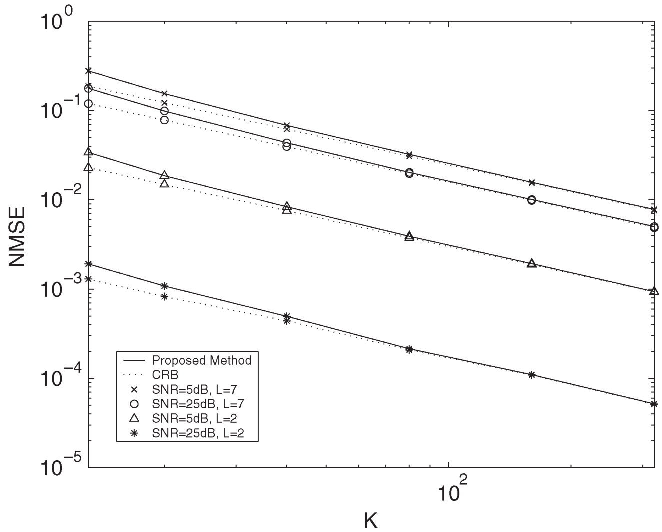

Fig. 7. Comparison of the simulated NMSE and the CRB versus K for different channel orders when $n_{\mathrm{t}} = 5$ and constant training sequences are used. The results are plotted for two different values of the SNR, namely 5 and 25 dB.

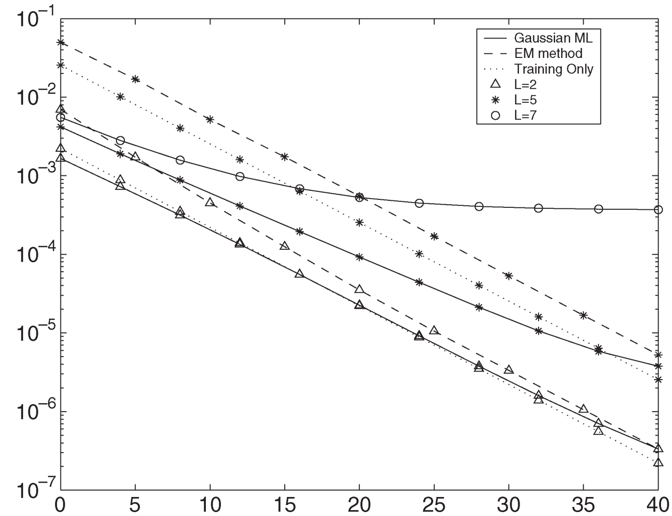

Fig. 8. Simulated NMSE versus SNR for the proposed Gaussian ML method and existing channel-estimation techniques for different channel orders when changing training sequences of length $n_{\mathrm{t}} = 6$ are used and the number of observed blocks is set to $K = 100$.

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 $n_{\mathrm{t}} \geqslant L+1$. In that case, one can define $\mathbf{T}'_k = \mathbf{T}_k(L+1:n_{\mathrm{t}},:)$ and $\mathbf{u}'_k = \mathbf{u}_k(L+1:n_{\mathrm{t}})$. The ML channel model is then known to be the LS fit of $\mathbf{T}'_k\mathbf{h}$ to $\mathbf{u}'_k$

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

When constant training sequences are used, the solution becomes $\mathbf{h}'_{\mathrm{ML}} = (\sum_{k=1}^{K} \mathbf{T}'^{\mathrm{H}}_k \mathbf{T}'_k)^{-1} \sum_{k=1}^{K} \mathbf{T}'_k \mathbf{u}'_k$, and only exists when $n_{\mathrm{t}} \geqslant 2L+1$. Note that the conditions on the training-sequence length n_{t} are much more stringent than in the proposed method.

We compare our algorithm with an adapted version of [14]. We adapt that method in order to estimate the channel relying on the channel-output samples contained in the set of $\mathbf{u}_k$ vectors, as exploiting all the channel output samples would yield an untractable complexity. Moreover, as the method cannot be readily adapted to jointly consider all $\mathbf{u}_k$'s, we compute a channel estimate for each received vector. Considering that the resulting modeling error has a Gaussian distribution, the final ML channel estimate is obtained after computing the mean of the K available channel estimates. Note that since we consider SISO channels rather than the original single-input multiple-output (SIMO) setup, the training sequences must have a minimal length of $n_{\mathrm{t}} \geqslant L+1$. If that condition is not fulfilled, the method cannot be applied because of rank-deficiency problems. Hence, it will not be possible to use this method when the channel order gets too large.

We can now compare the results obtained using these methods with the proposed Gaussian ML estimates. The NMSE of the different methods is presented as a function of the SNR in Fig. 8. We consider changing training sequences and compare the results for different channel orders when the length of the training sequences is set to $n_{\mathrm{t}} = 6$.

The proposed method outperforms existing ones in all situations, except for short channels and at high SNRs where other methods perform slightly better. We know however that increasing the number of observed blocks K or performing a few iterations would restore the advantage of our Gaussian ML method. When the channel order increases, the advantage of the new method increases, especially at low SNR. When the channel order increases and $n_{\mathrm{t}} < L$, the new method still provides reliable channel estimates while traditional methods cannot be applied anymore. Finally, note that the proposed Gaussian ML method has a complexity that is comparable to those of the training-based method, while the semiblind EM-based method has a significantly higher complexity (up to 200 iterations are needed to reach convergence).

IX. CONCLUSION

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 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 couple of iterations. The proposed method can be used in white as well as in colored noise conditions.

It clearly outperforms existing training-based and semiblind ML methods.

APPENDIX

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

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

Based on this observation, it is possible to check that the shorthand notation $\hat{\mathbf{Q}}'$ defined in (19) converges to the true $\mathbf{Q}$ when K tends to infinity

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

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

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

Since the training sequences have nonzero 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.

REFERENCES

[1] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "Optimal training for frequency-selective fading channels," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Salt Lake City, UT, May 2001, pp. 2105–2108.

[2] C. Fragouli, N. Al-Dhahir, and W. Turin, "Training-based channel estimation for multiple-antenna broadband transmissions," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 384–391, Mar. 2003.

[3] J. H. Manton, "On optimal channel identification by use of training sequences and pilot tones," in *Proc. 6th Int. Symp. Signal Processing and Applications (ISSPA)*, Kuala Lumpur, Malaysia, 2001, pp. 599–602.

[4] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.

[5] D. Gesbert, P. Duhamel, and S. Mayrargue, "Blind least-squares approaches for joint data/channel estimation," in *Proc. IEEE Digital Signal Processing Workshop*, Loen, Norway, Sep. 1–4, 1996, pp. 450–453.

[6] S. Schell and J. Bapat, "Partially blind identification of FIR channels for QAM signals," in *Proc. Military Communications Conf.*, San Diego, CA, 1995, vol. 2, pp. 592–596.

[7] V. Buchoux, O. Cappe, E. Moulines, and A. Gorokhov, "On the performance of semi-blind subspace-based channel estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1750–1759, Jun. 2000.

[8] E. de Carvalho and D. Slock, "Deterministic quadratic semi-blind FIR multichannel estimation algorithms and performance," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Istanbul, Turkey, Jun. 5–9, 2000, vol. 5, pp. 2553–2556.

[9] G. Leus and M. Moonen, "Semi-blind channel estimation for block transmission with non-zero padding," in *Proc. Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Nov. 4–7, 2001, pp. 762–766.

[10] J. Ayadi, E. de Carvalho, and D. T. M. Slock, "Blind and semi-blind maximum likelihood methods for FIR multichannel identification," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICCASP)*, Seattle, WA, 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. 1st IEEE Workshop Signal Processing Advances Wireless Communications (SPAWC)*, Paris, France, Apr. 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," in *Proc. 31st Asilomar Conf. Signal, Systems and Computers*, Pacific Grove, CA, Nov. 1997, pp. 1038–1042.

[13] E. de Carvalho and D. T. M. Slock, "Maximum likelihood blind fir multichannel estimation with Gaussian prior for the symbols," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Munich, Germany, Apr. 1997, pp. 3593–3596.

[14] J. L. Bapat, "Partially blind estimation: ML-based approaches and Cramer–Rao bound," *Signal Process.*, vol. 71, no. 3, pp. 265–277, Dec. 1998.

[15] H. A. Cipran and M. K. Tsatsanis, "Stochastic maximum likelihood methods for semi-blind channel estimation," *IEEE Signal Process. Lett.*, vol. 5, no. 1, pp. 21–24, Jan. 1998.

[16] L. Deneire, B. Gyselinckx, and M. Engels, "Training sequence vs. cyclic prefix: A new look on single carrier communication," *IEEE Commun. Lett.*, vol. 5, no. 7, pp. 292–294, Jul. 2001.

[17] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels (mobile radio)," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, Nov. 1991.

[18] T. Söderström and P. Stoica, *System Identification*, ser. Int. Series Syst. Control Eng.. New York: Prentice-Hall, 1989.

[19] O. Rousseaux, G. Leus, and M. Moonen, "A suboptimal iterative method for maximum likelihood sequence estimation in a multipath context," *EURASIP J. Appl. Signal Process. (JASP)*, vol. 2002, no. 12, pp. 1437–1447, Dec. 2002.

[20] L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.

[21] O. Rousseaux, G. Leus, P. Stoica, and M. Moonen, "A stochastic method for training based channel identification," in *7th Int. Symp. Signal Processing and Applications (ISSPA)*, Paris, France, Jul. 2003, pp. 657–660.

[22] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array—Part I: Algorithms," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1184–1197, May 1996.
[23] A. Dogandzic and A. Nehorai, "Space-time fading channel estimation and symbol detection in unknown spatially correlated noise," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 457–473, Mar. 2002.
[24] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, *Signal Processing Advances in Wireless and Mobile Communications*, vol. 1. Englewood Cliffs, NJ: Prentice-Hall, 2000.

Olivier Rousseaux received the M.S. degree in electromechanical engineering from the Universite Catholique de Louvain-la Neuve (UCL), Louvain-la Neuve, Belgium, in 1998. His thesis was in the field of automatic control theory. He is currently pursuing the Ph.D. degree in signal processing for wireless communications at the the Katholic University of Leuven (KUL), Leuven, Belgium, under the supervision of Marc Moonen.

He was formerly with Automatic Control Department, University of Linkoping, Linkoping, Sweden. He also worked for Alcatel from October 1998 to August 1999, before he joined the Electrical Engineering Department (ESAT), Katholic University of Leuven (KUL). He joined the Interuniversity Micro-Electronics Center (IMEC), Leuven, Belgium, in November 2004, where he has since then been working as a Project Officer in the Design Technology for Integrated Information and Communication Systems (DESICS) Division.

Geert Leus (M'01–SM'05) was born in Leuven, Belgium, in 1973. He received the Electrical Engineering degree and Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1996 and 2000, respectively.

He was a Research Assistant and a Postdoctoral Fellow of the Fund for Scientific Research—Flanders, Belgium, from October 1996 until September 2003. During that period, he was affiliated with the Electrical Engineering Department, Katholieke Universiteit Leuven. Currently, he is an Assistant Professor at the Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Delft, The Netherlands. During the summer of 1998, he visited Stanford University, and from March 2001 till May 2002, he was a Visiting Researcher and Lecturer at the University of Minnesota. His research interests are in the area of signal processing for communications. He is currently an Associate Editor of the *EURASIP Journal on Applied Signal Processing*.

Dr. Leus received a 2002 IEEE Signal Processing Society Young Author Best Paper Award. He is a member of the IEEE Signal Processing for Communications Technical Committee, and an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE SIGNAL PROCESSING LETTERS.

Petre Stoica (SM'91–F'94) received the M.Sc. and Ph.D. degrees in automatic control from the Polytechnic Institute of Bucharest, Bucharest, Romania, in 1972 and 1979, respectively.

He is currently a Professor of System Modeling at Uppsala University, Uppsala, Sweden.

Marc Moonen (M'94) received the Electrical Engineering degree and Ph.D. degree in applied sciences from the Katholieke Universiteit (K.U.) Leuven, Leuven, Belgium, in 1986 and 1990, respectively.

Since 2004, he has been a Full Professor at the Electrical Engineering Department, Katholieke Universiteit Leuven, where he is currently heading a research team of 16 Ph.D. candidates and postdocs, working in the area of signal processing for digital communications, wireless communications, digital subscriber line (DSL), and audio-signal processing.

Dr. Moonen was a member of the Editorial Board of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II—EXPRESS BRIEFS from 2002 to 2003. He was the Chairman of the IEEE Benelux Signal Processing Chapter from 1998 to 2002, and since 2000, he has been a EURASIP (European Association for Signal, Speech, and Image Processing) AdCom member.

He has been the Editor-in-Chief for the *EURASIP Journal on Applied Signal Processing* since 2003, and is a member of the Editorial Board of *Integration, the VLSI Journal*, *EURASIP Journal on Wireless Communications and Networking*, and *IEEE Signal Processing Magazine*. He received the 1994 K.U. Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with Piet Vandaele), the 2004 Alcatel Bell (Belgium) Award (with Raphael Cendrillon), and was a 1997 "Laureate of the Belgium Royal Academy of Science."