Efficient OFDM Channel Estimation via an Information Criterion

Alessandro Tomasoni, *Member, IEEE*, Devis Gatti, Sandro Bellini, Marco Ferrari, *Member, IEEE*, and Massimiliano Siti, *Member, IEEE*

***Abstract*—In this paper, we consider joint estimation of the channel length and of the impulse response for OFDM systems, exploiting information criteria to find the best trade-off, in terms of Kullback-Leibler divergence, between noise rejection and channel description accuracy. So far, information criteria have not been used for practical channel length estimation methods, due to their prohibitive complexity. We show how to make them affordable, performing channel estimation in a recursive way that allows to establish the optimal channel length with a moderate incremental cost.**

With reference to IEEE 802.11 OFDM-based standards for WLAN, we investigate several cases, applying the joint channel length and impulse response estimation to many scenarios, ranging from the simplest pilot-aided channel estimation based on training sequences, to the most challenging data-aided channel tracking, driven either by detected or by decoded symbols. In all cases, the achieved performance and robustness are very good, with a very small increase in complexity w.r.t. estimation methods that assume fixed channel length.

***Index Terms*—Channel estimation, channel length estimation, OFDM, Levinson recursion, Akaike criterion.**

I. Introduction

IN this paper, we focus on Channel Estimation (CE) and Channel Length Estimation (CLE) for Orthogonal Frequency Division Multiplexing (OFDM) systems, such as [1], [2]. CE plays a fundamental role in modern communication systems, especially for wireless devices. For a coherent communication, the channel must be estimated at the transmitter and/or receiver side. Knowledge of the Channel State Information (CSI) at the transmitter side is usually the most favorable condition, since the transmitter can apply smart techniques in advance, to adapt the communication to the environment conditions. Some examples are rate adaptation, power loading [3] [4], eigen-beamforming and water-filling [5], and dirty paper coding [6]. CSI at the transmitter side is hard to achieve in wireless communications, especially when mobility must be guaranteed, since the channel can change quickly. Reciprocity between transmitters and receivers cannot be assumed, and receivers must feedback their channel estimate to the transmitters, with a waste of throughput.

An easier task is to let the receivers tackle the channel effects. CE at the receiver should be as effective and cheap as possible, since battery life and device area are scarce resources in wireless communications. CE can be pilot or data aided. Pilot Aided CE (PACE) is based on known training symbols, usually inserted in the preamble or blended in the data payload [3], [7] to help CE. Data Aided CE (DACE) exploits decoded symbols as pilots, to refine a previous estimate or to track channel variations [8]. It is more complex but guarantees less throughput waste.

Focusing on OFDM systems, CE can be performed both in the frequency or in the time domain [3][9]. Frequency Domain (FD) CE is usually simpler, since it can be applied even independently subcarrier by subcarrier: this is the case of Zero Forcing (ZF), that delivers a noisy CE. More refined estimators exploit some additional information about the channel, to reject noise. Depending on the way this information is given, it could be convenient to perform CE either in the FD or in the Time Domain (TD). E.g., if the channel Power Delay Profile (PDP) is known, one can derive the channel covariance matrix in the FD and perform LMMSE estimation to reject more noise, or can even embed the PDP into a TD LMMSE estimator, focusing on the most relevant channel taps [9]. To reduce complexity, one can also also apply estimation to small groups of subcarriers [10], or just exploit the largest eigenvalues of the channel covariance matrix [3]. Reduced order models can also be applied in the TD: the most popular is ML estimation with knowledge of the channel length [7].

A problem with TD CE is to determine the proper channel length. The easiest solution is to set the estimated channel length equal to the maximum value (say L) that produces no Intra Carrier and Inter Symbol Interference, i.e. the OFDM cyclic prefix length plus one. This assumption simplifies the problem, but can lead to a performance degradation, in case of short channel impulse responses. To avoid this degradation some CLE must be implemented. Also CLE can be performed in the frequency or time domain. Frequency domain estimators look for some specific behavior of the channel estimates, such as the zero-level crossing rate [11]. They are simple, but do not exploit all the available information and typically require several OFDM symbols to become reliable. Time domain CLE usually starts from the maximal length and then ignores taps with amplitudes below some threshold [12]. Also this method

Manuscript received July 2, 2012; revised October 27, 2012; accepted December 25, 2012. The associate editor coordinating the review of this paper and approving it for publication was S, Wei.

A. Tomasoni and M. Ferrari are with Consiglio Nazionale delle Ricerche, Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni, Via G. Ponzio 34/5, 20133 Milano, Italy (e-mail: {Alessandro.Tomasoni, Marco.Ferrari}@ieiit.cnr.it).

S. Bellini is with Politecnico di Milano, Dipartimento di Elettronica e Informazione, Piazza L. da Vinci, 20133 Milano, Italy (e-mail: sandro.bellini@elet.polimi.it).

D. Gatti and M. Siti are with STMicroelectronics, Via Olivetti, 2, 20864 Agrate Brianza, Italy (e-mail: {Devis.Gatti, Massimiliano.Siti}@st.com).

Digital Object Identifier 10.1109/TWC.2013.022713.120961

typically has poor performance, and is sensitive to the chosen threshold.

Better results can be achieved by the Akaike Information Criterion (AIC) [13], [14]. In [15], AIC is applied to CE with the aim of minimizing the information lost by the channel estimator. AIC can be deployed also in more complicated scenarios, e.g. in case of DACE driven by unreliable pilots fed back by a channel decoder [8], or when not only the phases and amplitudes, but also the tap delays are parameters to be estimated [16]. Despite its remarkable performance, so far AIC has not been included in practical implementations of OFDM systems with virtual carriers, due to its complexity. A straight implementation requires many CE repetitions, FFTs/IFFTs and Euclidean distance computations. In this paper, we show how to perform CE and CLE jointly. Advantages are manifold. First, the system performance is improved and we can efficiently compute the ML estimates not only for PSK symbols, but also for QAM. Second, the above goals can be achieved with a moderate increase in complexity and memory. Finally, the method is flexible and can be easily generalized to many communication scenarios (e.g. channel tracking with slicing errors) and selection criteria (e.g., the Bayesian Information Criterion [17]).

The remainder of the paper is organized as follows. Section II introduces the system model and the classic CE theory. Section III discusses the application of the Levinson algorithm to the CE problem and estimates its complexity. Section IV introduces the AIC criterion, eventually embedding it into the Levinson algorithm. Section V reports the simulation results for several system configurations and channel conditions. Finally, Section VI concludes the paper.

II. System model and problem definition

In this paper, we denote column vectors and matrices with bold lower-case and capital letters, respectively. $(\cdot)^H$ is the transpose, complex-conjugate (Hermitian) operator. $\mathrm{diag}(\mathbf{x})$ puts the elements of $\mathbf{x}$ along the main diagonal of a matrix, equal to 0 elsewhere. $\mathrm{tr}(\cdot)$ computes the matrix trace. $\mathbf{0}$ and $\mathbf{1}$ are all-0 and all-1 matrices, respectively. $\mathbf{u}^{(l)}$ is an all-0 vector, but for the l-th entry equal to 1. $E[\cdot]$ denotes expectation. $\mathrm{I}(\cdot) \in \{0,1\}$ is the indicator function, returning 1 if the argument is true and 0 otherwise. When helpful, we state matrix and vector dimensions with subscripts, as $(\cdot)_{a\times b}$ and $(\cdot)_{a\times 1}$, respectively.

We consider OFDM symbols with N tones. The orthonormal matrix $\mathbf{W}$ of the Inverse Discrete Fourier Transform (I-DFT) has entries $w_{h,k} = \frac{1}{\sqrt{N}}\exp\left(j2\pi\frac{hk}{N}\right)$. The normalized Discrete Fourier Transform (DFT) is implemented by $\mathbf{W}^H$, and $\mathbf{W}^H\mathbf{W} = \mathbf{W}\mathbf{W}^H = \mathbf{I}_{N\times N}$.

Assume that we have a sequence of symbols p_k, namely *pilots*, known at the receiver. Depending on the context, they can be either a Long Training Sequence (LTS) belonging to a preamble, or the sequence of data symbols fed back to the channel estimator by a channel decoder. Only K out of N carriers are *active*. The others, such as DC and the tones close to Nyquist, are not used: they are called *virtual* carriers and their p_k is equal to 0. We can collect pilots in the diagonal matrix $\mathbf{P} = \mathrm{diag}\left(\begin{bmatrix} p_0 & p_1 & \dots & p_{N-1}\end{bmatrix}\right)$.

Usually, the channel length l is shorter than the OFDM cyclic prefix. For the time being, we also assume that l is known. In the time domain, let the channel response be $\mathbf{h}_{l\times 1}$. The N channel samples in the frequency domain are $\mathbf{h}^{(freq.)} = \mathbf{W}^H \begin{bmatrix}\mathbf{I}_{l\times l} \\ \mathbf{0}_{(N-l)\times l}\end{bmatrix} \mathbf{h}_{l\times 1}$. The transmitted pilot symbols $\mathbf{P}$ are distorted by the channel. The m-th received OFDM symbol in the time domain is

$$\begin{aligned}\mathbf{r}_{N\times 1}(m) &= \mathbf{W}\mathbf{P}\mathbf{h}^{(freq.)} + \mathbf{n}(m) \\ &= \mathbf{W}\mathbf{P}\mathbf{W}^H \begin{bmatrix}\mathbf{I}_{l\times l} \\ \mathbf{0}_{(N-l)\times l}\end{bmatrix} \mathbf{h}_{l\times 1} + \mathbf{n}(m) \end{aligned} \quad (1)$$

where $\mathbf{n}(m) \sim \mathcal{CN}\left(\mathbf{0}_{N\times 1}, N_0\mathbf{I}_{N\times N}\right)$ is Additive White Gaussian Noise (AWGN) with variance N_0 equal to the one-sided noise power spectral density (we assume unit-energy transmit and matched receive filters). Without loss of generality, we set $E\left[\left|h_k^{(freq)}\right|^2\right] = E\left[\left|p_k^{(freq)}\right|^2\right] = 1$ (over active carriers), so that the Signal to Noise Ratio (SNR) is equal to $\frac{1}{N_0}$.

The channel estimator exploits $\mathbf{r}(m)$ to estimate $\mathbf{h}$. M OFDM symbols are available for CE. The Least Squares (LS) estimate is computed applying the Moore-Penrose pseudo-inverse [3]

$$\begin{aligned}\tilde{\mathbf{h}}_{l\times 1}(m) = &\left(\begin{bmatrix}\mathbf{I}_{l\times l} & \mathbf{0}_{l\times(N-l)}\end{bmatrix} \mathbf{Q} \begin{bmatrix}\mathbf{I}_{l\times l} \\ \mathbf{0}_{(N-l)\times l}\end{bmatrix}\right)^{-1} \\ &\cdot \begin{bmatrix}\mathbf{I}_{l\times l} & \mathbf{0}_{l\times(N-l)}\end{bmatrix} \mathbf{W}\mathbf{P}^H\mathbf{W}^H\mathbf{r}(m)\end{aligned} \quad (2)$$

where $\mathbf{Q} = \mathbf{W}\mathbf{P}^H\mathbf{P}\mathbf{W}^H$ is Toeplitz and Hermitian. In (2), the rightmost part

$$\begin{aligned}\hat{\mathbf{h}}_{N\times 1}(m) &= \mathbf{W}\mathbf{P}^H\mathbf{W}^H\mathbf{r}(m) \\ &= \mathbf{Q}\begin{bmatrix}\mathbf{I}_{l\times l} \\ \mathbf{0}_{(N-l)\times l}\end{bmatrix}\mathbf{h}_{l\times 1} + \mathbf{W}\mathbf{P}^H\mathbf{W}^H\mathbf{n}(m)\end{aligned} \quad (3)$$

can be interpreted as a *coarse* frequency domain CE, based on a tone-by-tone Matched Filter (MF). Equations (2)-(3) are more readable with the "MATLAB-style" notation $1:l$:

$$\mathbf{Q}_{1:l,1:l} \triangleq \begin{bmatrix}\mathbf{I}_{l\times l} & \mathbf{0}_{l\times(N-l)}\end{bmatrix} \mathbf{Q} \begin{bmatrix}\mathbf{I}_{l\times l} \\ \mathbf{0}_{(N-l)\times l}\end{bmatrix} \quad (4)$$

$$\hat{\mathbf{h}}_{1:l}(m) \triangleq \begin{bmatrix}\mathbf{I}_{l\times l} & \mathbf{0}_{l\times(N-l)}\end{bmatrix} \hat{\mathbf{h}}(m) \quad (5)$$

Then, (2) becomes[1]

$$\tilde{\mathbf{h}}(m) = \mathbf{Q}_{1:l,1:l}^{-1}\hat{\mathbf{h}}_{1:l}(m) \quad (6)$$

where $\tilde{\mathbf{h}}(m) \sim \mathcal{CN}\left(\mathbf{h}_{l\times 1}, N_0\mathbf{Q}_{1:l,1:l}^{-1}\right)$ in case of PSK, as shown in the following (see (30)). The matrix $\mathbf{Q}_{1:l,1:l}^{-1}$ is a *smoothing* filter, and the low-pass behavior is due to the reduced-rank CE.

If more than one OFDM symbol is available for CE (e.g. in IEEE 802.11a/p standards [1], [2] there are two LTS), we can average partial noisy estimates:

$$\tilde{\mathbf{h}} \triangleq \frac{1}{M}\sum_{m=1}^{M}\tilde{\mathbf{h}}(m) \quad (7)$$

[1]Mind the difference between $X_{1:l,1:l}^{-1} = \left(X_{1:l,1:l}\right)^{-1}$ and $\left(X^{-1}\right)_{1:l,1:l}$.

If the matrix $\mathbf{Q}$ is the same for all OFDM symbols there is no need to perform channel smoothing many times: for linearity we can smooth the averaged coarse estimates and compute

$$\tilde{\mathbf{h}} = \mathbf{Q}_{1:l,1:l}^{-1} \underbrace{\left(\frac{1}{M} \sum_{m=1}^{M} \hat{\mathbf{h}}(m) \right)}_{\triangleq \hat{\mathbf{h}}} = \mathbf{Q}_{1:l,1:l}^{-1} \hat{\mathbf{h}}_{1:l} \tag{8}$$

In case of AWGN, the above computations coincide with ML estimation. There are other popular solutions, e.g. the Zero Forcing (ZF) CE. In this case, the received symbols are multiplied by a diagonal matrix $\check{\mathbf{P}}$ (rather than by $\mathbf{P}^H$), where $\check{\mathbf{P}}_{k,k} = \frac{1}{p_k}$ if $p_k > 0$, and 0 otherwise. Thus, one obtains $\check{\mathbf{h}}_{N\times 1}(m) = \mathbf{W}\check{\mathbf{P}}\mathbf{W}^H \mathbf{r}(m) = \check{\mathbf{Q}} \begin{bmatrix} \mathbf{I}_{l\times l} \\ \mathbf{0}_{(N-l)\times l} \end{bmatrix} \mathbf{h}_{l\times 1} + \mathbf{W}\check{\mathbf{P}}\mathbf{W}^H \mathbf{n}(m)$ with $\check{\mathbf{Q}} = \mathbf{W}\mathbf{A}\mathbf{W}^H$, being $\mathbf{A} = \mathrm{diag}\left(\begin{bmatrix} \mathrm{I}(|p_0| > 0) & \mathrm{I}(|p_1| > 0) & \ldots & \mathrm{I}(|p_{N-1}| > 0) \end{bmatrix}\right)$ the matrix of active carriers. After this processing, LS estimation is still performed, obtaining $\check{\tilde{\mathbf{h}}}_{l\times 1}(m) = \check{\mathbf{Q}}_{1:l,1:l}^{-1} \check{\mathbf{h}}_{1:l}(m)$. Notice the formal analogy with (6), due to the idempotence of $\check{\mathbf{Q}}$.

For PSK constellations (such as BPSK and QPSK), even the ZF filter leads to ML estimation, since $\mathbf{P}^H\mathbf{P} = \check{\mathbf{P}}\mathbf{P} = \mathbf{A}$. When the symbol amplitude is modulated (such as in case of QAM constellations), ZF becomes suboptimal, since the inversion of the weakest pilot symbols increases the average noise power. Yet, this approximation is attractive from a complexity point of view, since $\check{\mathbf{Q}}$ is independent of pilot symbols, and can be computed and inverted just once. Besides, when the position of active carriers is symmetric in the base-band spectrum (as in [1], [2] if we do not apply symbol selection [8]) $\check{\mathbf{Q}}$ is real. In the following, we will discuss the complexity of both MF and ZF solutions, for our joint CLE and CE.

III. Channel Estimation exploiting the Levinson Algorithm

In this Section, we discuss how to compute the smoothed channel $\tilde{\mathbf{h}}$ in (8) (and its analogous $\check{\tilde{\mathbf{h}}}$ in case of ZF), applying the Levinson Algorithm (LA) rather than multiplying by $\mathbf{Q}_{1:l,1:l}^{-1}$. The advantage is twofold. First, the LA will enable an efficient and effective channel length estimator, described in the Section IV. Second, the LA can also manage non constant matrices $\mathbf{Q}_{1:l,1:l}$, that otherwise would require the storage of a large number of inverse matrices. For this reason, the LA is appealing also on its own, and will be discussed first.

Having in mind (8), let us consider the solution of the system of equations

$$\mathbf{T}_{1:l,1:l}\mathbf{y}^{(l)} = \mathbf{x}_{1:l} \tag{9}$$

where $\mathbf{y}^{(l)}$ represents the smoothed channel estimate (either $\tilde{\mathbf{h}}$ or $\check{\tilde{\mathbf{h}}}$) having length l, $\mathbf{x}$ is the MF or ZF coarse channel estimate ($\mathbf{h}$ or $\check{\mathbf{h}}$), and $\mathbf{T}$ is $\mathbf{Q}$ or $\check{\mathbf{Q}}$. In principle, with the aim of jointly estimating the channel taps and their number, one should solve the above system for *all* possible channel lengths $l \leq L$, and choose the best result (the choice criterion will be discussed in the next Section).

In general, a matrix inversion costs $\mathcal{O}\left(l^3\right)$ multiplications. The overall cost to invert L matrices, with size l ranging from 1 to L, is $\mathcal{O}\left(L^4\right)$, unacceptable even for moderate L. A first improvement could be to compute off-line all the inverse matrices $\mathbf{T}_{1:l,1:l}^{-1}$, if known in advance, and just perform the L multiplications $\mathbf{T}_{1:l,1:l}^{-1}\mathbf{x}_{1:l}$, whose cost is $\mathcal{O}\left(l^2\right)$ each, thus leading to an overall complexity $\mathcal{O}\left(L^3\right)$. This cost is still high, and a large burden of memory is required to store all the inverses.

For our CE purposes, $\mathbf{T}_{1:l,1:l}$ is always Toeplitz, i.e. elements belonging to the same diagonal are constant. In this case, matrix inversion costs $\mathcal{O}\left(l^2\right)$ and can be performed by the LA [18], [19], [20]. The LA has a remarkable advantage: instead of computing the inverse matrix and eventually multiplying it by $\mathbf{x}_{1:l}$ it directly computes the solution $\mathbf{y}^{(l)}$ of (9). This is done recursively, solving the 1×1 subsystem first, then correcting the result at the previous step to find the solution of the 2×2 system and so on. The intermediate steps of the algorithm give the solutions of the smaller systems (9) with $l < L$. With overall complexity $\mathcal{O}\left(L^2\right)$ we compute the smoothed channel for every channel length l.

The LA can be split in two parts: a *core* algorithm, to be repeated for every different vector $\mathbf{x}$ (in our case, $\mathbf{x}$ is the coarse CE to be smoothed), and a *support* algorithm, assisting the core one in the computation of a set of support vectors $\mathbf{b}$, depending only on $\mathbf{T}$. This second algorithm is launched only if $\mathbf{T}$ changes, e.g. if we apply symbol selection [8] or we treat QAM modulations. When $\mathbf{T}$ is constant, e.g. when pilot symbols are PSK or the ZF coarse estimation is chosen, the computation of vectors $\mathbf{b}$ can be done only once, say off-line. In general, all variables of the LA are complex. When the matrix $\mathbf{T}$ is real (once again, only if symbol selection [8] is disregarded and the symbol power is uniform), also the support vectors are real.

A. Core and support algorithm description

In the core algorithm, the LA finds a solution of (9) assuming that the solution of the system $\mathbf{T}_{1:l-1,1:l-1}\mathbf{y}^{(l-1)} = \mathbf{x}_{1:l-1}$ has been previously computed. The algorithm augments $\mathbf{y}^{(l-1)}$ to the size of the new problem, measures the error and applies a correction to find $\mathbf{y}^{(l)}$. This is done finding a set of vectors, namely the *backward vectors*, such that

$$\mathbf{T}_{1:l,1:l}\mathbf{b}^{(l)} = \mathbf{u}^{(l)} \tag{10}$$

These vectors can be efficiently computed too, as reported in the following. The augmented solution is wrong only in the last position of $\mathbf{x}$, since

$$\mathbf{T}_{1:l,1:l} \begin{bmatrix} \mathbf{y}^{(l-1)} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{1:l-1} \\ \epsilon_y^{(l)} \end{bmatrix} \tag{11}$$

where $\epsilon_y^{(l)} \triangleq \mathbf{T}_{l,1:l-1}\mathbf{y}^{(l-1)}$. Thanks to linearity, the above result can be easily adjusted to be a solution of the $l \times l$

system, obtaining $\mathbf{y}^{(l)}$:

$$\underbrace{\mathbf{T}_{1:l,1:l}\left(\begin{bmatrix}\mathbf{y}^{(l-1)}\\ 0\end{bmatrix}+\left(x_l-\epsilon_y^{(l)}\right)\mathbf{b}^{(l)}\right)}_{=\mathbf{y}^{(l)}} = \begin{bmatrix}\mathbf{x}_{1:l-1}\\ \epsilon_y^{(l)}\end{bmatrix}+\left(x_l-\epsilon_y^{(l)}\right)\mathbf{u}^{(l)} \quad (12)$$

Focusing now on the support algorithm, the solution $\mathbf{b}^{(l-1)}$ at the $(l-1)$-th step is augmented first, and eventually corrected to obtain the backward vectors $\mathbf{b}^{(l)}$. Define the $l\times l$ matrix $\mathbf{R}^{(l)}$ having the secondary diagonal equal to 1, that reverses the elements of any vector that right-multiplies or left-multiplies it. Then, the backward vectors can be recursively computed as [18]:

$$\mathbf{b}^{(l)}=\frac{1}{1-\left|\epsilon_b^{(l)}\right|^2}\begin{bmatrix}0\\ \mathbf{b}^{(l-1)}\end{bmatrix}-\frac{\epsilon_b^{(l)}}{1-\left|\epsilon_b^{(l)}\right|^2}\mathbf{R}^{(l)}\begin{bmatrix}0\\ \left(\mathbf{b}^{(l-1)}\right)^*\end{bmatrix} \quad (13)$$

where $\epsilon_b^{(l)}=\mathbf{T}_{1,2:l}\mathbf{b}^{(l-1)}$.

B. Algorithm cost and memory requirements

In this Subsection, we estimate the complexity (measured as the number of scalar real multiplications) of the LA: $\epsilon_y^{(l)}$ costs $l-1$ complex-real multiplications; equation (12) costs l complex-real multiplications; $\epsilon_b^{(l)}$ costs $l-1$ real multiplications; equation (13) costs $3+2(l-1)$ real multiplications. The last two values refer to the support processing, and must be repeated only if the problem changes, e.g. if we apply symbol selection [8] or we treat QAM modulation. When $\mathbf{T}$ is constant the support algorithm is performed just once, so it is worth counting it separately: core operations require $2\sum_{l=1}^{L}(2l-1)=2L^2$ real multiplications; support operations require $3\sum_{l=1}^{L}l=3\frac{L(L+1)}{2}$ real multiplications.

The above results refer to the common case of matrix $\mathbf{T}$ and backward vectors $\mathbf{b}^{(l)}$ real, e.g. if symbol selection [8] is disregarded and the symbol power is uniform, or if the ZF estimation is adopted. Otherwise, the core operations cost is doubled, and the support operations cost is quadrupled (to be fair, also the multiplication by $\mathbf{T}_{1:L}^{-1}$ costs twice as much). One should also take into account the cost of the coarse estimates (in general, three FFTs and K complex multiplications, say $6N\log_2 N+4K$ real multiplications in the worst case), plus another FFT to compute one row of $\mathbf{T}$ ($2N\log_2 N$ real multiplications), when it is variable. Anyway, these costs are in common with all other CE methods with smoothing.

With almost the same complexity as the multiplication by $\mathbf{T}_{1:L,1:L}^{-1}$, we have now the freedom of choosing any intermediate solution produced by the smoothing algorithm, for channel lengths smaller than L. In the following Section, we discuss how to efficiently estimate the proper length $\hat{l}$. After that, Table I will summarize the overall costs.

Concerning memory requirements, the LA algorithm does not store the whole $\mathbf{T}$ matrix, required only by the computation of $\epsilon_y^{(l)}$ and $\epsilon_b^{(l)}$. The first row is enough for both of them, since $\epsilon_y^{(l)}=\mathbf{T}_{l,1:l-1}\mathbf{y}^{(l-1)}=\left(\mathbf{T}_{1,2:l}\right)^*\mathbf{R}^{(l-1)}\mathbf{y}^{(l-1)}$, being $\mathbf{T}$ Toeplitz and Hermitian.

In case of fixed $\mathbf{T}$, also L backward vectors of size l must be saved, approximately leading to $L+\sum_{l=1}^{L}l=\frac{L(L+3)}{2}$ scalar elements (e.g., 170 if $L=17$ as in [1], [2]). The storage of the inverse matrices would require a much larger burden of $\sum_{l=1}^{L}l^2=\frac{L^3}{3}+\frac{L^2}{2}+\frac{L}{6}$ elements (i.e. 1785 with $L=17$).

IV. The Akaike Information Criterion for Channel Length Estimation

In this Section, we introduce the Akaike Information Criterion (AIC) [13], that attempts to find the best model among a finite set. When the choice is among models with different channel lengths [15], [8], it implicitly estimates the most suitable $\hat{l}$. In the following, we show that the huge complexity burden typical of AIC can be almost avoided, deploying a modified AIC estimator that works recursively, fed by the LA results.

Assume that a "large enough" set of n observations, say $\mathbf{y}_{n\times 1}$, is available to estimate the true system pdf $g(\mathbf{y})$ by a model $f(\mathbf{y}|\boldsymbol{\theta})$, depending on p parameters $\boldsymbol{\theta}_{p\times 1}$. In many practical cases ML estimation $\hat{\boldsymbol{\theta}}(\mathbf{y})=\underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ f(\mathbf{y}|\boldsymbol{\theta})$ is satisfactory, in the sense that it is unbiased, its estimation error asymptotically becomes normal and achieves the Minimum Mean Square Error (MMSE) settled by the Cramér-Rao theorem. However, ML cannot choose among models with a different number of parameters, since the larger their number, the higher the likelihood we can achieve. An ML criterion would always choose the maximum allowed p, overfitting the observations and erroneously interpreting statistical fluctuations as relevant features of the system. For this reason, if two models provide similar likelihoods with a different number of parameters, the simplest has to be preferred (this is known as the minimum description length criterion), or equivalently a penalization should be assigned to the most complex. The AIC

$$AIC\left(\mathbf{y},\hat{\boldsymbol{\theta}}(\mathbf{y})\right)=-\ln f\left(\mathbf{y}|\hat{\boldsymbol{\theta}}(\mathbf{y})\right)+p \quad (14)$$

does exactly this.

AIC has a theoretical background [13]: its minimization is related to the Kullback-Leibler Divergence $\mathcal{D}_{KL}(g|f(\boldsymbol{\theta}))$, measuring the information lost by the model[2].

Finally, it is worth noticing that in case of AWGN the ML estimate

$$\tilde{\mathbf{h}}_{l\times 1}(m)=\underset{\mathbf{h}_{l\times 1}}{\operatorname{argmax}}\frac{1}{(\pi N_0)^N}\cdot\exp\left(-\frac{1}{N_0}\left\|\mathbf{W}^H\mathbf{r}(m)-\mathbf{P}\mathbf{W}^H\begin{bmatrix}\mathbf{h}_{l\times 1}\\ \mathbf{0}_{(N-l)\times 1}\end{bmatrix}\right\|^2\right) \quad (15)$$

matches the LS estimate (2). This can be easily proven putting the likelihood derivatives equal to zero. Therefore, the

[2]In detail, $\mathcal{D}_{KL}(g|f(\boldsymbol{\theta}))=-\int g(\mathbf{y})\ln f(\mathbf{y}|\boldsymbol{\theta})\,d\mathbf{y}+\int g(\mathbf{y})\ln g(\mathbf{y})\,d\mathbf{y}=\mathcal{X}(g|f(\boldsymbol{\theta}))-\mathcal{H}(g)$. $\mathcal{X}(g|f(\boldsymbol{\theta}))$ is the cross-entropy (i.e., the minimum number of bits we need to represent our signal assuming the model $f(\boldsymbol{\theta})$), and $\mathcal{H}(g)$ is the entropy of the true system (i.e., the minimum number of bits assuming perfect knowledge). $-\mathcal{X}(g|f(\boldsymbol{\theta}))$ and $-\mathcal{H}(g)$ can be also regarded as average log-likelihoods, and their difference is the lost likelihood, on average. $\mathcal{H}(g)$ depends on the unknown system and cannot be computed, but it is constant for all models and can be neglected during the minimization of $\mathcal{D}_{KL}(g|f(\boldsymbol{\theta}))$.

adoption of the AIC criterion is justified in case of channel smoothing.

AIC has some limits. The pdf in (14) must be exactly known, and AIC is optimal only asymptotically (in practice, n must be one or two order of magnitude larger than p). In the next subsections, we first circumvent the above two problems and then show how AIC can be applied with low complexity, thanks to the LA.

A. AIC and channel estimation

When the noise variance is unknown, or when other noise sources apart from AWGN have been neglected, AIC can be minimized w.r.t. this uncertain parameter. Assuming that the overall noise vector is still i.i.d. and Gaussian, the variance in the log-likelihood expression can be substituted by the Residual Sum of Squares (RSS) $\sigma^2(p)$:

$$\begin{aligned} AIC\,(p) &= -\ln\left(\frac{1}{\left(2\pi\frac{\sigma^2(p)}{n}\right)^{\frac{n}{2}}}\exp\left(-\frac{\sigma^2(p)}{2\frac{\sigma^2(p)}{n}}\right)\right)+p \\ &= \frac{n}{2}\ln\sigma^2(p)+p+const. \end{aligned} \tag{16}$$

Constant terms have no impact on the minimization, and will be neglected in the following.

In practical systems, the above "log-RSS" AIC expression is to be preferred to the original one, since it takes into account all the non-idealities (e.g. synchronization errors, mistaken data symbols used as pilots, phase noise, channel impulse responses exceeding the guard interval and so on), not included in the system model. When these unexpected noise sources arise, AIC chooses a shorter channel length, smoothing the received signal and rejecting as much noise as possible, conferring robustness to the overall system, as shown in the next Section for the case of unreliable pilots.

To apply AIC to our CE problem, we must set $p = 2l$ and $n = 2N$ (since samples are complex), obtaining in case of M OFDM symbols

$$\hat{l} = \underset{1\le l\le L}{\operatorname{argmin}}\, AIC(l) \tag{17}$$

$$AIC(l) = \underbrace{\ln\left(\frac{1}{M}\sum_{m=1}^{M}\sigma_m^2(l)\right)}_{=\sigma^2(l)}+\frac{2}{N}l \tag{18}$$

In the above expression, $\sigma_m^2(l)$ is the Residual Sum of Squares (RSS) of the m-th OFDM symbol and $\sigma^2(l)$ the average RSS. The RSS can be derived looking at the likelihood function, maximized by the optimal ML estimator (15). Thus, we obtain

$$\sigma_m^2(l) = \left\|\mathbf{W}^H\mathbf{r}(m)-\mathbf{P}\mathbf{W}^H\begin{bmatrix}\tilde{\mathbf{h}}(m)\\ \mathbf{0}_{(N-l)\times 1}\end{bmatrix}\right\|^2 \tag{19}$$

Elements embedded in (19) range from 1 to N, including all virtual carriers, whose $p_k = 0$. This is consistent with (15), where virtual carriers appear in the likelihood function. Since virtual carriers are never used, we would like to discard them from the RSS computation. A more careful inspection of (15) reveals that they are useless, since their contribution is constant in

$$\begin{aligned} \tilde{\mathbf{h}}(m) = \underset{\mathbf{h}_{l\times 1}}{\operatorname{argmax}}\, &\exp\left(-\frac{1}{N_0}\left\|(\mathbf{I}-\mathbf{A})\,\mathbf{W}^H\mathbf{r}(m)\right\|^2\right) \\ \cdot &\exp\left(-\frac{1}{N_0}\left\|\mathbf{A}\mathbf{W}^H\mathbf{r}(m)-\mathbf{P}\mathbf{W}^H\begin{bmatrix}\mathbf{h}_{l\times 1}\\ \mathbf{0}_{(N-l)\times 1}\end{bmatrix}\right\|^2\right) \end{aligned} \tag{20}$$

from which AIC derives. Thus, discarding virtual carriers does not affect the AIC CLE, and the corresponding "log-RSS" AIC can be re-stated as

$$AIC(l) = \ln\left(\frac{1}{M}\sum_{m=1}^{M}\sigma_m^2(l)\right)+\frac{2}{K}l \tag{21}$$

$$\sigma_m^2(l) = \left\|\mathbf{A}\mathbf{W}^H\mathbf{r}(m)-\mathbf{P}\mathbf{W}^H\begin{bmatrix}\tilde{\mathbf{h}}(m)\\ \mathbf{0}_{(N-l)\times 1}\end{bmatrix}\right\|^2 \tag{22}$$

Notice that in this case some difference may arise, since the "log-RSS" AIC is only an approximation of the exact one. Anyway, simulations show that the performance is very similar. The choice among different criteria mainly depends on complexity, summarized in the following.

Another issue is AIC optimality in case of small n. Indeed, AIC is unbiased only asymptotically (for small n, also AIC overfits). When $M \ge 2$, the average of different RSSs in (18) and (21) is a heuristic attempt to correct the AIC bias. An alternative, theoretically grounded solution is to employ the Corrected AIC [21], [22], [23] (AICc). AICc is a non-trivial extension of AIC, unbiased even for finite n. AICc is less general than AIC, since the penalty term depends on the problem under investigation. E.g., in case of linear regression (LS estimation), the AICc can be computed as

$$AICc(l) = \ln\sigma^2(l)+\frac{n+p}{n-p-2} = \ln\sigma^2(l)+\frac{K+l}{K-l-1} \tag{23}$$

or $\widetilde{AICc}(l) = \ln\sigma^2(l)+\frac{2}{n}p\left(1+\frac{p+1}{n-p-1}\right) = \ln\sigma^2(l)+\frac{2l}{K-l-0.5}$, with almost the same performance.

AICc is very attractive, since in any simulated case it provided better performance than AIC, as shown is Section V. This advantage is more relevant in case of DACE, when just one OFDM symbol is available for CE purposes. Besides, in case of PACE with multiple OFDM symbols ($M \ge 2$) AICc allows the computation of just one RSS referring to the average of M coarse channel estimates, rather than M RSSs, one per OFDM symbol, necessary to mitigate the AIC bias effect. This significantly reduces the algorithm complexity, passing from $2(MK + L)$ to $2(K + L)$ real multiplications, as shown in Table I.

Finally, though AIC assumes ML estimation, practical applications sometimes use it in conjunction with sub-optimal estimators, such as the coarse ZF of Section II. The resulting RSS

$$\check{\sigma}_m^2(l) = \left\|\check{\mathbf{P}}\mathbf{W}^H\mathbf{r}(m)-\mathbf{A}\mathbf{W}^H\begin{bmatrix}\check{\tilde{\mathbf{h}}}(m)\\ \mathbf{0}_{(N-l)\times 1}\end{bmatrix}\right\|^2 \tag{24}$$

still provides good results, as shown in Section V.

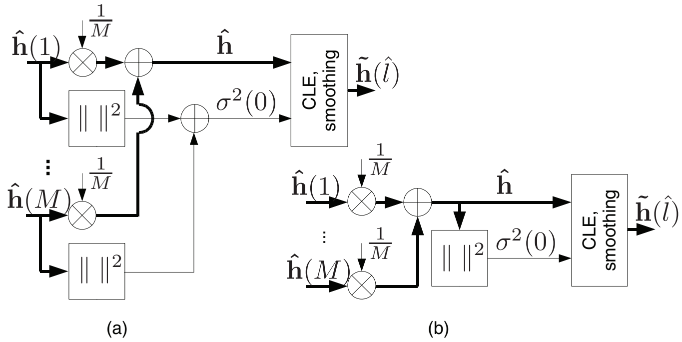

Fig. 1. Scheme of the processing applied to coarse CE. (a) refers to AIC CLE, (b) to AICc. The blocks denoted as CLE, smoothing implement the LA and AIC updates (12), (21)-(29). Bold lines concern vectors.

B. Embedding the AIC channel length estimation into the Levinson Algorithm

Regardless the simplifications introduced so far, the AIC burden seems still prohibitive, since *for every step* of the LA one has to compute i) the DFT of intermediate results $\tilde{\mathbf{h}}^{(l)}$, ii) $M \cdot K$ Euclidean distances (or K in case of AICc) among complex values, incurring in a complexity and latency much larger than the LA. In the following, we show that the modified RSS formulations (22)-(24) can be computed directly in the time domain, avoiding the FFT, and can be computed only once rather than L times, and easily updated at each step of the LA with a corrective term that resembles (12).

The RSS (22) can be rewritten, for any m, as

$$\begin{aligned}\sigma_m^2(l) &= \left\|\mathbf{A}\mathbf{W}^H\mathbf{r}(m)\right\|^2 + \left(\tilde{\mathbf{h}}^{(l)}\right)^H \mathbf{Q}_{1:l,1:l}\tilde{\mathbf{h}}^{(l)} \\ &\quad -2\Re\left(\left(\tilde{\mathbf{h}}^{(l)}\right)^H \left[\mathbf{I}_{l\times l} \quad \mathbf{0}_{l\times(N-l)}\right]\mathbf{W}\underbrace{\mathbf{P}^H\mathbf{A}}_{=\mathbf{P}^H}\mathbf{W}^H\mathbf{r}(m)\right) \\ &= \left\|\mathbf{A}\mathbf{W}^H\mathbf{r}(m)\right\|^2 + \left(\tilde{\mathbf{h}}^{(l)}\right)^H \hat{\mathbf{h}}_{1:l} \\ &\quad - 2\Re\left(\left(\tilde{\mathbf{h}}^{(l)}\right)^H \hat{\mathbf{h}}_{1:l}(m)\right) \qquad (25)\end{aligned}$$

Then, for linearity

$$\begin{aligned}\frac{1}{M}\sum_{m=1}^{M}\sigma_m^2(l) &= \frac{1}{M}\sum_{m=1}^{M}\left\|\mathbf{A}\mathbf{W}^H\mathbf{r}(m)\right\|^2 \\ &\quad + \left(\tilde{\mathbf{h}}^{(l)}\right)^H \hat{\mathbf{h}}_{1:l} - 2\Re\left(\left(\tilde{\mathbf{h}}^{(l)}\right)^H \frac{1}{M}\sum_{m=1}^{M}\hat{\mathbf{h}}_{1:l}(m)\right) \\ &= \underbrace{\frac{1}{M}\sum_{m=1}^{M}\left\|\mathbf{A}\mathbf{W}^H\mathbf{r}(m)\right\|^2}_{=\sigma^2(0)} - \left(\tilde{\mathbf{h}}^{(l)}\right)^H \hat{\mathbf{h}}_{1:l}\end{aligned} \qquad (26)$$

The first addend $\sigma^2(0)$ in the above expression is constant and can be computed at the beginning of the LA. It costs $2MK$ real multiplications in case of AIC and $2K$ real multiplications in case of AICc. Fig. 1a and 1b depict its computation in case of AIC and AICc, respectively.

Exploiting (12), the second addend of the above equation can be recursively updated:

$$\begin{aligned}\left(\tilde{\mathbf{h}}^{(l)}\right)^H \hat{\mathbf{h}}_{1:l} &= \left(\begin{bmatrix}\tilde{\mathbf{h}}^{(l-1)} \\ 0\end{bmatrix} + \left(\hat{h}_l - \epsilon_{\tilde{h}}^{(l)}\right)\mathbf{b}^{(l)}\right)^H \hat{\mathbf{h}}_{1:l} \\ &= \left(\tilde{\mathbf{h}}^{(l-1)}\right)^H \hat{\mathbf{h}}_{1:(l-1)} + \left(\hat{h}_l - \epsilon_{\tilde{h}}^{(l)}\right)^* \underbrace{\left(\mathbf{b}^{(l)}\right)^H \mathbf{Q}_{1:l,1:l}}_{=\left(\mathbf{u}^{(l)}\right)^H}\tilde{\mathbf{h}}^{(l)} \\ &= \left(\tilde{\mathbf{h}}^{(l-1)}\right)^H \hat{\mathbf{h}}_{1:(l-1)} + \left(\hat{h}_l - \epsilon_{\tilde{h}}^{(l)}\right)^* \tilde{h}_l^{(l)} \qquad (27)\end{aligned}$$

The above incremental update shows that *AIC and AICc complexity becomes affordable if the LA is adopted for the smoothing process, because the RSS at every step can be computed exploiting the LA by-product results* $\tilde{\mathbf{h}}$, $\hat{\mathbf{h}}$ and $\epsilon_{\tilde{h}}$. The update costs two *real* multiplications (since $\left(\tilde{\mathbf{h}}^{(l)}\right)^H \hat{\mathbf{h}}_{1:l}$ must be real) per recursion, i.e. $2L$ real multiplications, regardless of M both for AIC and AICc.

Finally, a similar sequential update can be performed in case of ZF coarse channel estimates:

$$\begin{aligned}\check{\sigma}_m^2(l) &= \left\|\check{\mathbf{P}}\mathbf{W}^H\mathbf{r}(m)\right\|^2 \\ &\quad + \left(\check{\tilde{\mathbf{h}}}^{(l)}\right)^H \check{\mathbf{h}}_{1:l} - 2\Re\left(\left(\check{\tilde{\mathbf{h}}}^{(l)}\right)^H \check{\mathbf{h}}_{1:l}(m)\right) \qquad (28)\end{aligned}$$

$$\frac{1}{M}\sum_{m=1}^{M}\check{\sigma}_m^2(l) = \underbrace{\frac{1}{M}\sum_{m=1}^{M}\left\|\check{\mathbf{P}}\mathbf{W}^H\mathbf{r}(m)\right\|^2}_{=\check{\sigma}^2(0)} - \left(\check{\tilde{\mathbf{h}}}^{(l)}\right)^H \check{\mathbf{h}}_{1:l} \qquad (29)$$

The above equations can be obtained following the footsteps of the derivation presented so far, and their proofs are omitted.

V. Simulation Results

In this Section, we provide numerical results for the performance of the proposed smoothing algorithm, driven by the AIC and AICc CLEs. We adopt many parameters from the IEEE 802.11a/p standards [1], [2]. In all the simulations, the number of active carriers is $K = 52$, out of the available $N = 64$. The guard interval length is $L-1 = 16$. We compare five different CE algorithms: coarse ZF CE, with an average of two tone-by-tone CEs in case of PACE; smoothed MF and ZF CEs with fixed channel length L; smoothed MF and ZF CEs with the AIC CLE (averaging the available RSSs); smoothed MF and ZF CEs with the AICc CLE (computing the RSS of the average coarse CE); smoothed MF and ZF CEs with a genie CLE, always choosing the correct channel length.

The channel PDP is exponentially decaying, and two power delay spreads have been tested: $\tau_{rms} = T_s$ and $\tau_{rms} = 2T_s$, where T_s is the sampling time (e.g., in [1] $T_s = 50$ns, in [2] $T_s = 100$ns).

For a complexity comparison, Table I summarizes the number of real multiplications of the above algorithms[3] (but

[3]Table I assumes a modified version of the AIC and AICc criteria, that does not require logarithms. $AIC(l)$ and $AICc(l)$ have been substituted by their exponentials: $\exp\left(AIC(l)\right) = \left(\frac{1}{M}\sum_{m=1}^{M}\sigma_m^2(l)\right)\exp\left(\frac{2l}{K}\right)$, $\exp\left(AICc(l)\right) = \sigma^2(l)\exp\left(\frac{K+l}{K-l-1}\right)$, $\exp\left(\widetilde{AICc}(l)\right) = \sigma^2(l)\exp\left(\frac{2l}{K-l-0.5}\right)$. The above modifications have no impact on performance. Since $\exp(\cdot)$ is monotonically increasing, each criterion and its exponential version are minimized by the same $\hat{l}$.

TABLE I
COMPLEXITY COMPARISON BETWEEN SEVERAL CE AND CLE METHODS. NUMBERS IN BRACKETS, AS WELL AS IN THE LAST COLUMN, REFER TO THE CASE $N = 64$, $K = 52$, $L = 17$.

Scenario	CE type	fft/ifft	smoothing	support algorithms	cle	total
PACE $M=2$ BPSK (ZF=ML)	Coarse	$2MN\log_2 N$ (1536)	0	0	0	1536
	Fixed L	$2(M+2)N\log_2 N$ (3072)	$2L^2$ (578)			3650
	Threshold				$2L$ (34)	3684
	LA+AICc				$2K+3L$ (155)	3805
	LA+AIC				$2MK+3L$ (259)	3909
	Straight AICc	$2(M+L+1)N\log_2 N$ (15360)	$2L^3$ (9826)		$L(2K+1)$ (1785)	26971
	Straight AIC				$L(2MK+1)$ (3553)	28739
DACE[a] $M=1$ QPSK (ZF=ML)	Coarse	$2N\log_2 N$ (768)	0	0	0	768
	Fixed L	$6N\log_2 N$ (2304)	$2L^2$ (578)			2882
	Threshold				$2L$ (34)	2916
	LA+AICc				$2K+3L$ (155)	3037
	LA+AIC					
	Straight AICc	$2(L+2)N\log_2 N$ (14592)	$2L^3$ (9826)		$L(2K+1)$ (1785)	26203
	Straight AIC					
DACE[a] $M=1$ 64-QAM ZF	Coarse	$2N\log_2 N+4K$ (976)	0	0	0	976
	Fixed L	$6N\log_2 N+4K$ (2512)	$2L^2$ (578)			3090
	Threshold				$2L$ (34)	3124
	LA+AICc				$2K+3L$ (155)	3245
	LA+AIC					
	Straight AICc	$2(L+2)N\log_2 N+4K$ (14800)	$2L^3$ (9826)		$L(6K+1)$ (5321)	29947
	Straight AIC					
DACE[a] $M=1$ 64-QAM MF	Coarse	$2N\log_2 N+4K$ (976)	0	0	0	976
	Fixed L	$6N\log_2 N+4K$ (2512)	$4L^2$ (1156)[b]	$2N\log_2 N$ (768)		4436[b]
	Threshold				$2L$ (34)	4470[b]
	LA+AICc		$4L^2$ (1156)	$2N\log_2 N+6L(L+1)$ (2604)	$2K+3L$ (155)	6427
	LA+AIC					
	Straight AICc	$2(L+2)N\log_2 N+4K$ (14800)	$4L^3$ (19652)[b]	$2N\log_2 N$ (768)	$L(6K+1)$ (5321)	40541[b]
	Straight AIC					

[a] To be repeated every OFDM symbol within the payload.
[b] Optimistic, it would involve a large memory burden.

for the unfeasible genie CLE). We also list the complexity of the AIC/AICc "straight" methods, not exploiting the deep connection with the LA, unveiled in this paper. Finally, we also report the complexity of CE methods exploiting thresholds to detect relevant channel taps [12].

A. Basic case: BPSK training sequences

With BPSK modulation and two training sequences (as in [1], [2]), the CE noise power is halved[4]. Fig. 2 and 3 refer to different τ_{rms} and CE algorithms introduced above. The subplots marked with (a) give the CE MSE (in dB) vs. $SNR = \frac{1}{N_0}$, while the subplots marked with (b) predict the impact of the noisy channel estimates on the overall receiver performance. As an example, in case of coarse ZF CE (red lines, square markers), the average of two LTS leads to an estimation noise 3 dB below N_0. This can be regarded as additional noise over the link, resulting in a system loss of $10\log_{10}\frac{N_0+N_0/2}{N_0} = 1.76$ dB. Similarly, we can derive all the other values of the right subplot.

The smoothing algorithm with fixed channel length L meets the expected performance, too. The estimation error $\mathbf{e}_{L\times 1} = \tilde{\mathbf{h}}_{L\times 1} - \begin{bmatrix}\mathbf{h}_{l\times 1}\\ \mathbf{0}_{(L-l)\times 1}\end{bmatrix}$ has covariance matrix

$$\begin{aligned}\mathbf{C_e} &= \frac{1}{M}E\Big[\mathbf{Q}^{-1}_{1:L,1:L}\left[\mathbf{I}_{L\times L} \quad \mathbf{0}_{L\times(N-L)}\right]\mathbf{W}\mathbf{P}^H\mathbf{W}^H \\ &\qquad \cdot\mathbf{n}\mathbf{n}^H\mathbf{W}\mathbf{P}\mathbf{W}^H\begin{bmatrix}\mathbf{I}_{L\times L}\\ \mathbf{0}_{(N-L)\times L}\end{bmatrix}\mathbf{Q}^{-H}_{1:L,1:L}\Big] \\ &= \frac{N_0}{M}\mathbf{Q}^{-1}_{1:L,1:L}\mathbf{Q}_{1:L,1:L}\mathbf{Q}^{-1}_{1:L,1:L} = \frac{N_0}{M}\mathbf{Q}^{-1}_{1:L,1:L} \end{aligned} \tag{30}$$

Considering the DFT noise covariance matrix, and averaging over the active carriers,

$$\begin{aligned}\sigma_e^2 &= \frac{N_0}{MK}\mathrm{tr}\left(\mathbf{A}\mathbf{W}^H\begin{bmatrix}\mathbf{Q}^{-1}_{1:L,1:L} & \mathbf{0}_{L\times(N-L)}\\ \mathbf{0}_{(N-L)\times L} & \mathbf{0}_{(N-L)\times(N-L)}\end{bmatrix}\mathbf{W}\right) \\ &= \frac{N_0}{MK}\mathrm{tr}\left(\mathbf{Q}\begin{bmatrix}\mathbf{Q}^{-1}_{1:L,1:L} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\right) = N_0\frac{L}{MK} \end{aligned} \tag{31}$$

where the third equality holds because the trace is invariant under cyclic permutations. In our case, $K = 52$ and $L = 17$, so we obtain an estimation error variance equal to $0.1635N_0$, i.e. $10\log_{10}(0.1635) = -7.87$ dB w.r.t. AWGN, as confirmed in Fig. 2 (magenta lines, triangle markers). In Fig. 3 the smoothing gain becomes smaller as the SNR increases. This is due to the tail of the PDP, that exceeds the guard interval. The tail power is the sum from $L+1$ to infinity of the (normalized) PDP $1 - \frac{\sum_{l=0}^{L}\exp\left(-\frac{l}{\tau_{rms}}\right)}{\sum_{l=0}^{\infty}\exp\left(-\frac{l}{\tau_{rms}}\right)} = \exp\left(-\frac{L+1}{\tau_{rms}}\right)$, introducing a CE error floor at $10\log_{10}\exp\left(-\frac{L}{\tau_{tms}}\right) \simeq -4.34\frac{L+1}{\tau_{rms}}$ dB, i.e. in our case -39 dB.

Comparing the performance of the different CEs, we see that even smoothing with fixed channel length L provides results remarkably better than coarse CE. At low SNRs,

[4] One could even optimize the pilot location and power, achieving better performance. Anyway, this is out of the scope of this paper, that focuses on efficient and effective CE and CLE. In many cases there is no freedom to modify pilots, as in [1] and [2].

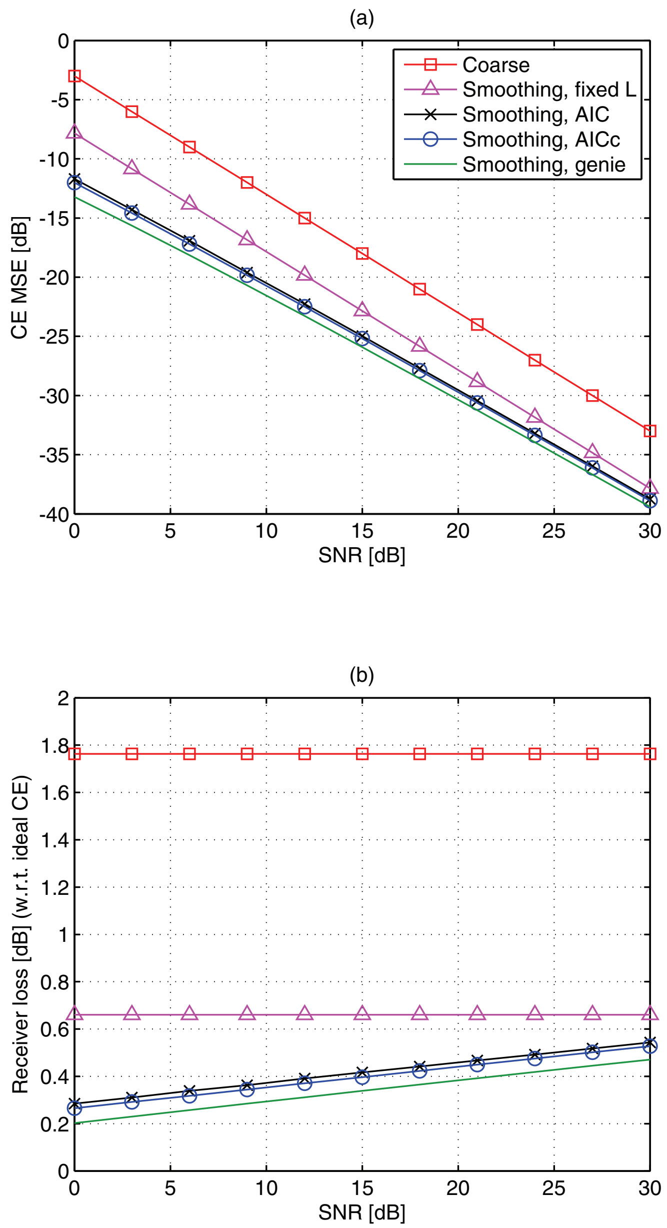

Fig. 2. A comparison among different PACEs (two LTSs), when the channel has exponential PDP. $N = 64$, $K = 52$, $L = 17$ and $\tau_{rms} = T_s$. Symbols are BPSK. Subplot (a) is the CE mean square error (MSE). Subplot (b) estimates the impact of the CE errors on the overall system performance, w.r.t. ideal CSI.

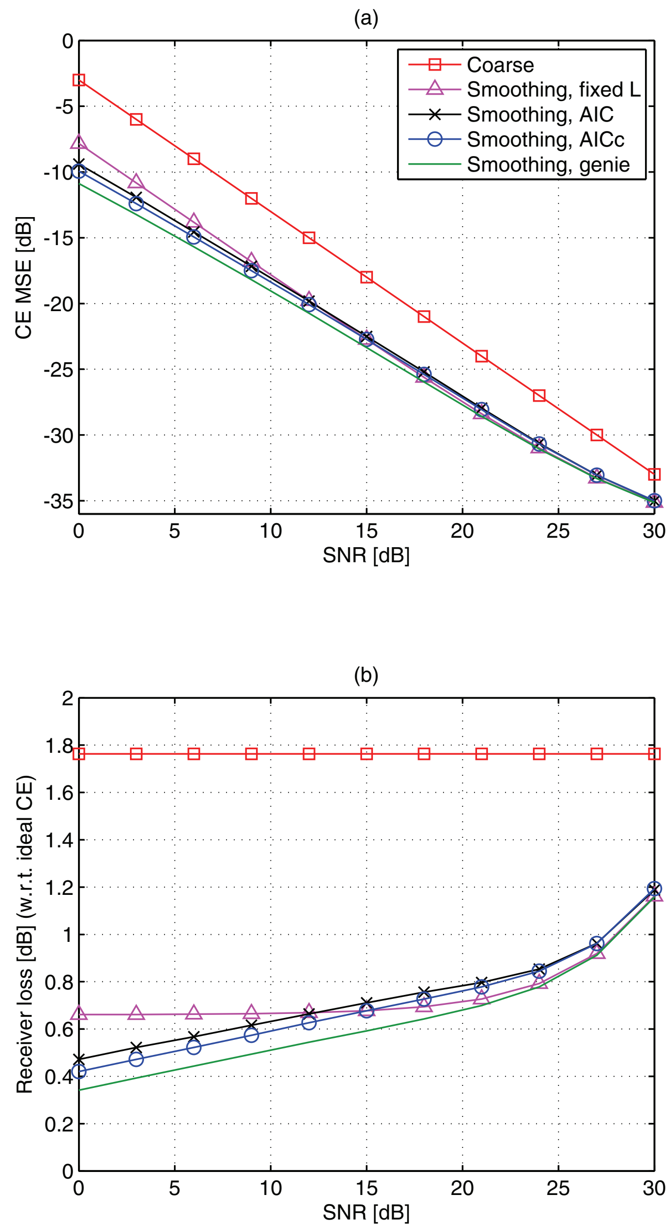

Fig. 3. A comparison among different PACEs (two LTSs), when the channel has exponential PDP. $N = 64$, $K = 52$, $L = 17$ and $\tau_{rms} = 2T_s$. Symbols are BPSK. Subplot (a) is the CE mean square error (MSE). Subplot (b) estimates the impact of the CE errors on the overall system performance, w.r.t. ideal CSI.

where a relevant fraction of noise can be discarded, AIC and AICc provide an improvement around 0.4 dB. This advantage disappears as the SNR increases, especially for AIC. AIC and AICc perform even better with shorter channel impulse responses. These results are not reported for lack of space.

B. AIC and channel tracking

DACE exploits data symbols to track the channel variations, as if they were new training sequences. In this paper, we are not interested in the Doppler effect on the system performance, thus we assume a relative speed between transmitter and receiver equal to 0 m/s. Nevertheless, we set the tracking algorithm on, so that we can evaluate the impact of the CE update process on the system performance. We consider two different update schemes run before each detection: 1) CE update based on the Viterbi decoder decisions (reencoded, interleaved and modulated again); 2) CE update based on the received symbols, hard-decided by the detector.

The former solution is more reliable, since the BER at the output of the channel decoder is remarkably smaller than at the input. Focusing on PER rather than BER, when decoded bits are mistaken at least one information bit is wrong, and this failure suffices to lose the entire packet, regardless of subsequent OFDM symbols. Thus, we can assume error-free data pilots and plot Fig. 4. The worse behavior w.r.t. PACE (-1.76 dB rather than -3 dB, see Fig. 2) can be explained by the lack of training repetition, which would have halved the CE noise power. Fig. 4 is meaningful only for PER, while

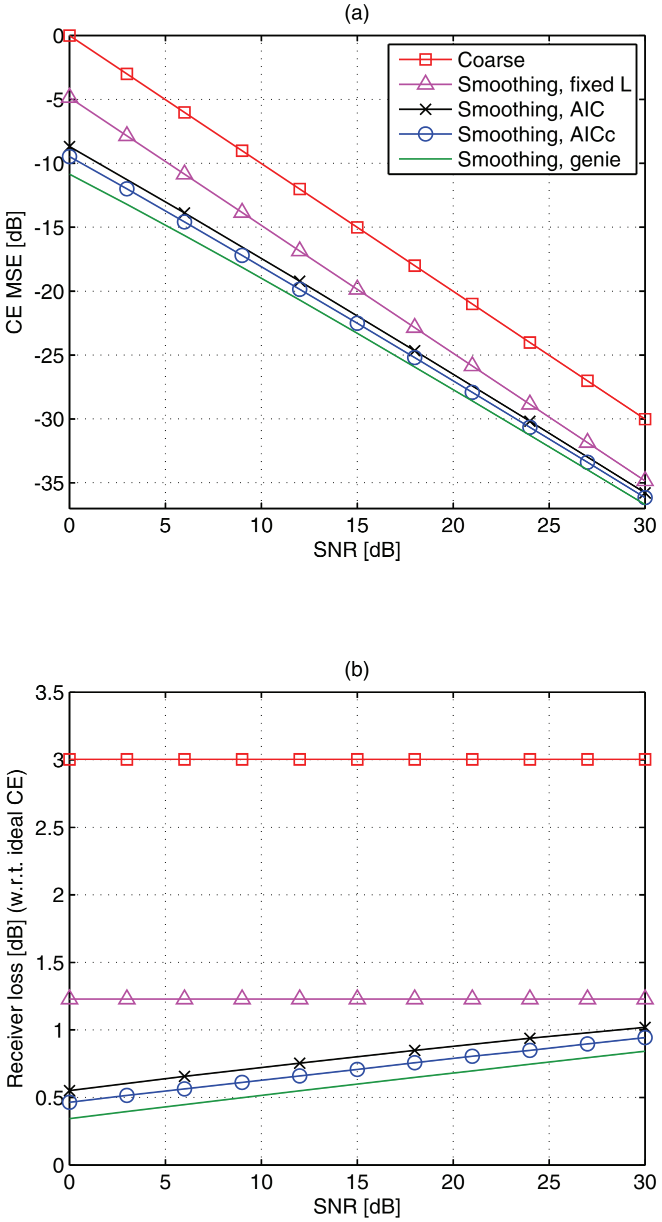

Fig. 4. A comparison among different DACEs, when the channel has exponential PDP. $N = 64$, $K = 52$, $L = 17$ and $\tau_{rms} = T_s$. Symbols are QPSK. Subplot (a) is the CE mean square error (MSE). Subplot (b) estimates the impact of the CE errors on the overall system performance, w.r.t. ideal CSI.

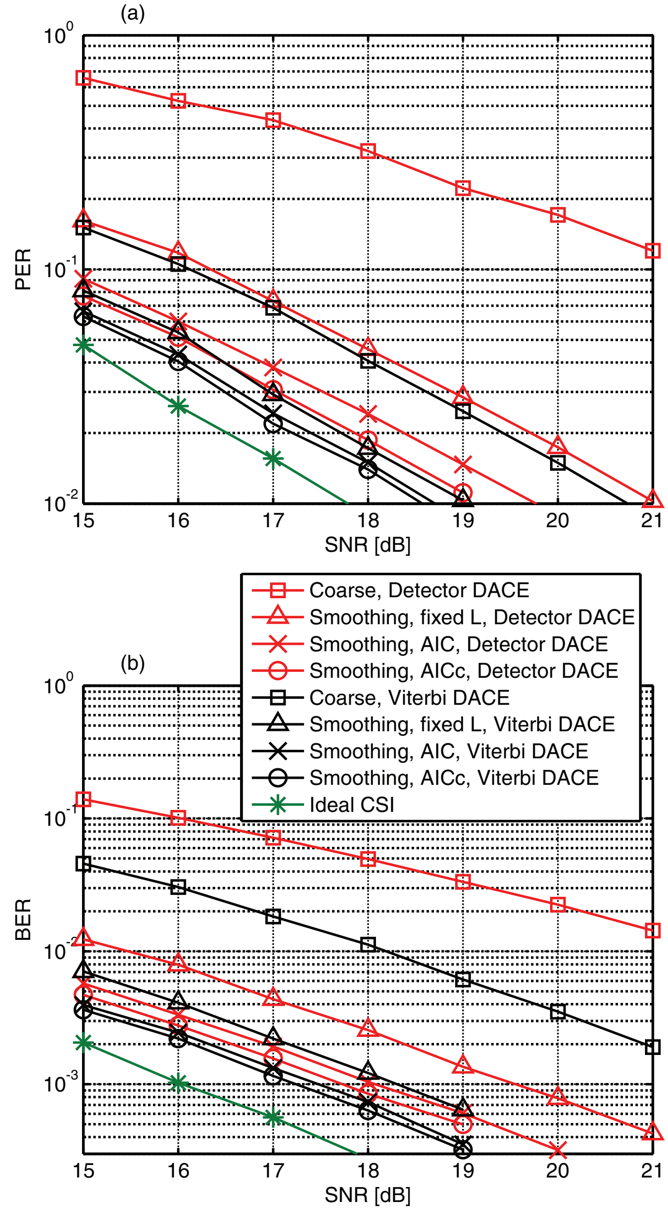

Fig. 5. System simulation, when $\tau_{rms} = T_s$. We consider data-aided channel estimations driven by the decoder or the detector, 40 OFDM symbols in the payload, QPSK modulation and a 64-states, $R_c = 0.75$ convolutional channel code. Subplot (a) estimates the PER, Subplot (b) the BER.

BER is expected to be worse. This is confirmed by simulations in Fig. 5 (black lines), where the gap between ideal CE and coarse CE is about 6 dB in case of BER (Fig. 5b), and "only" 3 dB in case of PER (Fig. 5a). In these simulations, we transmit 360 information bytes over 48 carriers (out of 52), corresponding to 40 OFDM symbols per packet. The target PER is 10^{-2}. The modulation is QPSK, the channel code is the standard 64-states, $R_c = 0.75$ convolutional code (octal polynomials: 133,171). The interleaver is borrowed from [1] and $\tau_{rms} = T_s$ as in Fig. 4.

When the mobile speed can be high, the delay of one OFDM symbol between the channel update and the detection process is harmful, since the estimations quickly become outdated. The channel estimate must be updated immediately, exploiting the detected symbols, despite their poor reliability. Even letting the speed equal to 0 m/s, the performance of such a system without countermeasures would be quite disappointing: in the case of Fig. 5a, approximately 8.5 dB worse than with ideal CE. Letting AIC choose shorter channel lengths becomes crucial, as smoothed frequency domain channel estimates reject not only part of the AWGN, but also those sparse CE errors due to erroneously detected symbols. As a consequence, the AIC CLE provides a gain larger than in other cases, around 1.3 dB. Besides, AICc is better than AIC by 0.5 dB, making the DACE driven by the detector a competitor of DACE driven by the decoder.

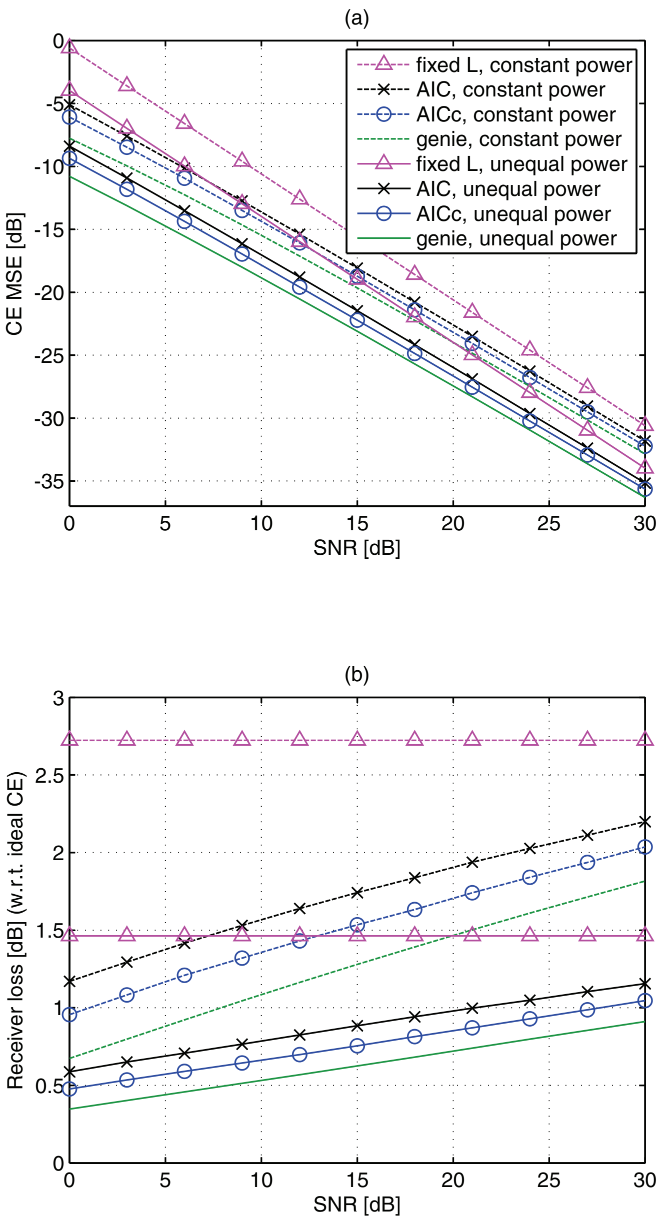

Fig. 6. A comparison among different DACEs, when the channel has exponential PDP. $N = 64$, $K = 52$, $L = 17$ and $\tau_{rms} = T_s$. Symbols are 64-QAM. Subplot (a) is the CE mean square error (MSE). Subplot (b) estimates the impact of the CE errors on the overall system performance, w.r.t. ideal CSI.

C. AIC and higher modulation orders

In this Subsection, we consider CE with QAM pilots. This case pertains to both PACE and DACE, thus all comments of the previous Subsections still hold. CE based on ZF matches the ML only for PSK constellations. For higher modulation orders the approach reported in Section II - performing a ZF estimation first and eventually smoothing it - is suboptimal. Figure 6 shows the impact of this approximation in case of a 64-QAM constellation. We can see that the exact LS and AIC perform better than the approximated LS and AIC, especially at high SNRs, where the system gain is about 1 dB, with $\tau_{rms} = T_s$. In this case a smoothing method based on matrix inversions would be quite unfeasible, since $\mathbf{Q}$ depends on the symbols power, and must be repeatedly inverted in real-time. The LA working in conjunction with the AIC estimator computes the solution without any inversion.

VI. Conclusions

In this paper, we have shown an efficient and effective way to perform joint channel length and channel estimation for OFDM systems, based on information criteria.

First, the Levinson recursion is used to solve the linear system related to the reduced-rank channel estimation problem. This method computes all the channel estimates up to the maximum allowed channel length, at the same cost as one single instance of the channel estimation algorithm, for the maximum channel length. Second, the AIC criterion is used to choose among these representations: only the model that loses the smallest amount of information is retained, implicitly determining the proper channel length.

Also the metric used to rank different models is obtained recursively, exploiting by-product results of the Levinson recursion. This leads to a small incremental complexity, w.r.t. a channel estimation with fixed, typically excessive, channel length.

Focusing on the performance of IEEE 802.11 OFDM-based wireless systems, the proposed scheme proves to be effective and flexible, suited in case of pilot aided and data aided channel estimation, as well as for PSK and QAM modulated symbols. Besides, the information criteria provide additional robustness to the system, since they aggressively reduce the estimated channel length whenever an additional, unexpected noise source arises, e.g. when erroneously detected symbols are used to track channel variations.

Finally, we remark that other criteria could be embedded in our joint channel length and CE framework, such as the Bayesian Information Criterion, not reported in this paper for lack of space. We tried also this criterion, but we did not find any advantage w.r.t. AIC.

References

[1] Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High-speed Physical Layer in the 5 GHZ Band, IEEE Std. 802.11a-1999, Sept. 1999.

[2] Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. Amendment 7: Wireless Access in Vehicular Environments, IEEE Std. 802.11pTM, June 2010.

[3] M. Engels, *Wireless OFDM Systems: How to Make Them Work?* Springer, 2002.

[4] L. Goldfeld, V. Lyandres, and D. Wulich, "Minimum BER power loading for OFDM in fading channel," *IEEE Trans. Commun.*, vol. 50, no. 11, p. 1729-1733, Nov. 2002.

[5] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[6] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplinkdownlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[7] Y. Li, N. Seshadri, and S. Ariyavisitakul, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 461–471, Mar. 1999.

[8] M. Siti, A. Assalini, E. Dall'Anese, and S. Pupolin, "Low complexity decision directed channel estimation based on a reliable symbol selection strategy for OFDM systems," in *Proc. 2010 IEEE Int. Conf. on Comm., Workshop on Vehicular Connectivity*, pp. 1–5.

[9] J.-J. van de Beek, O. Edfors, and M. Sandell, "On channel estimation in OFDM systems," in *Proc. 2005 IEEE Vehicular Techn. Conf.*, pp. 815–819.

[10] A. Tomasoni, E. Gallizio, and S. Bellini, "Low complexity and low latency training assisted channel estimation for MIMO-OFDM systems," in *Proc. 2007 IEEE Personal, Indoor and Mobile Radio Conf.*, pp. 1–5.

[11] K. Witrisal, "On estimating the RMS delay spread from the frequency-domain level crossing rate," *IEEE Commun. Lett.*, vol. 5, no. 7, pp. 287–289, July 2001.

[12] M. Jia, Z. Wang, and X. Gu, "Time domain channel and channel length estimation for OFDM system," in *Proc. 2007 IEEE Pac. Rim Conf. on Comm., Computers and Signal Proc.*, pp. 605–608.

[13] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automatic Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[14] T. Soderstrom and P. Stoica, *System Identification*, 1st edition. Prentice-Hall, 1989.

[15] E. G. Larsson, G. Liu, J. Li, and G. B. Giannakis, "Joint symbol timing and channel estimation for OFDM based WLANs," *IEEE Commun. Lett.*, vol. 5, no. 8, pp. 325–327, Aug. 2001.

[16] B. Yang, K. B. Letaief, R. S. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 467–479, Mar. 2001.

[17] K. Burnham and D. Anderson, "Multimodel inference: understanding AIC and BIC in model selection," *Sociological Methods & Research*, vol. 22, no. 2, pp. 260–301, Nov. 2004.

[18] N. Levinson, "The Wiener RMS (root mean square) error criterion in filter design and prediction," *Math and Phis.*, vol. 25, no. 4, pp. 261–278, 1947.

[19] J. G. Proakis, *Digital Communications*, 3rd edition. McGraw-Hill, 1995.

[20] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd edition. The Johns Hopkins Univesity Press, 1996.

[21] C. Hurvich and C. Tsai, "Regression and time series model selection in small samples," *Biometrika*, vol. 76, no. 2, pp. 297–307, June 1989.

[22] A.D.R. McQuarrie and C. Tsai, *Regression and Time Series Model Selection*, 2nd edition. World Scientific, 1999.

[23] J. Cavanaugh, "Unifying the derivations for the Akaike and corrected Akaike information criteria," *Statistics & Prob. Lett.*, vol. 33, no. 2, pp. 201–208, Apr. 1997.

Alessandro Tomasoni (M'09) was born in Milano, Italy, in 1980. He received the Telecommunications Engineering M.S. degree (cum laude) in 2005 and the Ph.D. degree in Information Engineering in 2009, from the Politecnico di Milano, Italy.

In 2012 he joined the Institute of Electronics, Computer and Telecommunication Engineering (IEIIT) of the Italian National Research Council (CNR), at the Dipartimento di Elettronica e Informazione (DEI) of the Politecnico di Milano, Italy. Between 2009 and 2011 he was Temporary Research Assistant at the same department. In 2008 he spent part of its Ph.D studies as a visiting researcher at Viterbi School of Engineering, University of Southern California, LA. He worked as consultant for leading industries, such as STMicroelectronics, Alcatel-Lucent and Numonyx (now Micron). He is a coauthor of several publications on journal papers and conference proceedings, as well as co-inventor of many patents. His main research interests are in wireless communications, optical communications and solid state data storage, with emphasis on information theory, advanced channel coding and channel estimation. Currently investigated topics are MIMO detection, Multi-User MIMO, LDPC codes and Power Line Communications. Dr. Tomasoni is a member of IEEE and IEEE Communications Society.

Devis Gatti was born in Italy, in 1975. He received the Electronic Engineering M.S. Degree in 2002 from the Pavia University, Italy. He made his Thesis Degree in Advanced System Technology (AST) research group, at STMicroelectronics, working on Turbo Codes for Satellite Communications. He is now Senior System Engineer in STM, dealing with Wireless Communication Systems and placing emphasis on digital signal processing of advanced detection, channel estimation, and channel coding techniques. He is the author or co-author of several patent applications and conference papers. Recently, he has extended his own activity research also to object tracking techniques for automotive RADAR applications. His current activity concerns Power Line communication topics for smart metering applications.

Sandro Bellini got a Dr. Ing. degree in Electronic Engineering (cum laude) in 1971 from the Politecnico di Milano, Italy. He began his research activity with the Italian National Research Council (CNR). In 1982 he got a position of associate professor. He is now full professor of Telecommunications, since 1990.

During these decades he has always had several research interests in the areas of telecommunication theory and digital signal processing. The main themes have been the analysis and design of pulse modulation systems, digital frequency modulation with continuous phase (CPM), efficient simulation techniques based on Importance Sampling, computer emission tomography, error correcting codes, multicarrier demodulation, channel equalization, blind equalization and deconvolution, symbol and carrier synchronization, digital data storage on optical support, and advanced coding techniques.

In the last years the research activity of Sandro Bellini has been mainly devoted to channel coding, with an emphasis on turbo codes, LDPC codes, and turbo product codes. In the last few years Sandro Bellini has also devoted some attention to MIMO wireless systems.

Marco Ferrari (M'06) was born in Milano, Italy, in 1971 and received the Laurea degree (M.S. equivalent) in Telecommunications Engineering (cum laude) and the Ph.D. degree in Electronics and Communication Engineering from the Politecnico di Milano, Italy, in 1996 and 2000, respectively.

Since 2001 he is a Researcher of the Institute of Electronics, Computer and Telecommunication Engineering (IEIIT) of the Italian National Research Council (CNR), at the Dipartimento di Elettronica e Informazione (DEI) of the Politecnico di Milano, Italy. In 2002 he was an EPRSC Research Fellow at the University of Plymouth (UK). He is coauthor in several scientific publications on leading international journals and conference proceedings and in a few patents. His main research interests are in wireless and optical digital transmission, information theory and channel coding. Dr. Ferrari is a member of IEEE and IEEE Communications Society.

Massimiliano Siti (M'06) was born in 1972. He received the Electrical Engineering M.S. degree from Politecnico di Milano in 1996. Since year 2000 he is with STMicroelectronics in Agrate Brianza (Italy). Previously (years 1998-2000) he joined Alcatel, working in an R&D optical laboratory. His background is in error correction codes and multiple antenna wireless systems (space-time coding and detection algorithms). His current research interests include communication systems for machine-to-machine communications (including smart grid applications), both wireless and power line. He is the author or co-author of several patent applications and technical papers. Mr. Siti was a visiting researcher in the wireless R&D laboratory at the University of California Los Angeles in 2004-2005, with the support of a Marie Curie fellowships obtained from the European Commission.