Performance of Iterative Data Detection and Channel Estimation for Single-Antenna and Multiple-Antennas Wireless Communications

Stefano Buzzi, *Member, IEEE*, Marco Lops, *Senior Member, IEEE*, and Stefania Sardellitti

***Abstract*—In iterative data-detection and channel-estimation algorithms, the channel estimator and the data detector recursively exchange information in order to improve the system performance. While a vast bulk of the available literature demonstrates the merits of iterative schemes through computer simulations, in this paper analytical results on the performance of an iterative detection/estimation scheme are presented. In particular, this paper focus is on uncoded systems and both the situations that the receiver and the transmitter are equipped with either a single antenna or multiple antennas are considered. With regard to the channel estimator, the analysis considers both the minimum mean square error and the maximum likelihood channel estimate, while, with regard to the data detector, linear receiver interfaces are considered. Closed-form formulas are given for the channel-estimation mean-square error and for its Cramér–Rao bound, as well as for the error probability of the data detector. Moreover, the problem of the optimal choice of the length of the training sequence is also addressed. Overall, results show that the considered iterative strategy achieves excellent performance and permits, at the price of some complexity increase, the use of very short training sequences without incurring any performance loss. Finally, computer simulations reveal that the experimental results are in perfect agreement with those predicted by the theoretical analysis.**

***Index Terms*—Channel estimation, Cramér–Rao bound (CRB), data detection, multiantenna systems, wireless communication.**

I. Introduction

THE USE OF iterative data-detection and channel-estimation schemes appears to be a suitable means to achieve excellent performance in wireless communication systems, where the length of the training sequence is to be kept as small as possible to increase data throughput and where the signal-to-noise ratio (SNR) may drop in deep channel fades [1]–[3]. Basically, in iterative schemes (which are usually denoted through the adjective "turbo"), the channel estimator and the data detector recursively exchange information in order to improve the system performance.

On the other hand, the explosive growth of wireless communications services, along with the birth of new applications, such as mobile computing and wireless Internet, as well as the deployment of wireless local area networks (WLAN), has resulted in an interest in bandwidth-efficient high-data-rate transmission systems. Recent results from information theory have shown that the capacity of a multiantenna wireless communication system operating in a rich scattering environment grows with a law approximately linear in the minimum between the number of transmit and receive antennas [4]–[6]. Likewise, high-performance space–time codes have been recently introduced, which permit us to achieve huge performance gains with respect to single-antenna communications systems (see, e.g., [7] and [8]).

In order to exploit the potential benefits that multiple-antenna systems promise in theory and their potential gains over single-antenna systems, some layered space–time architectures have been proposed. Among these, the most popular one has been termed *Bell Labs Layered Space–Time Architecture (BLAST)* [9], [10]. Layered space–time communication systems have attracted much attention in the recent years and several papers have appeared in the open literature, presenting theoretical findings and/or performance results for BLAST-like systems [5], [9]–[12]. In particular, the BLAST architecture is introduced and assessed in [9], [10], and [13], wherein the nulling and cancellation receiver is proposed and analyzed under the assumption that the frequency-flat fading channel is perfectly known to the receiver. In [14], some experimental results on the BLAST system are reported, while [12] proposed several data-detection/channel-equalization strategies for layered space–time communication systems operating over frequency-selective fading channels. Also, in this case, the propagation channel impulse response is assumed to be known at the receiver. The problem of channel estimation for BLAST systems has been tackled in [11], wherein the impact of channel-estimation errors on the outage probability of the nulling and cancellation receiver is evaluated. In [15], instead, the capacity of a multiple antenna wireless link versus the training data length is analyzed. It should be noted that much of the available literature on receiver design for BLAST-like systems shows results under the ideal assumption that the channel is known to the receiver. In practice, however, the channel gains are to be estimated at the receiver (based, e.g., on a known sequence of pilot symbols) and channel-estimation errors are to be accounted for when assessing the system performance. Likewise, the merits of iterative (turbo) detection algorithms are mainly demonstrated through computer-simulation results.

Manuscript received June 4, 2003; revised December 18, 2003 and February 19, 2004. This work was presented in part at the 2003 IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC '03), Rome, Italy, June 2003, and at the 37th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 2003.

The authors are with the Università degli Studi di Cassino, DAEIMI, I-03043 Cassino (FR), Italy (e-mail: buzzi@unicas.it; lops@unicas.it; s.sardellitti@unicas.it).

Digital Object Identifier 10.1109/TVT.2004.830144

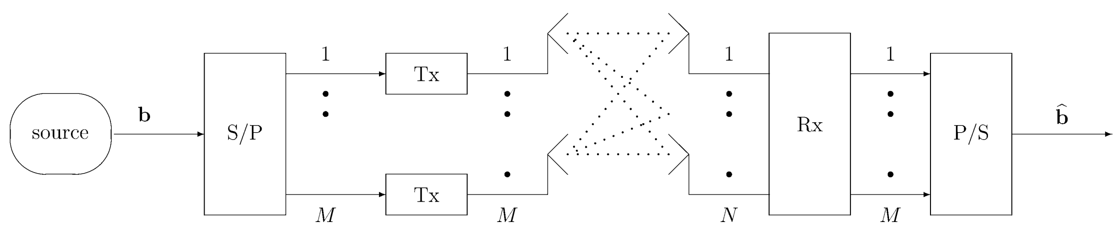

Fig. 1. Layered space–time wireless communication system diagram.

In this paper, we thus consider the problem of *joint* channel estimation and data detection in both single- antenna and multiantenna wireless system and propose and analyze a recursive data-detection/channel-estimation strategy based on an iterative exchange of information between the channel estimator and the data detector. We focus on uncoded transmission and assume that each transmitted data frame contains a preamble of known training symbols in order to perform channel estimation. Moreover, here we consider linear data-detection structures. The adoption of nonlinear detectors is an interesting topic that will possibly be considered in a future work and refer the reader to [16] and [17] for nonlinear iterative data-detection and channel-estimation algorithms based on the maximum-likelihood (ML) approach. The contributions of the present study may be summarized as follows.

- An iterative data-detection/channel-estimation procedure is introduced for uncoded systems. It is thus shown that uncoded systems may also benefit from the advantages of iterative data processing.
- Closed-form formulas for the channel-estimation mean square error (mse) are provided, both for the minimum mean square error (mmse) and ML channel-estimation algorithms. In particular, the channel-estimation mse is expressed as a simple function of the error probability at which the information symbols have been detected at the previous iteration.
- Cramér–Rao bounds (CRBs) on the variance of both biased and unbiased channel estimators are derived as a function of the error probability at which the information symbols have been detected at the previous iterations and also for the case that a prior probability density function (pdf) can be assigned to the parameter to be estimated (Bayesian CRBs).
- For single-antenna systems, an expression for the bit-error probability at a given iteration as a function of the bit-error probability at the previous iteration is provided.
- A graphical proof of the convergence of the proposed iterative scheme is provided.
- The problem of determining the optimal length of the training sequence is also investigated. In particular, since the training length is to be chosen as a compromise between the conflicting requirements of not reducing the system throughput and of reducing the channel-estimation mse, the problem of how to manage these opposite needs is examined.

For the sake of fairness, we note that techniques similar to our approach are reported in [18], which considers a multiantenna link equipped with convolutional coders and affected by frequency-selective fading and proposes a turbo-like iterative channel-estimation and data-detection procedure. Likewise, in [19], the iterative channel-estimation and data-detection procedure based on the expectation–maximization (EM) algorithm of [20] is extended to a multiantenna, possibly space–time coded, system. However, note that all of these works resort to extensive computer simulations to evaluate their algorithm performance, while in this paper a thorough theoretical analysis is provided.

The rest of this paper is organized as follows. In Section II the system and channel models are briefly illustrated, while in Section III the proposed iterative channel-estimation/data-detection procedure is described and details on the channel-estimators and data-detection structures are given. Section IV is devoted to the performance analysis of the channel estimators for both the single-antenna and the multiple-antennas scenarios and Section V contains the derivation of the CRBs. Section VI provides the analysis of the error probability for the single-antenna system, while Section VII is devoted to the graphical proof of the convergence of the proposed iterative strategy and to the discussion of the optimal setting of the length of the training sequence. Finally, concluding remarks are given in Section VIII.

II. System Model

Consider an uncoded single-user wireless communication system using a binary phase-shift keying (BPSK) modulation format.[1] Denote by M the number of transmitting antennas and by $N \geq M$ the number of receiving antennas.[2] At the transmitter, a data frame of length L is demultiplexed into M substreams of length L/M (we assume that L/M is an integer), to be transmitted one by each antenna at a data rate M times slower in the same frequency band; the same modulation format (i.e., BPSK) is used on each transmitting branch. A block diagram of the considered communication system is depicted in Fig. 1.

It is also assumed that the channel-coherence bandwidth is larger than the transmitted signal bandwidth and that the channel-coherence time is larger than the data-frame duration. Accordingly, the propagation channel introduces a flat fading in frequency and time and the classical discrete-time baseband-equivalent signal model for multiantenna systems subject to quasistatic flat fading is considered, i.e.,

$$\boldsymbol{r}(p) = A\boldsymbol{H}\boldsymbol{b}(p) + \boldsymbol{w}(p) \qquad p = 0, \ldots, \frac{L}{M} - 1 \qquad (1)$$

[1]Extending the following derivations to modulation schemes with larger cardinality is straightforward.

[2]The multiantenna scenario is illustrated here, since the single-antenna signal model is simply obtained by letting $N = M = 1$.

where $\boldsymbol{r}(p)$ is an N-dimensional vector obtained by stacking the matched-filtered and sampled outputs of the N receive antennas in the pth symbol interval. The $N \times M$-dimensional matrix $\boldsymbol{H}$ contains the channel-propagation coefficients. More precisely, the (n, m)th entry of $\boldsymbol{H}$, say $h_{n,m}$, is a complex gain accounting for the channel-propagation effects from the mth transmit antenna to the nth receive antenna. In keeping with the model in [9] and [10], this gain is modeled as a standard complex Gaussian random variate and $h_{n,m}$ and $h_{n',m'}$ are assumed to be statistically independent if either $n \neq n'$ or $m \neq m'$ (rich scattering environment). The constant A in (1) denotes the amplitude of the transmitted signal (note that the same power is transmitted on each antenna), while the M-dimensional column vector $\boldsymbol{b}(p) = [b_1(p),\ b_2(p),\ \ldots,\ b_M(p)]^T$, with $(\cdot)^T$ denoting transposition, contains the symbols transmitted in the pth symbol interval (with $b_m(p) \in \{-1, +1\}, \forall m, p$). Finally, it is also assumed that the additive thermal noise is uncorrelated across receive antennas (i.e., it is spatially white), thus implying that $\boldsymbol{w}(p)$ is a complex zero-mean Gaussian-random vector with covariance matrix $E\left\{\boldsymbol{w}(p)\boldsymbol{w}^H(p)\right\} = 2\mathcal{N}_0\boldsymbol{I}_N$, where $E\{\cdot\}$ and $(\cdot)^H$ denote statistical expectation and conjugate transposition, respectively, while $\boldsymbol{I}_N$ is the identity matrix of order N. Moreover, the vector $\boldsymbol{w}(p)$ is also assumed to be temporally white.

III. Iterative Data Detection and Channel Estimation

Assume that each transmitted packet contains T known training symbols (transmitted in the first T/M symbol intervals) and $L - T$ information symbols. At the receiver, first the training bits are exploited to obtain an estimate of the channel matrix, say $\widehat{\boldsymbol{H}}$, and this estimate is then exploited in the detection of the remaining $L - T$ information bits. The iterative strategy that is proposed here is based on the following idea. Once the $L - T$ information bits have been detected, they can be fed back, along with the first T training bits, to the channel estimator, which can treat them as a virtual training sequence of length L. Obviously, some of the $L - T$ detected bits may be erroneous; however, intuition suggests that, if the error probability is sufficiently low, the new channel estimate will expectedly be more reliable than the previous. Now, the new channel estimate can be fed back to the data detector and exploited to again detect the $L - T$ information bits with a lower error probability. Repeating this process several times expectedly yields some performance improvement on the system bit-error probability. In particular, the following analysis will show that this strategy is very effective and that few iterations (i.e., 2–3) suffice to achieve huge gains, even for very short training sequence lengths. A block scheme of this procedure is reported in Fig. 2. In the following, we briefly outline the channel estimators and data detectors that will be used in the analysis as building blocks of the scheme in Fig. 2. Thus, the rest of this section is not original, but is reported to make the paper self-contained and to ease equation referencing.

A. Channel Estimation

Define the following $N \times Q$-dimensional matrix:

$$\boldsymbol{R} = [\boldsymbol{r}(0), \ldots, \boldsymbol{r}(Q-1)] = A\boldsymbol{H}\boldsymbol{B} + \boldsymbol{W} \tag{2}$$

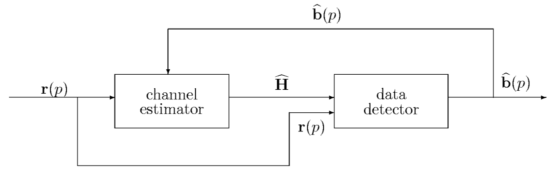

Fig. 2. Iterative channel-estimator and data-detector scheme.

where $\boldsymbol{W} = [\boldsymbol{w}(0), \ldots, \boldsymbol{w}(Q-1)]$ is $N \times Q$-dimensional, the quantity Q equals T/M at the 0th iteration and L/M at the following iterations. $\boldsymbol{B}$ is the following $M \times Q$-dimensional matrix:

$$\boldsymbol{B} = \begin{cases} [\boldsymbol{b}(0), \ldots, \boldsymbol{b}(Q-1)], & Q = \frac{T}{M} \\ \left[\boldsymbol{b}(0), \ldots, \boldsymbol{b}\left(\frac{T}{M}-1\right), \widehat{\boldsymbol{b}}\left(\frac{T}{M}\right), \ldots, \widehat{\boldsymbol{b}}(Q-1)\right], & Q = \frac{L}{M} \end{cases} \tag{3}$$

and $\widehat{\boldsymbol{b}}(T/M), \ldots, \widehat{\boldsymbol{b}}(Q-1)$ are the detected information bits in the symbol intervals $T/M, \ldots, Q-1$, respectively. Here we consider the ML and linear mmse channel estimators.

The ML estimate of the channel, say $\widehat{\boldsymbol{H}}_{\mathrm{ML}}$, can be obtained only if $Q \geq M$ [15] and assuming that the training bits matrix $\boldsymbol{B}$ is such that $\boldsymbol{B}\boldsymbol{B}^H$ is nonsingular. Indeed, if both these conditions are fulfilled, we can say that the channel is *identifiable* in the sense that, in the absence of noise, a perfect estimate of it can be recovered (see also [16] for general theorems on channel identifiability). Thus, assuming that the channel is identifiable, the ML estimate is written as

$$\widehat{\boldsymbol{H}}_{\mathrm{ML}} = \boldsymbol{R}\boldsymbol{B}^H\left(A\boldsymbol{B}\boldsymbol{B}^H\right)^{-1}. \tag{4}$$

The linear mmse estimate for the matrix $\boldsymbol{H}$ can instead be obtained as

$$\widehat{\boldsymbol{H}}_{\mathrm{mmse}} = A\boldsymbol{R}\left(A^2\boldsymbol{B}^H\boldsymbol{B} + 2\mathcal{N}_0\boldsymbol{I}_Q\right)^{-1}\boldsymbol{B}^H \tag{5}$$

with $\boldsymbol{I}_Q$ the identity matrix of order Q.

B. Data-Detection Structures

With regard to the data detector to be used in the outlined iterative strategy, we focus on linear receiver interfaces, particularly on the popular mmse and zero-forcing (ZF) receivers. Assuming that the channel estimate $\widehat{\boldsymbol{H}}$ is available at the receiver, in order to detect the symbols $\boldsymbol{b}(p)$ transmitted in the pth signaling interval, the vector $\boldsymbol{r}(p)$ in (1) is processed according to the decision rule

$$\widehat{\boldsymbol{b}}_{\mathrm{mmse}}(p) = \mathrm{sign}(\mathcal{R}\{A\widehat{\boldsymbol{H}}^H(A^2\widehat{\boldsymbol{H}}\widehat{\boldsymbol{H}}^H + 2\mathcal{N}_0\boldsymbol{I}_N)^{-1}\boldsymbol{r}(p)\}) \tag{6}$$

for the mmse strategy and

$$\widehat{\boldsymbol{b}}_{\mathrm{ZF}}(p) = \mathrm{sign}(\mathcal{R}\{\widehat{\boldsymbol{H}}^+\boldsymbol{r}(p)\}) \tag{7}$$

for the ZF strategy. In the above rules, sign$(\cdot)$ denotes the signum function, $\mathcal{R}(\cdot)$ denotes real part, and $(\cdot)^+$ denotes Moore–Penrose generalized inversion.[3]

[3]Note that, since $\widehat{\boldsymbol{H}}$ is a full-rank tall rectangular matrix, $\widehat{\boldsymbol{H}}^+ = (\widehat{\boldsymbol{H}}^H\widehat{\boldsymbol{H}})^{-1}\widehat{\boldsymbol{H}}^H$.

IV. Analysis of Channel-Estimation MSE

For the sake of clarity, the analysis of the channel-estimation mse is conducted first for the single-antenna scenario and then the results are generalized to systems with multiple antennas.

A. Single-Antenna Case: $N = M = 1$

If $N = M = 1$, the channel matrix $\boldsymbol{H}$ reduces to a scalar, say h, and we can define the following Q-dimensional observable column vector:

$$\boldsymbol{r} = Ah\boldsymbol{b} + \boldsymbol{w} \tag{8}$$

with $\boldsymbol{b}$ and $\boldsymbol{w}$ the transpose of the first row of the matrices $\boldsymbol{B}$ and $\boldsymbol{W}$, defined in Section III. The ML and mmse estimates of the channel h are now expressed as

$$\begin{aligned}\widehat{h}_{\mathrm{ML}} &= \frac{\boldsymbol{b}^H\boldsymbol{r}}{A\boldsymbol{b}^H\boldsymbol{b}} = \frac{\boldsymbol{b}^H\boldsymbol{r}}{AQ},\\ \widehat{h}_{\mathrm{mmse}} &= \frac{A\boldsymbol{b}^H\boldsymbol{r}}{2\mathcal{N}_0 + A^2\boldsymbol{b}^H\boldsymbol{b}} = \frac{A\boldsymbol{b}^H\boldsymbol{r}}{2\mathcal{N}_0 + A^2 Q}\end{aligned} \tag{9}$$

respectively.[4] Assuming that $Q = T$, i.e., we are at the 0th iteration, it is easily shown that the mean-squared errors are expressed as

$$\begin{aligned}E\left\{\left|h - \widehat{h}_{\mathrm{ML}}\right|^2\right\} &= \frac{2\mathcal{N}_0}{TA^2},\\ E\left\{\left|h - \widehat{h}_{\mathrm{mmse}}\right|^2\right\} &= \frac{2\mathcal{N}_0}{2\mathcal{N}_0 + A^2 T}.\end{aligned} \tag{10}$$

From (10), it is seen that the mean-squared estimation error is a decreasing function of the training bit length T or, equivalently, of the energy A^2T spent in the training phase. Let us now consider the case in which $Q = L$, i.e., the whole transmitted bit stream is fed back to the channel estimator in order to get a new channel estimate. Assuming that the $L-T$ information bits have been detected with an error probability $p(e)$, the vector $\boldsymbol{b}$ in (9) is now expressed as

$$\begin{aligned}\boldsymbol{b} =& [b(0), \ldots, b(T-1), b(T), \ldots, b(L-1)]^T\\ &+ \underbrace{\Big[\overbrace{0, \ldots, 0}^{T}, \overbrace{b_e(T), b_e(T+1), \ldots, b_e(L-1)}^{L-T}\Big]^T}_{=\boldsymbol{b}_e}\end{aligned} \tag{11}$$

where the random entries of the error vector $\boldsymbol{b}_e$ are defined as

$$b_e(p) = \begin{cases} 0, & \text{w.p. } 1 \quad \forall p = 0, \ldots, T-1\\ 0, & \text{w.p. } 1 - p(e) \ \forall p = T, \ldots, L-1\\ -2\mathrm{sign}(b(p)), & \text{w.p. } p(e) \quad \forall p = T, \ldots, L-1 \end{cases}. \tag{12}$$

Substituting expression (11) into (9) and evaluating the mean-squared channel-estimation errors leads, after some tedious manipulations, to the following expressions:[5]

$$E\left\{\left|h - \widehat{h}_{\mathrm{ML}}\right|^2\right\} = \frac{2\mathcal{N}_0}{LA^2} + \frac{4p(e)(L-T)}{L^2}\left[1 + (L-T-1)p(e)\right] \tag{13}$$

and

$$\begin{aligned}E\left\{\left|h - \widehat{h}_{\mathrm{mmse}}\right|^2\right\} =& 4A^4p(e)^2\left[\frac{(L-T-1)(L-T)}{(2\mathcal{N}_0 + A^2L)^2}\right] +\\ &+ 4A^2p(e)\left[\frac{(L-T)(2\mathcal{N}_0 + A^2)}{(2\mathcal{N}_0 + LA^2)^2}\right]\\ &+ \frac{2\mathcal{N}_0}{2\mathcal{N}_0 + LA^2}.\end{aligned} \tag{14}$$

Note that, as expected, the mean-squared channel-estimation error depends on the bit-error probability $p(e)$, with which the information bits have been detected. As $p(e)$ converges to zero, the decisions made are reliable training bits, whereby the mean-squared errors are

$$\begin{aligned}E\left\{\left|h - \widehat{h}_{\mathrm{ML}}\right|^2\right\} &\to \frac{2\mathcal{N}_0}{LA^2}\\ E\left\{\left|h - \widehat{h}_{\mathrm{mmse}}\right|^2\right\} &\to \frac{2\mathcal{N}_0}{2\mathcal{N}_0 + LA^2}.\end{aligned} \tag{15}$$

These relationships confirm that, as $p(e) \to 0$, all of the transmitted energy LA^2 contributes to the mean-squared-error reduction. Moreover, intuition suggests that there may exist some threshold value for the error probability, say $p(e)_{\mathrm{th}}$, such that use of the iterative procedure is beneficial (on the channel-estimation mean-squared error) only if the $L-T$ information bits are detected with an error probability lower than $p(e)_{\mathrm{th}}$. The value of $p(e)_{\mathrm{th}}$ for the ML estimate can be found by forcing the equality between (13) and the leftmost expression in (10). The solution is

$$[p(e)_{\mathrm{th}}]_{\mathrm{ML}} = \left(\frac{L}{2}\sqrt{\frac{1}{L^2} + \frac{2\mathcal{N}_0(L-T-1)}{(A^2LT)}} - \frac{1}{2}\right)(L-T-1)^{-1}. \tag{16}$$

Similarly, comparing (14) with the rightmost expression in (10) leads to

$$[p(e)_{\mathrm{th}}]_{\mathrm{mmse}} = \frac{\sqrt{(2\mathcal{N}_0+A^2)^2 + 2\mathcal{N}_0A^2(L-T-1)\frac{(2\mathcal{N}_0+A^2L)}{(2\mathcal{N}_0+A^2T)}} - (2\mathcal{N}_0+A^2)}{2A^2(L-T-1)}. \tag{17}$$

In Fig. 3, the threshold error probabilities $[p(e)_{\mathrm{th}}]_{\mathrm{ML}}$ and $[p(e)_{\mathrm{th}}]_{\mathrm{mmse}}$ are reported versus the average received energy

[4]In deriving $\widehat{h}_{\mathrm{mmse}}$, we have exploited the fact that h is a complex Gaussian-standard random variate.

[5]In deriving (13) and (14), we have assumed that the error vector $\boldsymbol{b}_e$ and the noise vector $\boldsymbol{w}$ are statistically independent. Computer simulations will show that this approximation has a very negligible effect on the final result.

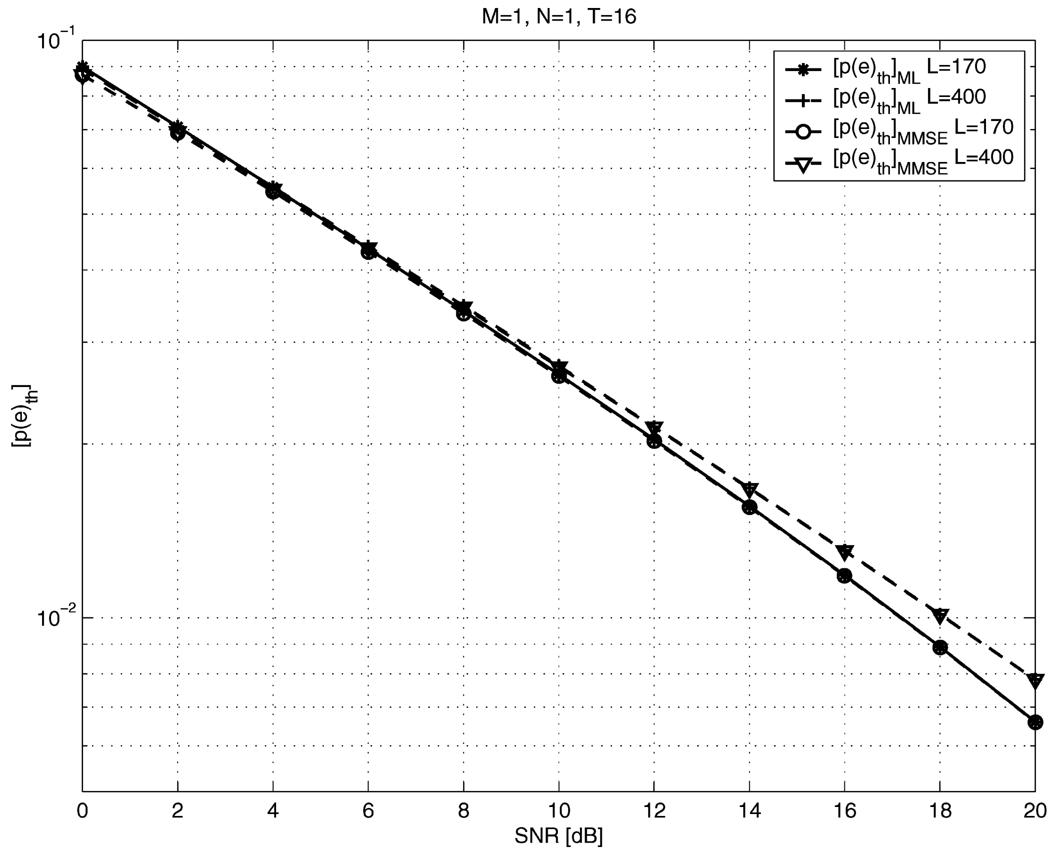

Fig. 3. Bit-error-rate threshold versus the SNR for two values of L in the single-antenna case.

contrast per bit SNR $= A^2/2\mathcal{N}_0$, in the case that $T = 16$ and $L = 170, 400$. Interestingly, it is seen that the threshold values increase with the packet length L and are in the interval $[10^{-2}, 10^{-1}]$ for SNR $\leq$ 17 dB. In Fig. 4, we instead report the mean-squared channel-estimation errors versus $p(e)$ at SNR $=$ 10 dB for both the cases of ML (upper plot) and mmse (lower plot) channel estimation. In particular, we report both simulation results and analytical formulas (13) and (14); moreover, in the same plot we also report the mean-squared channel-estimation errors for the cases in which either T or L training bits are exploited for channel estimation. Interestingly, the analytical formulas are in excellent agreement with the simulation results. Additionally, the errors in (13) and (14) intersect with the mean-squared channel-estimation error when T training bits are employed. The abscissa of the intersection point (which is approximately .028) obviously is the threshold-error probability: this is the same value that can be read in Fig. 3 for SNR $=$ 10 dB. As already commented, as $p(e)$ decreases, the channel-estimation errors converge to the error in the situation that the whole L-bit packet consists of known training bits.

B. Multiantenna Case: $N, M > 1$

Consider now the multiple-antennas scenario. Unfortunately, a direct generalization of the analysis presented for the single-antenna case is hardly tractable unless some further approximations are invoked. In the following, we will be approximating some fourth-order statistics of the matrix $\boldsymbol{B}$ in terms of second-order moments.[6] To begin, we consider the ML channel estimate (4) and assume that only the first T training bits are used by the channel estimator (i.e., $Q = T/M$). Simple algebraic manipulations lead to the following expression of the mean-squared channel-estimation error:

$$E\left\{\left\|\boldsymbol{H}-\widehat{\boldsymbol{H}}_{\mathrm{ML}}\right\|_F^2\right\}=\frac{2\mathcal{N}_0 N}{A^2}\mathrm{trace}\left[\left(\boldsymbol{B}\boldsymbol{B}^H\right)^{-1}\right]\approx\frac{2\mathcal{N}_0 N M^2}{TA^2} \quad (18)$$

where the rightmost approximate equality stems from the approximate relation[7] $(\boldsymbol{B}\boldsymbol{B}^H) \approx (T/M)\boldsymbol{I}_M$. With regard, instead, to the mmse channel estimate (5) with $Q = T/M$, letting

$$\boldsymbol{G}=\boldsymbol{B}A^2(A^2\boldsymbol{B}^H\boldsymbol{B}+2\mathcal{N}_0\boldsymbol{I}_{T/M})^{-1}\boldsymbol{B}^H\approx A^2\frac{T}{(TA^2+2\mathcal{N}_0 M)}\boldsymbol{I}_M \quad (19)$$

we have

$$\begin{aligned} &E\{\|\boldsymbol{H}-\widehat{\boldsymbol{H}}_{\mathrm{mmse}}\|_F^2\} \\ &= E\{\|\boldsymbol{H}\|_F^2\} - \mathrm{trace}(2\mathcal{R}(E\{\boldsymbol{H}^H\boldsymbol{H}\}\boldsymbol{G}))+ \\ &\quad + \mathrm{trace}(\boldsymbol{G}^2 E\{\boldsymbol{H}^H\boldsymbol{H}\}) \\ &\quad + \mathrm{trace}(2\mathcal{N}_0 N\boldsymbol{B}A^2(A^2\boldsymbol{B}^H\boldsymbol{B}+2\mathcal{N}_0\boldsymbol{I}_{T/M})^{-2}\boldsymbol{B}^H) \\ &\approx \frac{2\mathcal{N}_0 M^2 N}{(2\mathcal{N}_0 M + TA^2)} \end{aligned} \quad (20)$$

[6]This is a usual assumption in the analysis of adaptive filtering algorithms.
[7]Note that this relation is almost surely fulfilled as T/M grows large.

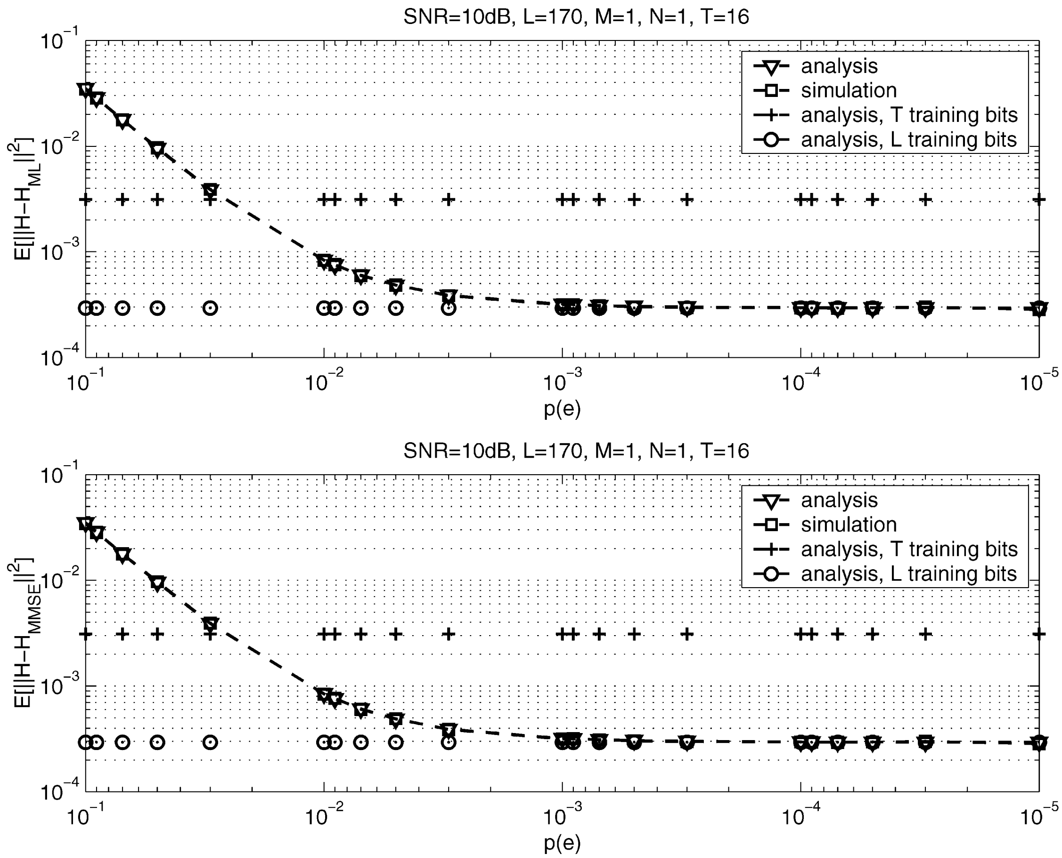

Fig. 4. Mean-squared channel-estimation error versus $\mathrm{p(e)}$ in the single-antenna case.

where the approximation

$$\boldsymbol{B}A^2(A^2\boldsymbol{B}^H\boldsymbol{B}+2\mathcal{N}_0\boldsymbol{I}_{T/M})^{-2}\boldsymbol{B}^H \approx A^2\frac{TM}{(TA^2+2\mathcal{N}_0M)^2}\boldsymbol{I}_M$$

was used. Also note that, for $N=M=1$, (18) and (20) reduce to that of (10).

Let us now consider the case that $Q=L/M$. Again assuming that the information bits have been detected with error probability $p(e)$, the matrix $\boldsymbol{B}$ in (2) is now expressed as

$$\boldsymbol{B}=\left[\boldsymbol{b}(0),\ldots,\boldsymbol{b}\left(\frac{L}{M}-1\right)\right]+\underbrace{\left[\boldsymbol{O}_{M,T/M},\boldsymbol{b}_e\left(\frac{T}{M}\right),\ldots,\boldsymbol{b}_e\left(\frac{L}{M}-1\right)\right]}_{=\boldsymbol{B}_e} \tag{21}$$

where $\boldsymbol{O}_{M,T/M}$ is an $M \times T/M$-dimensional matrix with all-zero entries, while the jth entry of the vector $\boldsymbol{b}_e(p)$ is a discrete random variate equal to 0 w.p. $(1-p(e))$ and equal to $-2\mathrm{sgn}([\boldsymbol{b}(p)]_j)$ w.p. $p(e)$, $\forall p = T/M,\ldots,L/M-1$, and $\forall j = 1,\ldots,M$. Substituting (21) into (4) and evaluating the mean-squared channel-estimation error, we have (22), shown at the bottom of the page. In order to simplify this expression, we now introduce the following approximations:[8]

$$(\boldsymbol{B}\boldsymbol{B}^H)^{-1} \approx \frac{M}{L}\boldsymbol{I}_M, \text{ and } (\boldsymbol{B}\boldsymbol{B}^H)^{-1}(\boldsymbol{B}\boldsymbol{B}^H)^{-1} \approx \frac{M^2}{L^2}\boldsymbol{I}_M. \tag{23}$$

We also approximate the matrix $(\boldsymbol{B}-\boldsymbol{B}_e)\boldsymbol{B}^H$ with a diagonal matrix whose jth diagonal element is written as

$$\frac{L}{M}+\sum_{i=T/M+1}^{L/M}(\boldsymbol{B}(j,i)-\boldsymbol{B}_e(j,i))\boldsymbol{B}_e(i,j). \tag{24}$$

These approximations finally lead to

$$E\{\|\boldsymbol{H}-\widehat{\boldsymbol{H}}_{\mathrm{ML}}\|_F^2\} \approx 4p(e)M^2N\frac{(L-T)}{L^2} + 4p(e)^2MN\frac{(L-T-M)(L-T)}{L^2}+\frac{2\mathcal{N}_0NM^2}{LA^2} \tag{25}$$

[8]Note that these relations are almost surely fulfilled as L/M grows large.

$$\begin{aligned}E\{\|\boldsymbol{H}-\widehat{\boldsymbol{H}}_{\mathrm{ML}}\|_F^2\} =& E\{\|\boldsymbol{H}\|_F^2\}\\ &-\mathrm{trace}(2\mathcal{R}(E\{(\boldsymbol{B}-\boldsymbol{B}_e)\boldsymbol{B}^H(\boldsymbol{B}\boldsymbol{B}^H)^{-1}\boldsymbol{H}^H\boldsymbol{H}\}))\\ &+\mathrm{trace}(E\{\boldsymbol{H}^H\boldsymbol{H}(\boldsymbol{B}-\boldsymbol{B}_e)\boldsymbol{B}^H(\boldsymbol{B}\boldsymbol{B}^H)^{-2}\boldsymbol{B}(\boldsymbol{B}-\boldsymbol{B}_e)^H\}\\ &+\boldsymbol{W}\boldsymbol{B}^H(\boldsymbol{B}\boldsymbol{B}^HA)^{-2}\boldsymbol{B}\boldsymbol{W}^H).\end{aligned} \tag{22}$$

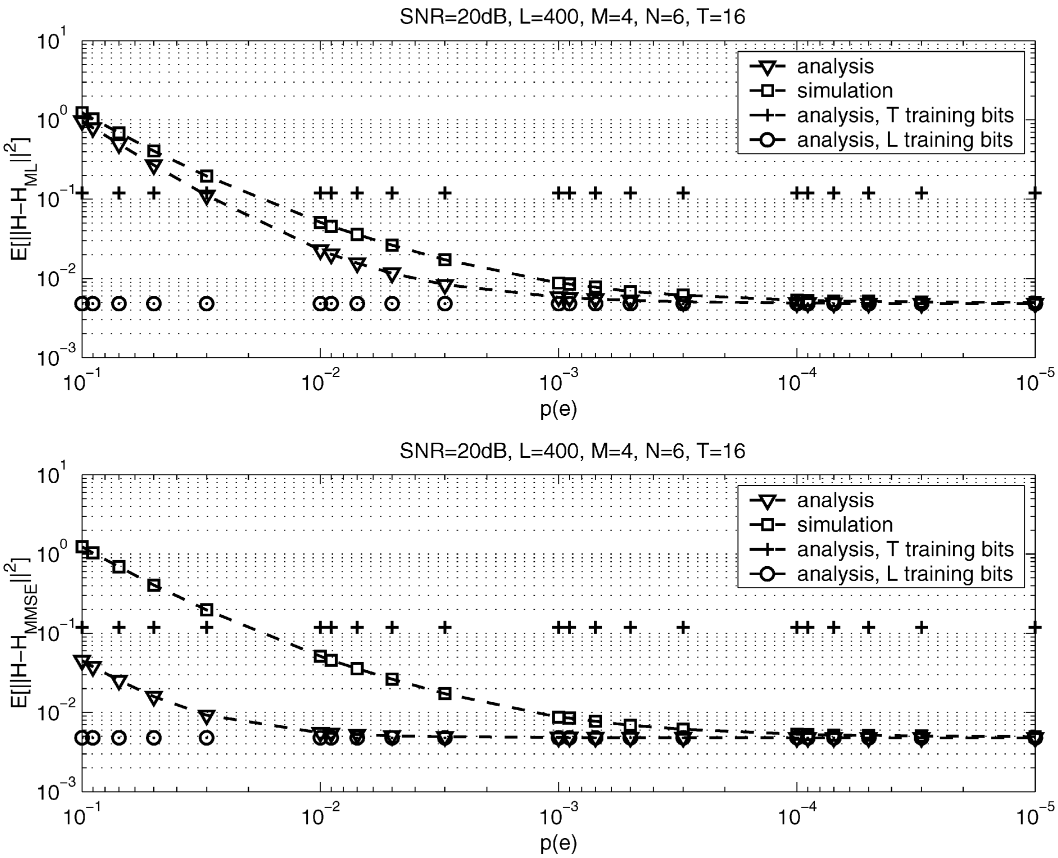

Fig. 5. Mean-squared channel-estimation error versus $\mathrm{p(e)}$ in the multiantenna case.

wherein the fact that $E\{\|\boldsymbol{H}\|_F^2\} = MN$ has been exploited. Note that, as $p(e) \to 0$, we have

$$E\{\|\boldsymbol{H} - \widehat{\boldsymbol{H}}_{\mathrm{ML}}\|_F^2\} = \frac{2\mathcal{N}_0 N M^2}{LA^2} \tag{26}$$

i.e., also in this case, the mean-squared channel-estimation error is lower bounded by the error achieved in the limiting situation that the whole L-bits packet is perfectly known at the channel estimator.

Let us now finally consider the mmse channel estimate. Substituting (21) into (5) and evaluating the mean-squared channel-estimation error gives us (27), shown at the bottom of the page, with $\boldsymbol{U} = A^2(\boldsymbol{B} - \boldsymbol{B}_e)(A^2\boldsymbol{B}^H\boldsymbol{B} + 2\mathcal{N}_0\boldsymbol{I}_{L/M})^{-1}\boldsymbol{B}^H$. Now, exploiting approximations (23) and (24), after some straightforward, but not trivial, algebraic manipulations, (27) is finally rewritten as

$$\begin{aligned} &E\{\|\boldsymbol{H} - \widehat{\boldsymbol{H}}_{\mathrm{mmse}}\|_F^2\} \\ &\quad\approx \frac{2\mathcal{N}_0 M^2 N}{(A^2L + 2\mathcal{N}_0 M)} + p(e)\frac{A^2}{(A^2L + 2\mathcal{N}_0 M)} \\ &\quad\cdot\left[4(L-T) + \frac{4A^2(L-T)(M-L)}{(A^2L + 2\mathcal{N}_0 M)}\right] \\ &\quad+ 4p(e)^2\frac{A^4(L-T)(L-T-M)}{(A^2L + 2\mathcal{N}_0 M)^2}. \end{aligned} \tag{28}$$

Again, we note that as $p(e) \to 0$, we obtain the same mean-squared channel-estimation error as in (20) with T replaced by L; moreover, if we let $N = M = 1$, (25) and (28) reduce to expressions (13) and (14), respectively. In Fig. 5, we report the mean-squared channel-estimation errors (25) and (28) versus $p(e)$. In the same plots, we also report these mean-squared channel-estimation errors obtained through computer simulations, along with the mean-squared errors corresponding to the situation that either T or L training bits are employed for channel estimation. A system with $M = 4$ transmit antennas and $N = 6$ receiving antennas is considered; the training length is $T = M^2 = 16$ and $L = 400$; the SNR is 20 dB. First of all, note that the analytical performance is in good agreement with the simulation results, even though, in regard to mmse channel estimation, the analytical results

$$\begin{aligned} E\{\|\boldsymbol{H} - \widehat{\boldsymbol{H}}_{\mathrm{mmse}}\|_F^2\} = &E\{\|\boldsymbol{H}\|_F^2\} - \mathrm{trace}(2\mathcal{R}(E\{\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{U}\})) \\ &+ \mathrm{trace}(E\{\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{U}\boldsymbol{U}^H\}) \\ &+ \mathrm{trace}(E\{(2\mathcal{N}_0 N\boldsymbol{B}A^2(A^2\boldsymbol{B}^H\boldsymbol{B} + 2\mathcal{N}_0\boldsymbol{I}_{L/M})^{-2}\boldsymbol{B}^H)\}) \end{aligned} \tag{27}$$

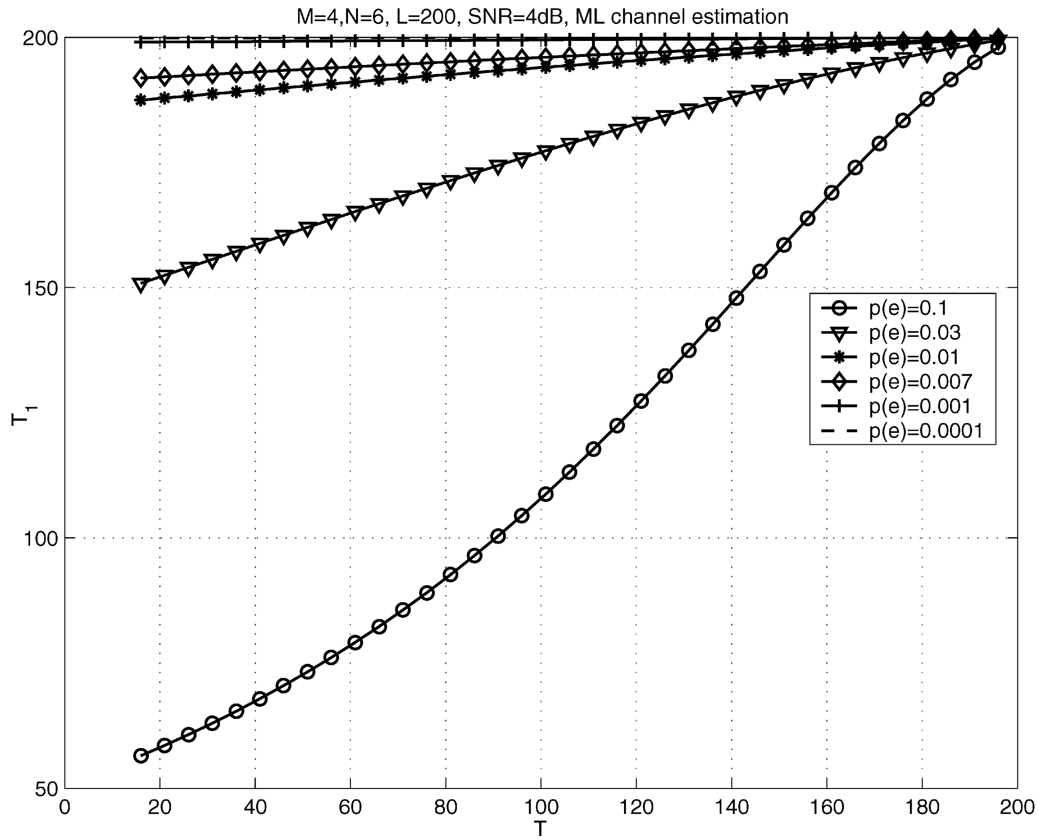

Fig. 6. T_1 versus T for several values of $\mathrm{p(e)}$.

are revealed to be not accurate for large values of $p(e)$. Additionally, also in this case, a threshold-error probability can be devised and, as $p(e) \to 0$, the mean-squared errors converge to the bound given by (20) with T replaced by L.

Comparing the two equations in (10) with (13) and (14), and (18) and (20) with (25) and (28), respectively, a relationship between the training lengths for the iterative and noniterative strategies that achieve the same performance can be established. In particular, assume that, for ML channel estimation, T_1 training bits are adopted in a noniterative strategy and that T training bits are adopted in the iterative strategy. Equating the corresponding channel estimation mses (18) and (25) with T_1 in place of T in (18) and solving for T_1, we have

$$T_1 = \frac{2\mathcal{N}_0 L^2}{A^2 4p(e)(1-p(e))(L-T) + \frac{A^2 4p(e)^2(L-T)^2}{M} + 2\mathcal{N}_0 L} \tag{29}$$

which provides a relation between the training lengths that achieve the same channel-estimation mse for the iterative and noniterative strategies. Fig. 6 reports (29) for several values of $p(e)$ and for a system with $M = 4$, $N = 6$, $L = 200$, and $\mathrm{SNR} = 4$ dB. It is seen that, the lower the $p(e)$, the bigger the saving in the training length brought by the iterative strategy. Interestingly, results show that even an as large as 0.1 error probability permits reducing the training length T. Note also that, as $p(e) \to 0$, from (29) we have $T_1 = L$; namely, the iterative strategy achieves the same channel-estimation mse that would be achieved if the data frame contained only training bits.

V. Derivation of the CRBs

In the following, we will derive the CRB for both unbiased and biased estimators; moreover, the Bayesian CRB, wherein the parameter to be estimated is a random variate with known pdf, will be also addressed.

A. CRB for the Single-Antenna Case: $N = M = 1$

Consider again the signal model in (8). Denoting by $\widehat{\boldsymbol{b}}$ the L-dimensional vector containing the T training bits and the $L-T$ detected information bits that are fed back to the channel estimator at a given iteration, the vector $\boldsymbol{r}$ in (8) can be written as

$$\boldsymbol{r} = Ah\boldsymbol{b} + \boldsymbol{w} = Ah\widehat{\boldsymbol{b}} + \underbrace{\boldsymbol{w} + Ah(\boldsymbol{b} - \widehat{\boldsymbol{b}})}_{\boldsymbol{n}}. \tag{30}$$

Strictly speaking, the vector $\boldsymbol{n}$ is the sum of a Gaussian vector and of a discrete-valued random vector; in the following, however, in order to derive the CRB, we assume that $\boldsymbol{n}$ follows a Gaussian distribution. This will result in an approximated CRB. Note, however, that for $p(e) \to 0$ the vector $\boldsymbol{n}$ converges to a Gaussian distribution, thus implying that our results will be accurate in the region of interest of low error probabilities. Accordingly, given the parameter to be estimated h and the detected bits $\widehat{\boldsymbol{b}}$, the L-dimensional vector $\boldsymbol{r}$ is conditionally Gaussian and its pdf is written as

$$p(\boldsymbol{r}|h, \widehat{\boldsymbol{b}}) = \frac{1}{\det(\pi \boldsymbol{C}(h))} e^{-[(\boldsymbol{r}-\boldsymbol{\mu_r}(h))^H \boldsymbol{C}^{-1}(h)(\boldsymbol{r}-\boldsymbol{\mu_r}(h))]}. \tag{31}$$

Since the entries of $\boldsymbol{b}$ have the mass probability distribution

$$b(p) = \begin{cases} \widehat{b}(p), & \text{w.p. } 1 \quad \forall p = 0, \dots, T-1, \\ \widehat{b}(p), & \text{w.p. } 1 - p(e) \quad \forall p = T, \dots, L-1, \\ -\widehat{b}(p), & \text{w.p. } p(e) \quad \forall p = T, \dots, L-1 \end{cases} \quad (32)$$

it is easily shown that

$$\boldsymbol{\mu_r}(h) = E_{\boldsymbol{b}}[Ah\widehat{\boldsymbol{b}} + Ah(\boldsymbol{b} - \widehat{\boldsymbol{b}})] = Ah\widetilde{\boldsymbol{b}} \quad (33)$$

with $\widetilde{\boldsymbol{b}}$ defined as

$$\widetilde{\boldsymbol{b}} = [\widehat{b}(0), \dots, \widehat{b}(T-1), \widehat{b}(T)(1-2p(e)), \widehat{b}(T+1)(1-2p(e)), \dots, \widehat{b}(L-1)(1-2p(e))]^T \quad (34)$$

and that

$$\begin{aligned} \boldsymbol{C}(h) &= E\{[\boldsymbol{r} - E\{\boldsymbol{r}\}][\boldsymbol{r} - E\{\boldsymbol{r}\}]^H\} \\ &= 2\mathcal{N}_0 \boldsymbol{I}_L + A^2|h|^2 E_{\boldsymbol{b}}\{(\boldsymbol{b} - \widehat{\boldsymbol{b}})(\boldsymbol{b} - \widehat{\boldsymbol{b}})^H\} \\ &\quad - 4A^2|h|^2 p(e)^2 \overline{\boldsymbol{b}}\overline{\boldsymbol{b}}^H \end{aligned} \quad (35)$$

with

$$\overline{\boldsymbol{b}} = \Big[\overbrace{0, \dots, 0}^{T}, \overbrace{\widehat{b}(T), \widehat{b}(T+1), \dots, \widehat{b}(L-1)}^{L-T}\Big]^T. \quad (36)$$

Since the off-diagonal entries of the matrix $E_{\boldsymbol{b}}\{(\boldsymbol{b}-\widehat{\boldsymbol{b}})(\boldsymbol{b} - \widehat{\boldsymbol{b}})^H\}$ are proportional to $p(e)^2$, while its nonzero diagonal entries are proportional to $p(e)$, for a sufficiently small error probability the following approximation holds:

$$E_{\boldsymbol{b}}\{(\boldsymbol{b} - \widehat{\boldsymbol{b}})(\boldsymbol{b} - \widehat{\boldsymbol{b}})^H\} \approx \begin{bmatrix} \boldsymbol{O}_{T,T} & \boldsymbol{O}_{T,L-T} \\ \boldsymbol{O}_{L-T,T} & 4p(e)\boldsymbol{I}_{L-T} \end{bmatrix}. \quad (37)$$

Likewise, we can neglect the off-diagonal terms of the rightmost summand in (35), so as to obtain

$$\boldsymbol{C}(h) \approx \begin{bmatrix} 2\mathcal{N}_0 \boldsymbol{I}_T & \boldsymbol{O}_{T,L-T} \\ \boldsymbol{O}_{L-T,T} & (2\mathcal{N}_0 + A^2|h|^2\sigma_{e_b}^2)\boldsymbol{I}_{L-T} \end{bmatrix} \quad (38)$$

where we have let $\sigma_{e_b}^2 = 4p(e)(1 - p(e))$.

Now, (31), (33), and (38) are what suffices to derive the CRB. We start by considering the bound for unbiased estimators. First, note that since the parameter to be estimated is complex, it is useful to consider the real two-dimensional (2-D) vector containing its real part and the coefficient of the imaginary part

$$\boldsymbol{h} = [h_R, h_I]^T. \quad (39)$$

Introducing the log-likelihood ratio

$$\begin{aligned} \Lambda(\boldsymbol{r}|h, \widehat{\boldsymbol{b}}) = \ln(p(\boldsymbol{r}|h, \widehat{\boldsymbol{b}})) &= -\ln(\pi)^L - \ln(\det[\boldsymbol{C}(h)]) \\ &\quad -(\boldsymbol{r} - \boldsymbol{\mu_r}(h))^H \boldsymbol{C}^{-1}(h)(\boldsymbol{r} - \boldsymbol{\mu_r}(h)) \end{aligned} \quad (40)$$

and denoting by $\widehat{\boldsymbol{h}}$ an unbiased estimator of $\boldsymbol{h}$, the following bound on the variance of $\widehat{\boldsymbol{h}}$ holds [21]:

$$E\{(\boldsymbol{h} - \widehat{\boldsymbol{h}})(\boldsymbol{h} - \widehat{\boldsymbol{h}})^T\} \geq \boldsymbol{J}^{-1}(h) \quad (41)$$

where $\boldsymbol{J}(h)$, the Fisher information matrix (FIM), is defined as

$$\boldsymbol{J}(h) = -E\left\{\begin{bmatrix} \frac{\partial^2\Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial h_R^2} & \frac{\partial^2\Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial h_I \partial h_R} \\ \frac{\partial^2\Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial h_R \partial h_I} & \frac{\partial^2\Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial h_I^2} \end{bmatrix}\right\}. \quad (42)$$

As shown in Appendix A, the entries of the matrix $\boldsymbol{J}(h)$ are given by

$$\begin{aligned} -E\left\{\frac{\partial^2\Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial h_R^2}\right\} &= \frac{4(L-T)A^4\sigma_{e_b}^4 {h_R}^2}{(2\mathcal{N}_0 + A^2({h_R}^2 + {h_I}^2)\sigma_{e_b}^2)^2} + \frac{A^2T}{\mathcal{N}_0} \\ &\quad + \frac{2A^2(L-T)(1-2p(e))^2}{(2\mathcal{N}_0 + A^2({h_R}^2 + {h_I}^2)\sigma_{e_b}^2)} \\ -E\left\{\frac{\partial^2\Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial h_R \partial h_I}\right\} &= \frac{4(L-T)A^4\sigma_{e_b}^4 h_R h_I}{(2\mathcal{N}_0 + A^2({h_R}^2 + {h_I}^2)\sigma_{e_b}^2)^2} \\ -E\left\{\frac{\partial^2\Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial h_I \partial h_R}\right\} &= \frac{4(L-T)A^4\sigma_{e_b}^4 h_R h_I}{(2\mathcal{N}_0 + A^2({h_R}^2 + {h_I}^2)\sigma_{e_b}^2)^2} \\ -E\left\{\frac{\partial^2\Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial h_I^2}\right\} &= \frac{4(L-T)A^4\sigma_{e_b}^4 {h_I}^2}{(2\mathcal{N}_0 + A^2({h_R}^2 + {h_I}^2)\sigma_{e_b}^2)^2} + \frac{A^2T}{\mathcal{N}_0} \\ &\quad + \frac{2A^2(L-T)(1-2p(e))^2}{(2\mathcal{N}_0 + A^2({h_R}^2 + {h_I}^2)\sigma_{e_b}^2)}. \end{aligned} \quad (43)$$

Thus, it is seen that the FIM has a nondiagonal structure, i.e., the estimators of h_R and h_I are, in general, coupled. Moreover, for $L - T = 0$, the FIM reduces to

$$\boldsymbol{J}(h) = \begin{bmatrix} \frac{A^2T}{\mathcal{N}_0} & 0 \\ 0 & \frac{A^2T}{\mathcal{N}_0} \end{bmatrix} \quad (44)$$

and provides the CRB when only T known training bits are employed for channel estimation; in this case, the CRB on the channel-estimation mse is $\text{trace}(\boldsymbol{J}^{-1}(h)) = 2\mathcal{N}_0/(A^2T)$, thus proving that the ML channel estimator achieves the CRB when no iterative strategy is adopted.

The CRB (41) provides a bound on the variance of any unbiased channel estimator. On the other hand, it is seen that taking the statistical expectation of the estimators (9) we have

$$E\{\widehat{h}_{\text{ML}}\} = \frac{hE\{\widehat{\boldsymbol{b}}^H\boldsymbol{b}\}}{L} = \frac{h[T + (L-T)(1-2p(e))]}{L} \quad (45)$$

and

$$E\{\widehat{h}_{\text{mmse}}\} = \frac{AE\{\widehat{\boldsymbol{b}}^H\boldsymbol{r}\}}{2\mathcal{N}_0 + A^2L} = \frac{A^2h[T + (L-T)(1-2p(e))]}{2\mathcal{N}_0 + A^2L}. \quad (46)$$

Otherwise stated, the proposed estimators are such that $E\{\widehat{h}\} = kh$, with

$$k = \begin{cases} k_{\text{ML}} = \frac{[T + (L-T)(1-2p(e))]}{L} \\ k_{\text{mmse}} = \frac{A^2[T + (L-T)(1-2p(e))]}{2\mathcal{N}_0 + A^2L} \end{cases} \quad (47)$$

for the ML and mmse estimators. Thus, it is seen that only the ML channel estimator with $T = L$, i.e., when no iterative strategy is adopted, is an unbiased estimator. As a consequence, it is of interest to derive the CRB for biased estimators. Considering the 2-D column vector $\widehat{\boldsymbol{h}} = [\widehat{h}_R \widehat{h}_I]^T$, with $\widehat{h}_R$ and $\widehat{h}_I$ the real part and the coefficient of the imaginary part of $\widehat{h}$, respectively, we have

$$\boldsymbol{a}(h)=E\{\widehat{\boldsymbol{h}}\}=[E\{\widehat{h}_R\},E\{\widehat{h}_I\}]^T=k[h_R,h_I]^T=k\boldsymbol{h}. \tag{48}$$

The CRB for the biased estimator $\widehat{\boldsymbol{h}}$ can thus be written as [22], [23, p. 146]

$$E\{(\widehat{\boldsymbol{h}}-E\{\widehat{\boldsymbol{h}}\})(\widehat{\boldsymbol{h}}-E\{\widehat{\boldsymbol{h}}\})^H\}\geq\nabla_{\boldsymbol{h}}\boldsymbol{a}(h)\boldsymbol{J}(h)^{-1}[\nabla_{\boldsymbol{h}}\boldsymbol{a}(h)]^T. \tag{49}$$

Since, for the case at hand, $\nabla_{\boldsymbol{h}}\boldsymbol{a}(h)=\nabla_{\boldsymbol{h}}[k\boldsymbol{h}]=k\boldsymbol{I}_2$, we have

$$E\{(\widehat{\boldsymbol{h}}-E\{\widehat{\boldsymbol{h}}\})(\widehat{\boldsymbol{h}}-E\{\widehat{\boldsymbol{h}}\})^H\}\geq k^2\boldsymbol{J}(h)^{-1}. \tag{50}$$

Finally, we consider the "Bayesian" CRB for biased estimators, which is a bound on the variance of biased estimators of quantities whose pdf is assumed to be known. Note that this is just the case of interest to the subject of this paper, since the channel coefficients are modeled as complex Gaussian random variates. The Bayesian CRB is expressed as

$$E\{(\widehat{\boldsymbol{h}}-E\{\widehat{\boldsymbol{h}}\})(\widehat{\boldsymbol{h}}-E\{\widehat{\boldsymbol{h}}\})^H\}\geq(\boldsymbol{J}_C+\boldsymbol{J}_A)^{-1} \tag{51}$$

with $\boldsymbol{J}_C=E_{h_R,h_I}[\boldsymbol{J}(h)/k^2]$ and $\boldsymbol{J}_A=-E_{h_R,h_I}[\nabla_{\boldsymbol{h}}(\nabla_{\boldsymbol{h}}(\ln p_{h_R,h_I}(h_R,h_I)))]=2\boldsymbol{I}_2$, with $p_{h_R,h_I}(h_R,h_I)$ the pdf of the vector $\boldsymbol{h}$.

B. CRB for the Multiantenna Case: $N,M>1$

Consider now the multiantenna scenario and the observable in (2). First, it is convenient to stack the columns of the $N\times L/M$-dimensional matrix $\boldsymbol{R}$ into the NL/M-dimensional column vector

$$\widetilde{\boldsymbol{r}}=A\boldsymbol{F}\widetilde{\boldsymbol{h}}+\widetilde{\boldsymbol{w}}. \tag{52}$$

In (52), $\widetilde{\boldsymbol{w}}=\left[\boldsymbol{W}^T(:,1),\ldots\boldsymbol{W}^T(:,L/M)\right]^T$, $\widetilde{\boldsymbol{h}}=[\boldsymbol{H}^T(:,1),\ldots\boldsymbol{H}^T(:,M)]^T$ is an NM-dimensional column vector and, finally, $\boldsymbol{F}=\boldsymbol{B}^T\otimes\boldsymbol{I}_N$ ($\otimes$ denotes Kronecker product) is a $NL/M\times NM$-dimensional matrix containing the information symbols. We start by deriving the CRB for unbiased estimators. Let $\widehat{\boldsymbol{F}}=\widehat{\boldsymbol{B}}^T\otimes\boldsymbol{I}_N$, with $\widehat{\boldsymbol{B}}$ an $M\times L/M$-dimensional matrix containing, in the first T/M columns, the T known training bits and, in the remaining columns, the $L-T$ detected symbols that are fed back to the channel estimator. The observable (52) can, thus, be expressed as

$$\widetilde{\boldsymbol{r}}=A\widehat{\boldsymbol{F}}\widetilde{\boldsymbol{h}}+\underbrace{A(\boldsymbol{F}-\widehat{\boldsymbol{F}})\widetilde{\boldsymbol{h}}+\widetilde{\boldsymbol{w}}}_{\widetilde{\boldsymbol{n}}}. \tag{53}$$

The entries of the matrix $\boldsymbol{B}$ are such that

$$\boldsymbol{B}(i,j)=\begin{cases}\widehat{\boldsymbol{B}}(i,j) & \text{w.p. } 1 \quad \forall i,\forall j\leq\frac{T}{M}\\ \widehat{\boldsymbol{B}}(i,j) & \text{w.p. } 1-p(e)\ \forall i,\forall j>\frac{T}{M}\\ -\widehat{\boldsymbol{B}}(i,j) & \text{w.p. } p(e) \quad \forall i,\forall j>\frac{T}{M}.\end{cases} \tag{54}$$

Again, we assume that $\widetilde{\boldsymbol{n}}$ is a complex Gaussian-random vector. Accordingly, conditioned on $\widetilde{\boldsymbol{h}}$ and $\widehat{\boldsymbol{F}}$, the vector $\widetilde{\boldsymbol{r}}$ is a complex Gaussian vector and its conditional pdf is

$$p(\widetilde{\boldsymbol{r}}|\widetilde{\boldsymbol{h}},\widehat{\boldsymbol{B}})=\frac{1}{\det(\pi\widetilde{\boldsymbol{C}}(\widetilde{\boldsymbol{h}}))}e^{-(\widetilde{\boldsymbol{r}}-\boldsymbol{\mu}_{\widetilde{\boldsymbol{r}}}(\widetilde{\boldsymbol{h}}))^H\widetilde{\boldsymbol{C}}^{-1}(\widetilde{\boldsymbol{h}})(\widetilde{\boldsymbol{r}}-\boldsymbol{\mu}_{\widetilde{\boldsymbol{r}}}(\widetilde{\boldsymbol{h}}))}. \tag{55}$$

Upon defining

$$\overline{\boldsymbol{B}}=\left[\boldsymbol{O}(:,1),\ldots,\boldsymbol{O}\left(:,\frac{T}{M}\right),\widehat{\boldsymbol{B}}\left(:,\frac{T}{M}+1\right),\ldots,\widehat{\boldsymbol{B}}\left(:,\frac{L}{M}\right)\right] \tag{56}$$

and

$$\widetilde{\boldsymbol{B}}=\left[\widehat{\boldsymbol{B}}(:,1),\ldots,\widehat{\boldsymbol{B}}\left(:,\frac{T}{M}\right),\widehat{\boldsymbol{B}}\left(:,\frac{T}{M}+1\right)(1-2p(e))\right.\\ \left.\ldots,\widehat{\boldsymbol{B}}\left(:,\frac{L}{M}\right)(1-2p(e))\right] \tag{57}$$

in (55) we have

$$\begin{aligned}\boldsymbol{\mu}_{\widetilde{\boldsymbol{r}}}(\widetilde{\boldsymbol{h}})&=E\{\widetilde{\boldsymbol{r}}\}=A\widehat{\boldsymbol{F}}\widetilde{\boldsymbol{h}}-2Ap(e)(\overline{\boldsymbol{B}}^T\otimes\boldsymbol{I}_N)\widetilde{\boldsymbol{h}}\\&=A(\widetilde{\boldsymbol{B}}^T\otimes\boldsymbol{I}_N)\widetilde{\boldsymbol{h}}\end{aligned} \tag{58}$$

and

$$\begin{aligned}&\widetilde{\boldsymbol{C}}(\widetilde{\boldsymbol{h}})\\&=E\{(\widetilde{\boldsymbol{n}}-E\{\widetilde{\boldsymbol{n}}\})(\widetilde{\boldsymbol{n}}-E\{\widetilde{\boldsymbol{n}}\})^H\}=\\&=E\{\widetilde{\boldsymbol{w}}\widetilde{\boldsymbol{w}}^H\}-4A^2p(e)^2(\overline{\boldsymbol{B}}^T\otimes\boldsymbol{I}_N)\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}^H(\overline{\boldsymbol{B}}^T\otimes\boldsymbol{I}_N)^H\\&\quad+E\{A^2(\boldsymbol{F}-\widehat{\boldsymbol{F}})\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}^H(\boldsymbol{F}-\widehat{\boldsymbol{F}})^H\}\\&=2\mathcal{N}_0\boldsymbol{I}_{NL/M}-4A^2p(e)^2(\overline{\boldsymbol{B}}^T\otimes\boldsymbol{I}_N)\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}^H(\overline{\boldsymbol{B}}^T\otimes\boldsymbol{I}_N)^H\\&\quad+A^2(\widehat{\boldsymbol{B}}^T\otimes\boldsymbol{I}_N)\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}^H(\widehat{\boldsymbol{B}}^T\otimes\boldsymbol{I}_N)^H+\\&\quad-A^2(\widetilde{\boldsymbol{B}}^T\otimes\boldsymbol{I}_N)\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}^H(\widehat{\boldsymbol{B}}^T\otimes\boldsymbol{I}_N)^H\\&\quad-A^2(\widehat{\boldsymbol{B}}^T\otimes\boldsymbol{I}_N)\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}^H(\widetilde{\boldsymbol{B}}^T\otimes\boldsymbol{I}_N)^H+E\{A^2\boldsymbol{F}\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}^H\boldsymbol{F}^H\}.\end{aligned} \tag{59}$$

In order to come up with a simpler formula, we approximate the expressions[9] for the summands in (59). First of all, note that the matrix $E\{\boldsymbol{F}\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}^H\boldsymbol{F}^H\}$ is well approximated by a block-diagonal matrix, say $\boldsymbol{D}$, that is written as

$$E\{\boldsymbol{F}\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}^H\boldsymbol{F}^H\}\approx\boldsymbol{D}=\text{Diag}\left(\underbrace{\boldsymbol{D}_1\ldots\boldsymbol{D}_1}_{L/M\text{ times}}\right) \tag{60}$$

with $\boldsymbol{D}_1$ an $N\times N$-dimensional matrix whose (i,j)th element is written as

$$\boldsymbol{D}_1(i,j)=\sum_{l=0}^{M-1}(\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}^H)_{i+lN,j+lN}\quad\forall\ i,j=1,\ldots,N. \tag{61}$$

[9] We do not dwell on the proofs of the subsequent approximations for the sake of brevity.

Moreover, we also have

$$(\overline{\boldsymbol{B}}^T \otimes \boldsymbol{I}_N)\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}^H(\overline{\boldsymbol{B}}^T \otimes \boldsymbol{I}_N)^H \approx \begin{bmatrix} \boldsymbol{O}_{NT/M,NT/M} & \boldsymbol{O}_{NT/M,N(L-T)/M} \\ \boldsymbol{O}_{N(L-T)/M,NT/M} & \boldsymbol{P} \end{bmatrix} \quad (62)$$

with

$$\boldsymbol{P} = \mathrm{Diag}\left(\underbrace{\boldsymbol{D}_1 \ \ldots \ \boldsymbol{D}_1}_{(L-T)/M \text{ times}}\right)$$

and $(\widehat{\boldsymbol{B}}^T \otimes \boldsymbol{I}_N)\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}^H(\widehat{\boldsymbol{B}}^T \otimes \boldsymbol{I}_N)^H \approx \boldsymbol{D}$ and

$$(\widetilde{\boldsymbol{B}}^T \otimes \boldsymbol{I}_N)\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}^H(\widehat{\boldsymbol{B}}^T \otimes \boldsymbol{I}_N)^H = \mathrm{Diag}\left(\underbrace{\boldsymbol{D}_1 \ldots \boldsymbol{D}_1}_{T/M \text{ times}} \ \underbrace{(1-2p(e))\boldsymbol{D}_1 \ldots (1-2p(e))\boldsymbol{D}_1}_{(L-T)/M \text{ times}}\right).$$

Based on the above approximations, it is seen that (59) can be given the more compact representation

$$\widetilde{\boldsymbol{C}}(\widetilde{\boldsymbol{h}}) = \mathrm{Diag}\left(\underbrace{2\mathcal{N}_0\boldsymbol{I}_N \ldots 2\mathcal{N}_0\boldsymbol{I}_N}_{T/M \text{ times}} \ \underbrace{2\mathcal{N}_0\boldsymbol{I}_N + \overline{\boldsymbol{R}_2} \ldots 2\mathcal{N}_0\boldsymbol{I}_N + \overline{\boldsymbol{R}_2}}_{(L-T)/M \text{ times}}\right) \quad (63)$$

with

$$\overline{\boldsymbol{R}_2}(i,j) = 4A^2 p(e)(1-p(e)) \cdot \sum_{l=0}^{M-1}\left(\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}^H\right)_{i+lN,j+lN} \text{ for } i,j=1,\ldots,N. \quad (64)$$

Now, armed with (55), (58), and (63), deriving the unbiased CRB is straightforward. Once again, since the parameter to be estimated is complex, we consider the $2NM$-dimensional real vector

$$\widetilde{\boldsymbol{h}}_r = [\,\mathcal{R}(\widetilde{\boldsymbol{h}})^T \quad \mathcal{I}(\widetilde{\boldsymbol{h}})^T\,]^T \quad (65)$$

with $\mathcal{I}(\cdot)$ denoting the coefficient of the imaginary part. Introducing the log-likelihood ratio

$$\Lambda(\widetilde{\boldsymbol{r}}|\widetilde{\boldsymbol{h}},\widehat{\boldsymbol{B}}) = \ln\left[p(\widetilde{\boldsymbol{r}}|\widetilde{\boldsymbol{h}},\widehat{\boldsymbol{B}})\right] = -\ln(\det(\pi\widetilde{\boldsymbol{C}}(\widetilde{\boldsymbol{h}})) - \left(\widetilde{\boldsymbol{r}} - \boldsymbol{\mu}_{\widetilde{\boldsymbol{r}}}(\widetilde{\boldsymbol{h}})\right)^H \widetilde{\boldsymbol{C}}^{-1}\left(\widetilde{\boldsymbol{h}}\right)\left(\widetilde{\boldsymbol{r}} - \boldsymbol{\mu}_{\widetilde{\boldsymbol{r}}}(\widetilde{\boldsymbol{h}})\right) \quad (66)$$

and denoting by $\widehat{\widetilde{\boldsymbol{h}}}_r$ an unbiased estimator of $\widetilde{\boldsymbol{h}}_r$, the CRB is written as

$$E\left\{\left(\widehat{\widetilde{\boldsymbol{h}}}_r - \widetilde{\boldsymbol{h}}_r\right)\left(\widehat{\widetilde{\boldsymbol{h}}}_r - \widetilde{\boldsymbol{h}}_r\right)^H\right\} \geq \boldsymbol{J}^{-1}(\widetilde{\boldsymbol{h}}_r) \quad (67)$$

where $\boldsymbol{J}(\widetilde{\boldsymbol{h}}_r)$ is the FIM whose (i,j)th element is now defined as

$$\left[\boldsymbol{J}(\widetilde{\boldsymbol{h}}_r)\right]_{i,j} = -E\left\{\frac{\partial^2\Lambda(\widetilde{\boldsymbol{r}}|\widetilde{\boldsymbol{h}},\widehat{\boldsymbol{B}})}{\partial\widetilde{\boldsymbol{h}}_r(j)\partial\widetilde{\boldsymbol{h}}_r(i)}\right\}. \quad (68)$$

Further details on how to compute the entries of the FIM are reported in Appendix B.

Consider now the biased CRB. First of all, stacking in NM-dimensional column vectors, say $\widehat{\widetilde{\boldsymbol{h}}}_{\mathrm{ML}}$ and $\widehat{\widetilde{\boldsymbol{h}}}_{\mathrm{mmse}}$ the estimators (4) and (5), respectively, and taking the statistical expectation yields

$$E\left\{\widehat{\widetilde{\boldsymbol{h}}}_{\mathrm{ML}}\right\} \approx \left[1 - \frac{(L-T)2p(e)}{L}\right]\widetilde{\boldsymbol{h}}$$

and

$$E\left\{\widehat{\widetilde{\boldsymbol{h}}}_{\mathrm{mmse}}\right\} \approx \left[\frac{A^2(L-(L-T)2p(e))}{LA^2 + 2\mathcal{N}_0 M}\right]\widetilde{\boldsymbol{h}}. \quad (69)$$

Accordingly, if we assume that $E\{\widehat{\widetilde{\boldsymbol{h}}}\} = k\widetilde{\boldsymbol{h}}$, then following the same steps as in the single-antenna case leads to the bound

$$E\left\{\left(\widehat{\widetilde{\boldsymbol{h}}}_r - E\left\{\widehat{\widetilde{\boldsymbol{h}}}_r\right\}\right)\left(\widehat{\widetilde{\boldsymbol{h}}}_r - E\left\{\widehat{\widetilde{\boldsymbol{h}}}_r\right\}\right)^H\right\} \geq k^2(\boldsymbol{J}(\widetilde{\boldsymbol{h}}_r))^{-1}. \quad (70)$$

Finally, consider the Bayesian CRB for biased estimators. We now have

$$E\left\{\left(\widehat{\widetilde{\boldsymbol{h}}}_r - E\left\{\widehat{\widetilde{\boldsymbol{h}}}_r\right\}\right)\left(\widehat{\widetilde{\boldsymbol{h}}}_r - E\left\{\widehat{\widetilde{\boldsymbol{h}}}_r\right\}\right)^H\right\} \geq (\boldsymbol{J}_{C_1} + \boldsymbol{J}_{A_1})^{-1} \quad (71)$$

with $\boldsymbol{J}_{C_1} = E_{\widetilde{\boldsymbol{h}}_r}[\boldsymbol{J}(\widetilde{\boldsymbol{h}}_r)/k^2]$ and $\boldsymbol{J}_{A_1} = -E_{\widetilde{\boldsymbol{h}}_r}[\nabla_{\widetilde{\boldsymbol{h}}_r}(\nabla_{\widetilde{\boldsymbol{h}}_r}(\ln(p_{\widetilde{\boldsymbol{h}}_r}(\widetilde{\boldsymbol{h}}_r))))]$. In our context, since $\widetilde{\boldsymbol{h}}_r$ is a vector with independent and identically distributed (i.i.d.) real Gaussian variates, we have $\boldsymbol{J}_{A_1} = 2\boldsymbol{I}_{2NM}$.

As a special case of the previous derivations, we can obtain an expression for the CRB when no iterative strategy is employed. Indeed, letting $L = T$ and $p(e) = 0$, the FIM is now written as

$$\boldsymbol{J} = \begin{bmatrix} \frac{A^2(\boldsymbol{B}^T\otimes\boldsymbol{I}_N)^H(\boldsymbol{B}^T\otimes\boldsymbol{I}_N)}{\mathcal{N}_0} & \boldsymbol{O}_{NM\times NM} \\ \boldsymbol{O}_{NM\times NM} & \frac{A^2(\boldsymbol{B}^T\otimes\boldsymbol{I}_N)^H(\boldsymbol{B}^T\otimes\boldsymbol{I}_N)}{\mathcal{N}_0} \end{bmatrix}. \quad (72)$$

Assuming now that $E\{\widehat{\widetilde{\boldsymbol{h}}}\} = k\widetilde{\boldsymbol{h}}$, the biased CRB is now expressed as

$$\begin{aligned} &E\left\{\left\|\widehat{\widetilde{\boldsymbol{h}}}_r - E\left\{\widehat{\widetilde{\boldsymbol{h}}}_r\right\}\right\|^2\right\} \\ &\geq \mathrm{trace}\{k^2\boldsymbol{J}^{-1}\} \\ &= \frac{2\mathcal{N}_0 k^2}{A^2}\mathrm{tr}\{[(\boldsymbol{B}^T\otimes\boldsymbol{I}_N)^H(\boldsymbol{B}^T\otimes\boldsymbol{I}_N)]^{-1}\} \\ &\approx \frac{2\mathcal{N}_0 NM^2k^2}{A^2T} \end{aligned} \quad (73)$$

in keeping with (18) for the ML channel estimator.

In Fig. 7, we report the variance of the mmse channel estimator and the biased Bayesian CRB (71) versus the error probability of the detected bits. We consider both a multiantenna system with $N = M = 2$ and a single-antenna system. The SNR is 5 dB and the packet length is $L = 200$, while the training-sequence length T is 1 and 4 for the single-antenna and multiantenna cases, respectively. We also report the CRB corresponding to the cases that either T or L training bits are used to estimate the channel. Interestingly, the CRB is an

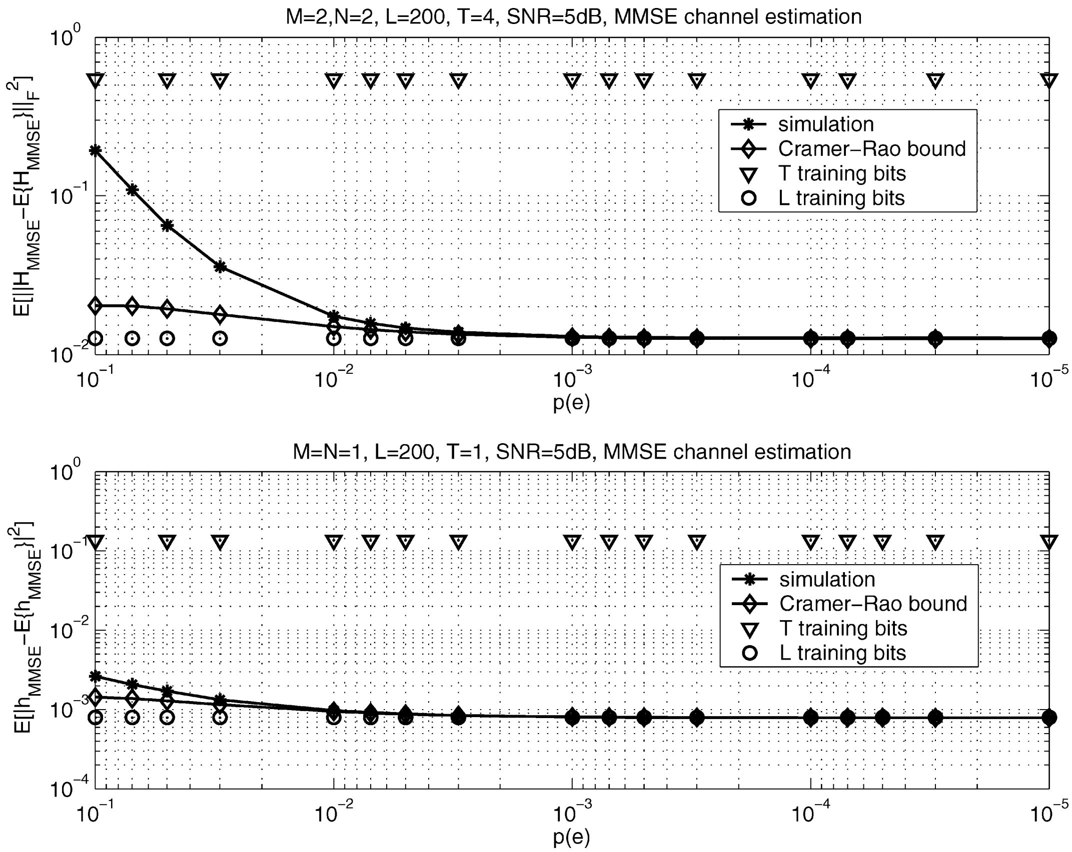

Fig. 7. Variance of mmse channel-estimation mse and its Bayesian CRB versus p(e) for the multiple- and single-antenna systems.

increasing function of $p(e)$, i.e., the lower the $p(e)$, the lower the resulting CRB. This behavior confirms, at a theoretical level, the intuition that adopting the iterative strategy yields a performance improvement. Moreover, since the CRB curve is always below the CRB corresponding to the case that only T training bits are used to estimate the channel, we can claim that the iterative strategy may also yield some advantages when the bits have been detected with a not-so-small error probability. Finally, note that the mmse estimation procedure achieves the CRB in the region of low $p(e)$.

VI. Error-Probability Analysis for the Single-Antenna System

Now consider the system error-probability analysis for the single-antenna scenario.[10] The information bits can be detected according to the rule

$$\widehat{b}(p) = \mathrm{sgn}\left(\mathcal{R}\left\{\widehat{h}^* r(p)\right\}\right) = \mathrm{sgn}\left(\mathcal{R}\left\{A\widehat{h}^* h b(p) + \widehat{h}^* w(p)\right\}\right), \ \forall p = T, \ldots, L-1 \tag{74}$$

with $r(p)$, $b(p)$, and $w(p)$ the pth entry of the vectors $\boldsymbol{r}$, $\boldsymbol{b}$, and $\boldsymbol{w}$, respectively. Let us first assume that $Q = T$. Denoting by ϕ and $\widehat{\phi}$ the phases of the complex factors h and $\widehat{h}$, respectively, the bit-error probability conditioned on h and $\widehat{h}$ is written as

$$P(e|h,\widehat{h}) = \frac{1}{2}\mathrm{erfc}\left(\frac{\mathcal{R}\left(A\widehat{h}^* h\right)}{\sqrt{2\mathcal{N}_0}|\widehat{h}|}\right) = \frac{1}{2}\mathrm{erfc}\left(\frac{A|h|\cos(\phi - \widehat{\phi})}{\sqrt{2\mathcal{N}_0}}\right) \tag{75}$$

with $\mathrm{erfc}(\cdot)$ the complementary error function. To obtain the unconditional error probability, note that, based on (9), we have

$$\widehat{h} = \beta h + z \tag{76}$$

where either $\beta = 1$, for the ML estimate, or $\beta = A^2T/(2\mathcal{N}_0 + A^2T)$, for the mmse estimate, and z is a zero-mean complex Gaussian-random variate with variance equal to either $2\mathcal{N}_0/A^2T$, for the ML estimate, or $2\mathcal{N}_0 A^2 T/(2\mathcal{N}_0 + A^2T)^2$, for the mmse estimate. Accordingly, denoting by

$$f_{\widehat{h},h}(\widehat{h},h) = f_{\widehat{h}|h}(\widehat{h}|h) f_h(h) = \frac{1}{2\pi^2\sigma^2} e^{-(|\widehat{h}-\beta h|^2)/2\sigma^2} e^{-|h|^2}$$

the joint pdf of $(\widehat{h}, h)$, the unconditional error probability can be shown to be written as

$$P(e) = \int P(e|h,\widehat{h}) f_{\widehat{h}|h}(\widehat{h}|h) f_h(h)\, dh\, d\widehat{h}$$

[10]Due to the lack of a manageable closed-form formula, only computer-simulation results will be reported for the multiantenna case.

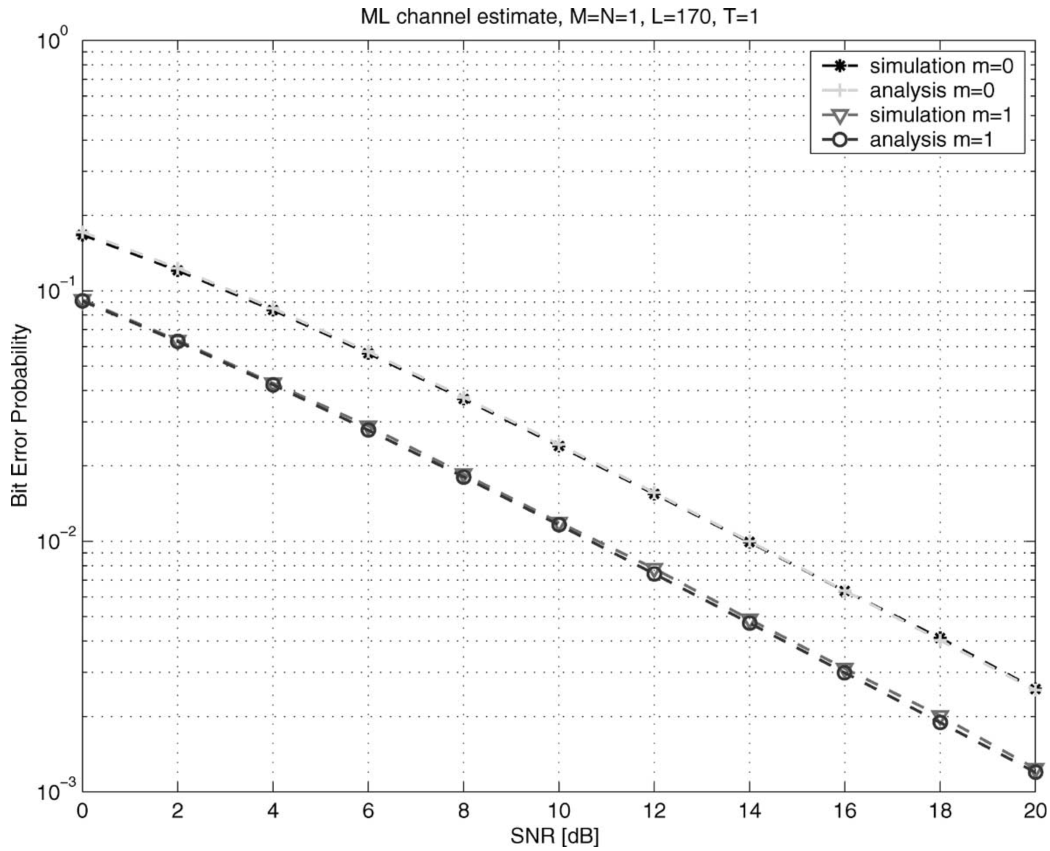

Fig. 8. Bit-error probability versus the SNR compared with the numerical results in the single-antenna case.

$$= \int_0^\infty d|\widehat{h}| \int_0^\infty d|h| \int_0^{2\pi} d\widehat{\phi} \int_0^{2\pi} d\phi \times \left[\frac{1}{4\pi^2\sigma^2} \mathrm{erfc}\left(\frac{A|h|\cos(\phi - \widehat{\phi})}{\sqrt{2\mathcal{N}_0}} \right) |h| e^{-|h|^2} |\widehat{h}| \exp\left(-\frac{|\widehat{h}|^2 + \beta^2|h|^2 - 2\beta|\widehat{h}||h|\cos(\phi-\widehat{\phi})}{2\sigma^2} \right) \right]. \quad (77)$$

Let us finally consider the case that $Q = L$ and let us denote the bit-error probability after m iterations by p_m. Given h and $\widehat{h}$, the conditional error probability at the $(m+1)$th iteration is still expressed as in (75); however, the pdf of $\widehat{h}$ now depends on the detected information-bits realizations. In particular, it can be shown that

$$f_{\widehat{h}|h,\boldsymbol{b},\boldsymbol{b}_e}(\widehat{h}|h,\boldsymbol{b},\boldsymbol{b}_e) = \frac{1}{2\pi\sigma^2} e^{-(|\widehat{h} - \beta h (L - \boldsymbol{b}_e^H \boldsymbol{b})|^2 / 2\sigma^2)} \quad (78)$$

with either $2\sigma^2 = 2\mathcal{N}_0/A^2L$ for the ML estimate or $2\sigma^2 = 2\mathcal{N}_0 A^2 L/(2\mathcal{N}_0 + A^2L)^2$ for the mmse estimate and $\beta = 1/L$ or $\beta = A^2/(2\mathcal{N}_0 + A^2L)$ for the ML and mmse estimates, respectively. Interestingly, the pdf in (78) is a function of the inner product $\boldsymbol{b}_e^H\boldsymbol{b}$, whose pdf is

$$f_{\boldsymbol{b}_e^H\boldsymbol{b}}(x) = \sum_{\ell=0}^{L-T} \binom{L-T}{\ell} p_m^\ell (1-p_m)^{L-T-\ell} \delta(x + 2\ell). \quad (79)$$

Neglecting the statistical dependence of $\boldsymbol{b}_e^H\boldsymbol{b}$ on the channel h, i.e., assuming that $f_{\boldsymbol{b}_e^H\boldsymbol{b}|h}(x|h) \approx f_{\boldsymbol{b}_e^H\boldsymbol{b}}(x)$, we have

$$f_{\widehat{h}|h}(\widehat{h}|h) = \int f_{\widehat{h}|h,\boldsymbol{b}_e^H\boldsymbol{b}}(\widehat{h}|h,x) f_{\boldsymbol{b}_e^H\boldsymbol{b}}(x)dx. \quad (80)$$

Finally, given (80), it is easy to obtain (81), given at the bottom of the next page. In Fig. 8, we report the system-error probabilities $P(e)$ and p_1 versus the SNR in both the cases that either formulas (77) and (81) are used or that Montecarlo simulations are performed. The integrations in (77) and (81) have been carried out numerically. A system with $T = 1$ training bit and $L = 170$ has been considered and the ML channel estimator has been adopted. The theoretical analysis is again in excellent agreement with the computer simulations. Moreover, adopting the iterative strategy permits us to improve the system-error probability; in particular, a performance gain of more than 3 dB is achieved after one iteration. In Fig. 9, we report the error probability at the $(m+1)$th iteration p_{m+1} in (81) versus p_m for two different values of SNR. We consider a system with $T = 1$ training bit and $L = 170$. Interestingly, even if the starting error probability p_m is close to 0.5, the error probability at the following iteration (i.e., p_{m+1}) is largely decreased; conversely, it is seen that $p_{m+1} > p_m$ for only $p_m > 0.5$. Thus, the proposed strategy is also effective in improving the system performance when the information bits are detected with a large (i.e., close to 0.5) error probability.

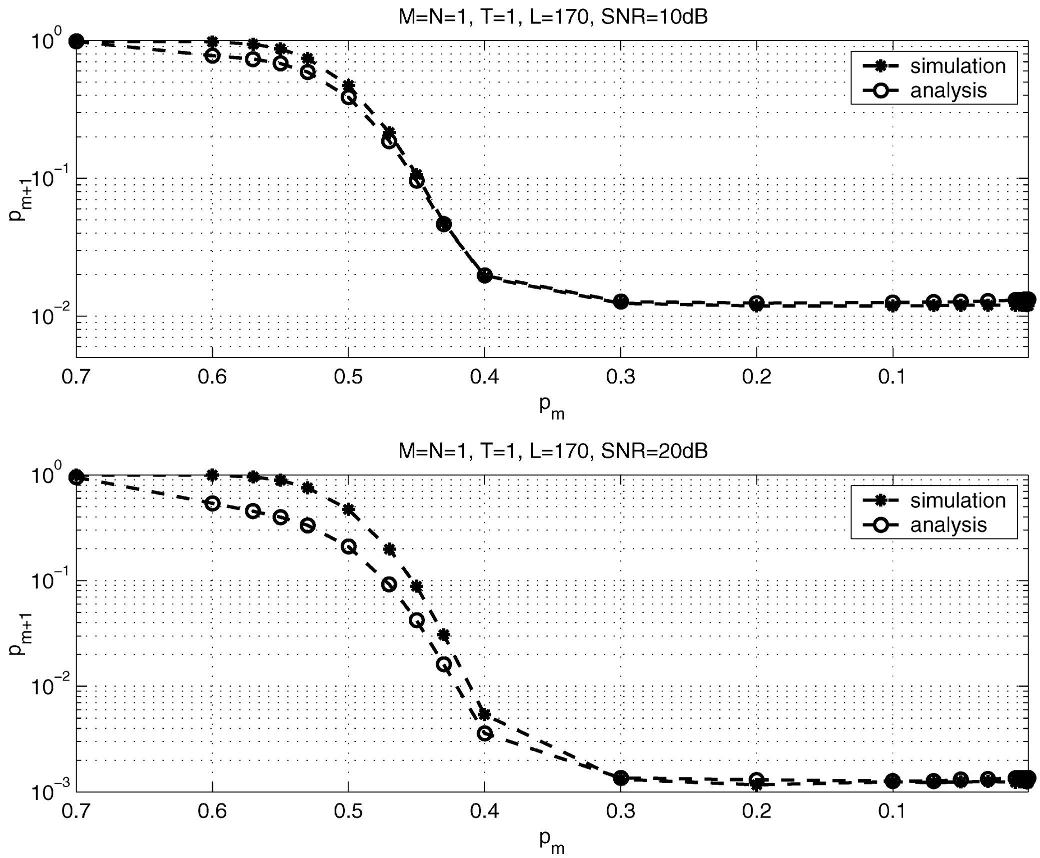

Fig. 9. p_{m+1} versus p_m in the single-antenna case and for two values of the SNR.

A. Multiantenna Case: Simulation Example

Consider a system of $M = 4$ transmit and $N = 6$ receive antennas; let $T = M^2$ and $L = 200$. Fig. 10 demonstrates the beneficial effects of the proposed strategy. Indeed, here we report two subplots showing, for the ZF and mmse detectors, the performance improvements that the iterative strategy can grant. The results have been averaged over 10^4 independent data-frame realizations. The parameter m denotes the number of times that the detected bits are fed back to the channel estimator (i.e., $m = 0$ means that no iterative strategy has been pursued). In these plots, we also report the curves corresponding to both the ideal situations that the channel is known and that has been estimated using L error-free detected bits: both these curves are a lower bound to the error probability that can be achieved by the iterative strategy. Overall, the results demonstrate two remarkable facts: first, even a small number of iterations (i.e., $m = 2$) enables a gain of about 5 dB with respect to the case that $m = 0$ and second, even when the length of the training sequence is kept at its minimum (i.e., $T = M^2$), the error probability after three iterations is less than 2 dB, far from its lower bound at an error probability of 10^{-3}.

VII. Convergence Analysis and Training-Length Optimization

A. Convergence Analysis of the Iterative Algorithm

Consider now the issue of investigating the existence and stability of possible fixed points of the algorithm. To this end, first note that upon defining the bit estimation error

$$e_b(p) = b(p) - \widehat{b}(p) \tag{82}$$

$$p_{m+1} = \sum_{k=0}^{L-T} \binom{L-T}{k} (1-p_m)^{L-T-k} p_m^k \int_0^\infty d|\widehat{h}| \int_0^\infty d|h| \int_0^{2\pi} d\widehat{\phi} \int_0^{2\pi} d\phi \cdot \left[\frac{1}{4\pi^2\sigma^2} \operatorname{erfc}\left(\frac{A|h|\cos(\phi - \widehat{\phi})}{\sqrt{2N_0}} \right) \times |h| e^{-|h|^2} |\widehat{h}| \exp - \frac{|\widehat{h}|^2 + \beta^2 |h|^2 (L-2k)^2 - 2\beta |\widehat{h}||h| \cos(\phi - \widehat{\phi})(L-2k)}{2\sigma^2} \right]. \tag{81}$$

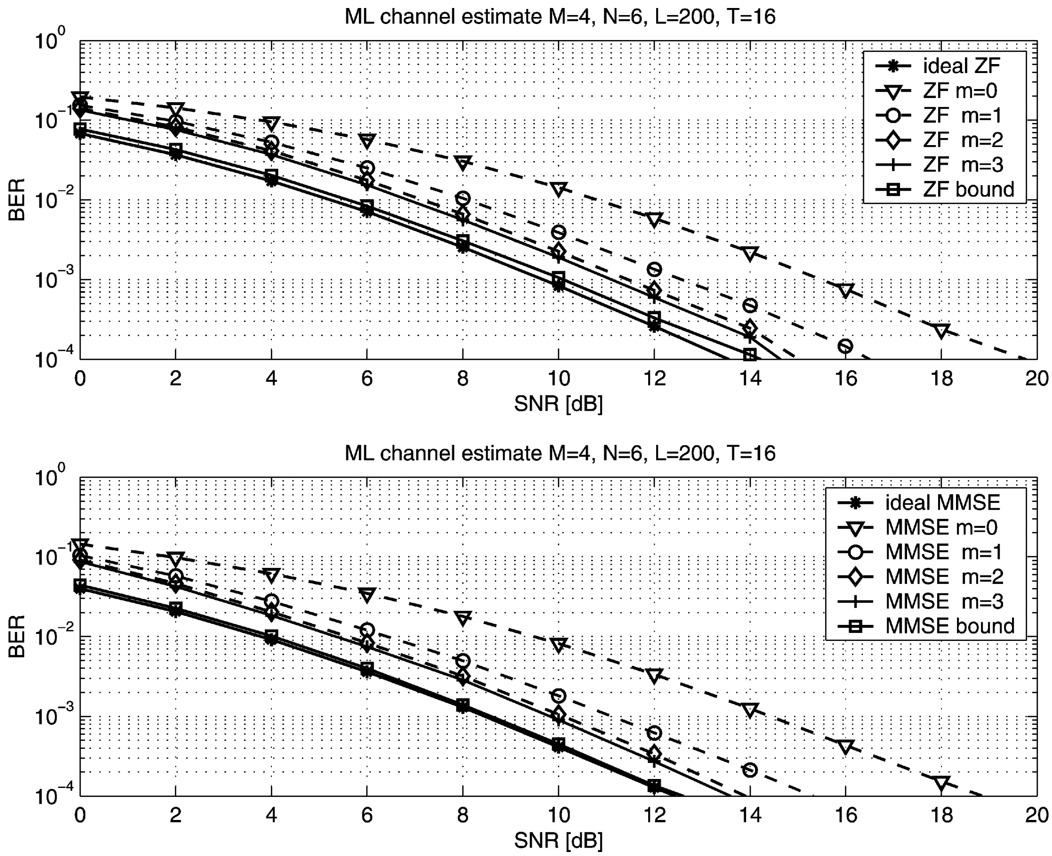

Fig. 10. System BER for the ZF and mmse receivers versus the SNR, in the case of ML channel estimation and for several values of m.

it is easily shown that since $\widehat{b}(p)$ is the opposite of $b(p)$ w.p. $p(e)$ and equals $b(p)$ w.p. $1-p(e)$, it follows that the mean of $e_b(p)$ is $E\{e_b(p)\}=2b(p)p(e)$ and its variance is $\sigma_{e_b}^2 = E\{[e_b - E\{e_b(p)\}]^2\} = 4p(e)(1-p(e))$. Inverting this relation with $p(e)\in[0,0.5]$ yields

$$p(e)=\frac{(1-\sqrt{1-\sigma_{e_b}^2})}{2}. \quad (83)$$

Thus, there is a one-to-one correspondence between the error probability and the variance of the bit-estimation error. Accordingly, (13) and (25) provide a relationship between the ML[11] channel-estimation mse, which we denote here by σ_c^2, and the variance of the bit-estimation error. In particular, given (83), these relations can be rewritten as

$$\sigma_c^2=\frac{2\mathcal{N}_0}{LA^2}+\frac{2\left(1-\sqrt{1-\sigma_{e_b}^2}\right)(L-T)}{L^2} \cdot\left[1+(L-T-1)\frac{(1-\sqrt{1-\sigma_{e_b}^2})}{2}\right] \quad (84)$$

and

$$\sigma_c^2\approx 2\left(1-\sqrt{1-\sigma_{e_b}^2}\right)M\frac{(L-T)}{L^2} +\left(1-\sqrt{1-\sigma_{e_b}^2}\right)^2\frac{(L-T-M)(L-T)}{L^2}+\frac{2\mathcal{N}_0M}{LA^2} \quad (85)$$

[11]We focus here on the ML channel estimate, but these considerations also apply to (14) and (28), which refer to an mmse channel estimate.

for the single-antenna and multiantenna situations, respectively. Equations (84) and (85) can be written compactly as $\sigma_c^2 = f(\sigma_{e_b}^2)$ and the channel estimator can be viewed as a black box accepting at its input a given bit-estimation error variance and producing at its output a channel estimate with mse σ_c^2.

In regard to the data detector, instead note that the bit-error probability and, consequently, the variance of the bit-estimation error, generally depend not only on the second-order moment σ_c^2, but on the pdf of the channel-estimation error too. Nonetheless, by virtue of computer simulations, it can be obtained an empirical plot relating the mse of the channel estimate at the input of the data detector with the bit-estimation error variance at the output of the data detector, i.e., $\sigma_{e_b}^2 = g(\sigma_c^2)$.

Given the above formulation, denoting by $\sigma_c^2(m)$ and by $\sigma_{e_b}^2(m)$ the channel-estimation mse and the bit-estimation error variance at the mth iteration, a single iteration of the algorithm is described as

$$\sigma_c^2(m+1)=f(g(\sigma_c^2(m))) \text{ and } \sigma_{e_b}^2(m+1)=g(f(\sigma_{e_b}^2(m))). \quad (86)$$

Fixed points of $f(g(\cdot))$ and of $g(f(\cdot))$ and their stability represent the asymptotic convergence points of the iterative algorithm. Unfortunately, lacking an analytical expression for the function $g(\cdot)$, no theoretical considerations can be done on the fixed points of the iterative processing. However, in keeping with the analysis on turbo-equalization algorithms in [24] and [25], the existence of a stable fixed point can be shown graphically. In Fig. 11, the functions $f(\cdot)$ and $g(\cdot)$ are reported for SNR $=5$ dB; in particular, the upper plot refers to a single-

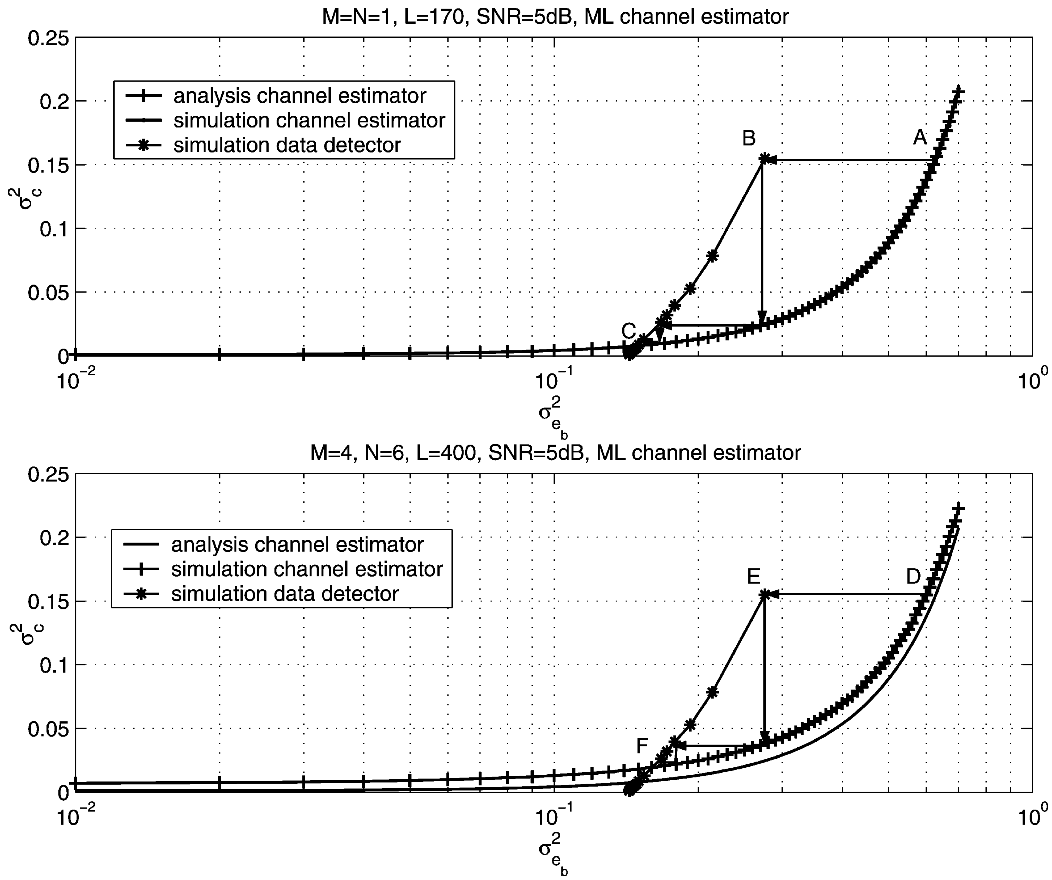

Fig. 11. ML channel-estimation mse versus the variance of the bit-estimation error for the channel estimator and data detector in the two cases of single-antenna and multiple-antenna systems.

antenna system, while the lower plot refers to a multiantenna system with $M = 4$ and $N = 6$. The function $g(\cdot)$, which refers to the data detector and relates the input channel-estimation mse with the output variance of the bit-estimation error, has been obtained through computer simulations; as instead regards the function $f(\cdot)$, we have reported both the analytical formulas (84) and (85) and the curves obtained through simulations. The plots can be interpreted in the following way. At the 0th iteration, the channel estimator achieves a certain channel-estimation mse (ordinate of the points A and D) based on the training bits only. The data detector, in turn, based on the channel estimate, achieves a certain error variance in the bit-estimation error (abscissa of the points B and E) and the detected bits are then fed back to the channel estimator, which thus achieves a new channel-estimation mse. Iterating this procedure, it is seen that after few recursions the crossing point between the functions $f(\cdot)$ and $g(\cdot)$ is reached, representing a fixed stable point for the iterative algorithm.

B. Training-Length Optimization: Estimation Accuracy versus Throughput

In previous derivations, it has been shown that the channel-estimation error is a decreasing function of the training length T. On the other hand, for a fixed L, increasing T leads to a reduction in the net data throughput $(L-T)/L$, which is the fraction of information symbols in the data frame. Accordingly, the training length T is chosen to be a compromise between the conflicting requirements of achieving satisfactory performance and of not nulling the system data throughput. Thus, it is reasonable to introduce the objective function

$$\eta(T) = \frac{E\left\{\left\|\boldsymbol{H} - \widehat{\boldsymbol{H}}\right\|_F^2\right\}}{\frac{(L-T)}{L}} \tag{87}$$

which is the ratio of the channel-estimation mse to the data throughput. Using (18) and (25), $\eta(T)$ is written, for the ML channel estimator, as

$$\eta(T) = \frac{2\mathcal{N}_0 N M^2 L}{A^2 T(L-T)} \tag{88}$$

and as

$$\eta(T) \approx \frac{4p(e)M^2 N}{L} + \frac{4p(e)^2 MN(L-T-M)}{L} + \frac{2\mathcal{N}_0 N M^2}{A^2(L-T)} \tag{89}$$

for the noniterative and iterative strategies, respectively. Now, (88) and (89) can be minimized with respect to T to determine the optimal value of the training length. It is easily shown that (88) is minimized for $T = L/2$. Unfortunately, the minimization of (89) cannot be carried out through a derivative with respect to T, since the error probability is itself a function of T. Accordingly, the minimum of $\eta(T)$ can be derived through a numerical procedure. In Fig. 12, we report $\eta(T)$ versus T for several values of the SNR. A multiantenna system with $M = 4$

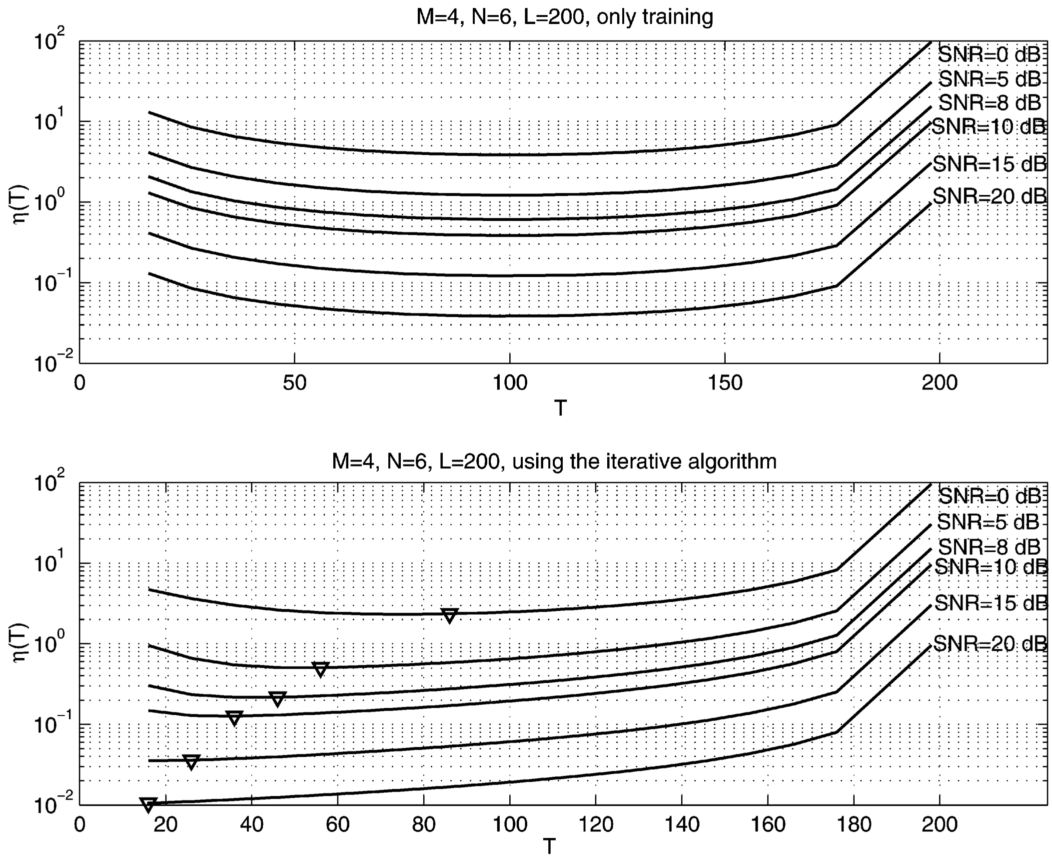

Fig. 12. $\eta(T)$ versus T for several values of the SNR in the multiple-antenna case.

and $N=6$ is considered, while the packet length is $L=200$. The upper plot refers to the noniterative system, while the lower plot refers to the iterative procedure; here, we have marked the minimum of each curve with a triangle. Comparing the upper and lower plots, several interesting remarks can be made. First, for a fixed SNR, the iterative procedure permits attaining lower values of $\eta(T)$. Moreover, the lower plot reveals that, for increasing SNR, the optimal training length is reduced; in particular, for SNR $=20$ dB, the optimal training length is equal to its minimum value M^2. Overall, it can again be concluded that the iterative strategy permits achieving improved performance with very short training lengths.

VIII. Conclusion

In this paper, we have considered the problem of joint channel estimation and data detection in wireless communication systems. An iterative procedure has been proposed, whose basic idea is to recursively exploit the information detected bits in order to improve the channel-estimate accuracy and, eventually, the system-error probability. A detailed statistical analysis of the strategy has been developed, showing that, under mild conditions, the proposed iterative strategy brings substantial performance improvements over the conventional noniterative approach. Moreover, extensive computer simulations have shown that the experimental results are in good agreement with the results predicted by the theoretical analysis.

It is worth noting that an interesting open problem is to extend the performance analysis to coded systems operating over frequency-selective fading channels, which arise when the signal bandwidth exceeds the channel coherence bandwidth. While the consideration of frequency-selective channels appears to be quite straightforward, the consideration of convolutionally or turbo-coded system is, instead, much more challenging. On the other hand, the proposed techniques can be extended with moderate effort to the case that an orthogonal space–time block code (such as the well-known Alamouti scheme) is employed. These topics are currently being considered by the authors.

Appendix A
Proof of (43)

In order to compute the entries of the FIM $\boldsymbol{J}(h)$ in (42), first the entries of the matrix

$$\boldsymbol{J}'(h,h^*) = \begin{bmatrix} -E\left\{\frac{\partial^2\Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}}))}{\partial^2 h}\right\} & -E\left\{\frac{\partial^2\Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial h^*\partial h}\right\} \\ -E\left\{\frac{\partial^2\Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial h\partial h^*}\right\} & -E\left\{\frac{\partial^2\Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial^2 h^*}\right\} \end{bmatrix} \tag{90}$$

are computed and then the FIM $\boldsymbol{J}(h)$ is obtained through the identity [26, p. 1403]

$$\boldsymbol{J}(h) = \begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix} \boldsymbol{J}'(h,h^*) \begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix}^T. \tag{91}$$

First of all, recall that the following identities hold[12] [21, pp. 1400–1401]:

$$\frac{\partial \ln(\det[\boldsymbol{C}(h)])}{\partial h} = \text{trace}\left[\left(\frac{\partial \ln(\det[\boldsymbol{C}(h)])}{\partial \boldsymbol{C}}\right)^H \frac{\partial \boldsymbol{C}(h)}{\partial h}\right] \tag{92}$$

$$\frac{\partial \ln(\det[\boldsymbol{C}(h)])}{\partial \boldsymbol{C}} = \left(\boldsymbol{C}^{-1}(h)\right)^H \tag{93}$$

and

$$\frac{\partial (\boldsymbol{C}(h))^{-1}}{\partial h} = -\boldsymbol{C}^{-1}(h)\frac{\partial \boldsymbol{C}(h)}{\partial h}\boldsymbol{C}^{-1}(h). \tag{94}$$

Exploiting the above identities, it is easily shown that

$$\begin{aligned}-\frac{\partial \Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial h} &= \frac{\partial(\ln(\det[\boldsymbol{C}(h)]))}{\partial h}\\ &-\frac{\partial \boldsymbol{\mu}_r(h)^H}{\partial h}\boldsymbol{C}^{-1}(h)(\boldsymbol{r}-\boldsymbol{\mu}_r(h))+\\ &+(\boldsymbol{r}-\boldsymbol{\mu}_r(h))^H\left[-\boldsymbol{C}^{-1}(h)\frac{\partial \boldsymbol{C}(h)}{\partial h}\boldsymbol{C}^{-1}(h)\right](\boldsymbol{r}-\boldsymbol{\mu}_r(h))+\\ &-(\boldsymbol{r}-\boldsymbol{\mu}_r(h))^H\boldsymbol{C}^{-1}(h)\frac{\partial \boldsymbol{\mu}_r(h)}{\partial h}.\end{aligned} \tag{95}$$

Again deriving the above relation and taking the statistical expectation yields, after some lengthy algebra

$$-E\left\{\frac{\partial^2 \Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial h^2}\right\} = \text{trace}\left[\boldsymbol{C}^{-1}(h)\frac{\partial \boldsymbol{C}(h)}{\partial h}\boldsymbol{C}^{-1}(h)\frac{\partial \boldsymbol{C}(h)}{\partial h}\right]. \tag{96}$$

In deriving (96), we have used the fact that the derivative $\partial\boldsymbol{\mu}_r(h)^H/\partial h$ is zero. With similar steps, it can be also shown that

$$-E\left\{\frac{\partial^2 \Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial h^{*2}}\right\} = \text{trace}\left[\boldsymbol{C}^{-1}(h)\frac{\partial \boldsymbol{C}(h)}{\partial h^*}\boldsymbol{C}^{-1}(h)\frac{\partial \boldsymbol{C}(h)}{\partial h^*}\right] \tag{97}$$

where the fact that the derivative $\partial\boldsymbol{\mu}_r(h)/\partial h^*$ is zero has been exploited. Moreover, we also have

$$\begin{aligned}-E\left\{\frac{\partial^2 \Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial h^*\partial h}\right\} &= \text{trace}\left[\boldsymbol{C}^{-1}(h)\frac{\partial \boldsymbol{C}(h)}{\partial h}\boldsymbol{C}^{-1}(h)\frac{\partial \boldsymbol{C}(h)}{\partial h^*}\right]\\ &+\frac{\partial \boldsymbol{\mu}_r(h)^H}{\partial h^*}\boldsymbol{C}^{-1}(h)\frac{\partial \boldsymbol{\mu}_r(h)}{\partial h}\end{aligned} \tag{98}$$

and

$$\begin{aligned}-E\left\{\frac{\partial^2 \Lambda(\boldsymbol{r}|h,\widehat{\boldsymbol{b}})}{\partial h\partial h^*}\right\} &= \text{trace}\left[\boldsymbol{C}^{-1}(h)\frac{\partial \boldsymbol{C}(h)}{\partial h^*}\boldsymbol{C}^{-1}(h)\frac{\partial \boldsymbol{C}(h)}{\partial h}\right]\\ &+\frac{\partial \boldsymbol{\mu}_r(h)^H}{\partial h^*}\boldsymbol{C}^{-1}(h)\frac{\partial \boldsymbol{\mu}_r(h)}{\partial h}.\end{aligned} \tag{99}$$

[12]Note that, although not explicitly indicated, the matrix $\boldsymbol{C}(h)$ can be viewed as a function of h and h^*, since it is seen from (38) that it depends on $|h|^2 = hh^*$.

Now, since

$$\begin{aligned}\frac{\partial \boldsymbol{\mu}_r(h)^H}{\partial h^*}\boldsymbol{C}^{-1}(h)\frac{\partial \boldsymbol{\mu}_r(h)}{\partial h} &= A^2\widetilde{\boldsymbol{b}}^H\boldsymbol{C}^{-1}(h)\widetilde{\boldsymbol{b}}\\ &= \frac{A^2 T}{2\mathcal{N}_0} + \frac{A^2(L-T)(1-2p(e))^2}{\left(2\mathcal{N}_0 + A^2|h|^2\sigma_{e_b}^2\right)}\end{aligned} \tag{100}$$

it is easily shown that entries (96)–(99) of the matrix $\boldsymbol{J}'(h,h^*)$ can be expressed as shown in (101) at the bottom of the page. Substituting (101) into (91) leads to (43).

Appendix B
FIM Derivation for the Multiantenna Case

In order to obtain an expression for the FIM (68), it is convenient to define the $2NM$-dimensional complex vector $\widetilde{\boldsymbol{h}}_c = [\widetilde{\boldsymbol{h}}^T \quad (\widetilde{\boldsymbol{h}}^*)^T]^T$ and the matrix $\boldsymbol{J}'(\widetilde{\boldsymbol{h}}_c)$, whose (i,j)th element is defined as

$$\left[\boldsymbol{J}'(\widetilde{\boldsymbol{h}}_c)\right]_{i,j} = -E\left\{\frac{\partial^2\Lambda(\widetilde{\boldsymbol{r}}|\widetilde{\boldsymbol{h}},\widehat{\boldsymbol{B}})}{\partial\widetilde{h}_c(j)\partial\widetilde{h}_c(i)}\right\} \tag{102}$$

with $\widetilde{h}_c(i)$ denoting the ith entry of the vector $\widetilde{\boldsymbol{h}}_c$. Now, following a procedure similar to that for the single-antenna scenario, it is seen that

$$\begin{aligned}-E\left\{\frac{\partial^2\Lambda(\widetilde{\boldsymbol{r}}|\widetilde{\boldsymbol{h}},\widehat{\boldsymbol{B}}))}{\partial\widetilde{h}_c(m)\partial\widetilde{h}_c(n)}\right\} &= \text{trace}\left[\widetilde{\boldsymbol{C}}^{-1}(\widetilde{\boldsymbol{h}})\frac{\partial\widetilde{\boldsymbol{C}}(\widetilde{\boldsymbol{h}})}{\partial\widetilde{h}_c(n)}\widetilde{\boldsymbol{C}}^{-1}(\widetilde{\boldsymbol{h}})\frac{\partial\widetilde{\boldsymbol{C}}(\widetilde{\boldsymbol{h}})}{\partial\widetilde{h}_c(m)}\right]\\ &+\frac{\partial\boldsymbol{\mu}_{\widetilde{\boldsymbol{r}}}^H(\widetilde{\boldsymbol{h}})}{\partial\widetilde{h}_c(m)}\widetilde{\boldsymbol{C}}^{-1}(\widetilde{\boldsymbol{h}})\frac{\partial\boldsymbol{\mu}_{\widetilde{\boldsymbol{r}}}(\widetilde{\boldsymbol{h}})}{\partial\widetilde{h}_c(n)}+\frac{\partial\boldsymbol{\mu}_{\widetilde{\boldsymbol{r}}}^H(\widetilde{\boldsymbol{h}})}{\partial\widetilde{h}_c(n)}\widetilde{\boldsymbol{C}}^{-1}(\widetilde{\boldsymbol{h}})\frac{\partial\boldsymbol{\mu}_{\widetilde{\boldsymbol{r}}}(\widetilde{\boldsymbol{h}})}{\partial\widetilde{h}_c(m)}.\end{aligned} \tag{103}$$

In order to use (103), an explicit expression for its derivatives is to be worked out. It can be shown that

$$\frac{\partial\widetilde{\boldsymbol{C}}(\widetilde{\boldsymbol{h}})}{\partial\widetilde{h}_c(m)} = \begin{cases}\text{Diag}\Bigg(\underbrace{\boldsymbol{O}_N\ldots\boldsymbol{O}_N}_{T/M\ \text{times}}\quad\underbrace{\boldsymbol{R}_3\ldots\boldsymbol{R}_3}_{(L-T)/M\ \text{times}}\Bigg), & \forall m\le NM\\ \text{Diag}\Bigg(\underbrace{\boldsymbol{O}_N\ldots\boldsymbol{O}_N}_{T/M\ \text{times}}\quad\underbrace{\boldsymbol{R}_3^H\ldots\boldsymbol{R}_3^H}_{(L-T)/M\ \text{times}}\Bigg) & \forall m> NM\end{cases} \tag{104}$$

$$\boldsymbol{J}'(h,h^*) = \begin{bmatrix}\frac{A^4(L-T)\sigma_{e_b}^4 h^{*2}}{\left(2\mathcal{N}_0+A^2|h|^2\sigma_{e_b}^2\right)^2} & \frac{A^4(L-T)\sigma_{e_b}^4|h|^2}{\left(2\mathcal{N}_0+A^2|h|^2\sigma_{e_b}^2\right)^2}+\frac{A^2T}{2\mathcal{N}_0}+\frac{A^2(L-T)(1-2p(e))^2}{\left(2\mathcal{N}_0+A^2|h|^2\sigma_{e_b}^2\right)}\\ \frac{A^4(L-T)\sigma_{e_b}^4|h|^2}{\left(2\mathcal{N}_0+A^2|h|^2\sigma_{e_b}^2\right)^2}+\frac{A^2T}{2\mathcal{N}_0}+\frac{A^2(L-T)(1-2p(e))^2}{\left(2\mathcal{N}_0+A^2|h|^2\sigma_{e_b}^2\right)} & \frac{A^4(L-T)\sigma_{e_b}^4 h^2}{\left(2\mathcal{N}_0+A^2|h|^2\sigma_{e_b}^2\right)^2}\end{bmatrix}. \tag{101}$$

$$\frac{\partial \boldsymbol{\mu}_{\widetilde{\boldsymbol{r}}}{}^{H}(\widetilde{\boldsymbol{h}})}{\partial \widetilde{h}_c(n)} = \begin{cases} 0, & n=1,\ldots,NM \\ A^* \boldsymbol{I}_{NM}(n-NM,:)(\widetilde{\boldsymbol{B}}^T \otimes \boldsymbol{I}_N)^H, & n=NM+1,\ldots,2NM \end{cases} \quad (107)$$

where $\boldsymbol{R}_3$ is a matrix whose (i,j)th element is defined as

$$\boldsymbol{R}_3(i,j) = \begin{cases} 0 & \forall\, i,j=1,\ldots,N, i \neq p \\ A^2\sigma_{e_b}^2 \widetilde{h}_{j+kN}^* & \text{for } i=p,\ j=1,\ldots,N \end{cases} \quad (105)$$

with

$$p = \begin{cases} m \bmod N, & \text{if } m \bmod N \neq 0 \\ N, & \text{otherwise} \end{cases} \quad (106)$$

and $k = \lfloor (m-1)/N \rfloor$. Moreover, we have (107), shown at the top of the page, and

$$\frac{\partial \boldsymbol{\mu}_{\widetilde{\boldsymbol{r}}}(\widetilde{\boldsymbol{h}})}{\partial \widetilde{h}_c(n)} = \begin{cases} A(\widetilde{\boldsymbol{B}}^T \otimes \boldsymbol{I}_N)\boldsymbol{I}_{NM}(:,n), & n=1,\ldots,NM \\ 0, & n=NM+1,\ldots,2NM \end{cases}. \quad (108)$$

Substituting (104)–(108) in (102), a closed-form expression for the matrix $\boldsymbol{J}'(\widetilde{\boldsymbol{h}}_c)$ can be obtained. Finally, using the transformation [26, p. 1403]

$$\boldsymbol{J}(\widetilde{\boldsymbol{h}}_r) = \begin{bmatrix} \boldsymbol{I}_{NM} & \boldsymbol{I}_{NM} \\ j\boldsymbol{I}_{NM} & -j\boldsymbol{I}_{NM} \end{bmatrix} \boldsymbol{J}'(\widetilde{\boldsymbol{h}}_c) \begin{bmatrix} \boldsymbol{I}_{NM} & \boldsymbol{I}_{NM} \\ j\boldsymbol{I}_{NM} & -j\boldsymbol{I}_{NM} \end{bmatrix}^T \quad (109)$$

the FIM in (68) is obtained.

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor, Dr. K. J. Molnar, for his kind and collaborative assistance and for his efforts in ensuring a timely review process.

REFERENCES

[1] A. Lampe, "Iterative multiuser detection with integrated channel estimation for coded DS-CDMA," *IEEE Trans. Commun.*, vol. 50, pp. 1217–1223, Aug. 2002.
[2] A. O. Berthet, B. S. Sayrac, and R. Visoz, "Iterative decoding of convolutionally encoded signals over multipath Rayleigh fading channels," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1729–1743, Sept. 2001.
[3] R. Visoz, A. O. Berthet, and J. J. Boutros, "Reduced-complexity iterative decoding and channel estimation for space time BICM over frequency-selective wireless channels," in *IEEE Int. Symp. Personal, Indoor, Mobile Radio Commun. (PIMRC'02)*, vol. 3, 2002, pp. 1017–1022.
[4] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, 1999.
[5] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," in *Wireless Pers. Commun.*, 1998, vol. 6, pp. 311–335.
[6] B. Hochwald and T. L. Marzetta, "Capacity of a mobile multiple-antenna communication link in a Rayleigh flat-fading environment," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.
[7] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.
[8] G. Raleigh and J. M. Cioffi, "Spatial-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1999.
[9] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple-element antennas," *Bell Labs. Tech. J.*, vol. 1, pp. 41–59, 1996.
[10] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communications employing multi-element arrays," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1841–1852, Nov. 1999.
[11] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annu. Allerton Conf. Commun., Control, Computing*, Monticello, IL, Sept. 1999.
[12] A. Lozano and C. Papadias, "Layered space–time receivers for frequency-selective wireless channels," *IEEE Trans. Commun.*, vol. 50, pp. 65–73, Jan. 2002.
[13] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. Union Radio-Scientifique Int. (URSI) Symp. Signals, Systems, Electronics*, 1998, pp. 295–300.
[14] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space–time communication architecture," *Electron. Lett.*, vol. 35, pp. 14–16, 1999.
[15] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, pp. 951–963, Apr. 2003.
[16] 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 Processing*, vol. 44, pp. 1184–1197, May 1996.
[17] S. Talwar and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array—Part II: Performance analysis," *IEEE Trans. Signal Processing*, vol. 45, pp. 706–718, Mar. 1997.
[18] M. Loncar *et al.*, "Iterative joint detection, decoding, and channel estimation for dual antenna arrays in frequency selective fading," in *Proc. 5th Int. Symp. Wireless Personal Multimedia Communication*, Honolulu, HI, Oct. 2002, pp. 125–129.
[19] C. Cozzo and B. L. Hughes, "Joint channel estimation and data detection in space-time communications," *IEEE Trans. Commun.*, vol. 51, pp. 1266–1270, Aug. 2003.
[20] C. N. Georghiades and J. C. Han, "Sequence estimation in the presence of random parameters via the EM algorithm," *IEEE Trans. Commun.*, vol. 45, pp. 300–308, Mar. 1997.
[21] H. L. Van-Trees, *Optimum Array Processing*. New York: Wiley, 2002, pt. IV.
[22] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1285–1301, Nov. 1990.
[23] H. L. Van-Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 2001, pt. I.
[24] A. Roumy *et al.*, "Turbo-equalization: Convergence analysis," in *IEEE Int. Conf. Acoustistics, Speech, Signal Processing (ICASSP'01)*, vol. 4, 2001, pp. 2645–2648.
[25] S. ten Brink, "Convergence of iterative decoding," *Electron. Lett.*, vol. 35, pp. 1117–1119, June 1999.
[26] E. Aktas and U. Mitra, "Complexity reduction in subspace-based blind channel identification for DS/CDMA systems," *IEEE Trans. Commun.*, vol. 48, pp. 1392–1404, Aug. 2000.

Stefano Buzzi (M'98) was born in Piano di Sorrento, Italy, on December 10, 1970. He received the Dr. Eng. (Hons.) degree in 1994 and the Ph.D. degree in electronic engineering and computer science in 1999, both from the University of Naples "Federico II," Naples, Italy.

He spent six months at Centro Studi e Laboratori Telecomunicazioni (CSELT), Turin, Italy, in 1996. From November 1999 through December 2001, he spent eight months with the Department of Electrical Engineering, Princeton University, Princeton, NJ, as a Visiting Research Fellow. He currently is an Assistant Professor with the University of Cassino, Cassino, Italy. His current research and study interests are in statistical signal processing, with emphasis on signal detection in non-Gaussian noise and multiple access communications. He currently serves as an Associate Editor for the *Journal of Communications and Networks*.

Dr. Buzzi was awarded the "G. Oglietti" scholarship by the Associazione Elettrotecnica ed Elettronica Italiana (AEI) in 1996 and was the recipient of a North Atlantic Treaty Organization (NATO)/National Research Council (CNR) Advanced Fellowship in 1999 and of a CNR Short-Term Mobility Grant in 2000 and 2001.

Marco Lops (M'96–SM'01) was born in Naples, Italy, on March 16, 1961. He received the Dr. Eng. and Ph.D. degrees in electronic engineering from the University of Naples in 1986 and 1992, respectively.

He was an Engineer with the Air Traffic Control Systems Group, Selenia, Rome, Italy, from 1986 to 1987. In 1987, he joined the Department of Electronic and Telecommunications Engineering, University of Naples, as a Ph.D. student in electronic engineering. From 1991 to 2000, he was an Associate Professor of Radar Theory and Digital Transmission Theory with the University of Naples. Since March 2000, he has been a Full Professor with the University of Cassino, Cassino, Italy, where he is engaged in research in the field of statistical signal processing, with emphasis on radar processing and spread spectrum multiuser communications. He has also held teaching positions with the University of Lecce, Lecce, Italy, and, during 1991, 1998, and 2000, he was on sabbatical leave at the University of Connecticut, Storrs; Rice University, Houston, TX; and Princeton University, respectively. He currently serves as an Associate Editor for the *Journal of Communications and Networks*.

Stefania Sardellitti was born in Pontecorvo, Italy, on November 14, 1970. She received the Dr. Eng. degree in electronic engineering from the University of Rome, La Sapienza, Italy, in 1998. She currently is working toward the Ph.D. degree in electrical and information engineering at the University of Cassino, Cassino, Italy.

Her current research interests are in the area of signal processing for wireless communications, with emphasis on multiple-antenna and multiple-access systems.