Soft-iterative estimation of structured channels: performance analysis and comparison

S. Ferrara (*), M. Nicoli, U. Spagnolini
Dip. di Elettronica ed Informazione, Politecnico di Milano, Italy
e-mail: astersim@ee.oulu.fi, nicoli@elet.polimi.it, spagnoli@elet.polimi.it

***Abstract*—In this paper we propose a new soft-iterative method for the estimation of block-fading channels based on multi-block (MB) processing. The MB estimator exploits the invariance of the subspace spanned by the multipath components of the channel and it estimates the channel subspace by sample averaging over a frame of blocks. Here the MB method is extended to incorporate soft information, which is available in iterative equalizers. A performance evaluation and comparison with other soft methods in the literature is carried out showing the benefits of the proposed approach on the turbo equalizer convergence.**

***Index Terms*— Iterative receivers, Maximum-likelihood channel estimation, Multipath channels, Performance analysis, Soft processing, Subspace methods.**

I. INTRODUCTION

Iterative (turbo) equalization is a powerful technique that can be adopted at the receiver when data, protected by an error correction code, is transmitted over a frequency selective channel causing inter-symbol interference. The equalization and decoding tasks are performed iteratively on the same block of received signals, with exchange of soft information [1][2][3].

The convergence of these iterative receivers depends on the quality of channel state information (CSI) [4]. To improve the CSI accuracy, soft-iterative techniques refine the channel estimate using the soft information provided by the channel decoder. In this paper we focus on soft-iterative channel estimation for block-based wireless systems, where turbo processing is performed on a set of $L > 1$ data blocks, each containing both training and data symbols. We propose a new soft-based technique that performs a multi-block maximum-likelihood (MB-ML) estimation of the channel relying on the quasi-stationarity of the multipath delays within the L blocks. As in the block-by-block (or single-block, SB) estimation methods [4][5][6], the channel is estimated by combining both the hard-valued training and soft-valued information-bearing data. But differently from the above listed methods, the MB-ML method exploits multi-block measurements to improve the estimate accuracy for the slowly-varying subspace spanned by the multipath components. At first iteration the channel estimate is obtained as in [7] from the training signals only, while in the subsequent iterations it is refined using the a-priori statistics on the information-bearing symbols. The ML SB-MB methods are compared to [4][5][6], both analytically, by evaluating the mean square error (MSE) of the estimates, and through simulation results. To allow the comparison, the estimation is carried out for single-input-single-output (SISO) channels, but it can be also extended to SIMO/MIMO (single/multiple input multiple output) systems (see also [8]).

(*) S. Ferrara is now with CWC (Centre for Wireless Communication), University of Oulu, Finland.

The paper is organized as follows. Sec. II defines the signal model and the receiver structure. Soft-iterative channel estimation is in Sec. III, the analytical MSE evaluation and comparison in Sec. IV. Sec. V gives the simulation results and Sec. VI the concluding remarks.

II. SYSTEM DESCRIPTION

We consider a block-based transmission of convolutionally coded data over a frequency selective channel. A sequence of binary information symbols is convolutionally encoded with code rate R, permuted by a random interleaver $\Pi[\cdot]$ and mapped into quadrature phase-shift-keying (QPSK) symbols $x_\mathrm{d}(i) = (b(2i) + jb(2i+1))/\sqrt{2}$, where $b(i) \in \mathcal{D} = \{+1, -1\}$ denotes the ith code bit after interleaving. The transmission of the obtained QPSK sequence $\mathbf{x}_\mathrm{d} = [x_\mathrm{d}(1) \cdots x_\mathrm{d}(N_\mathrm{d}^{'} L)]^\mathrm{T}$ is organized (as shown in Fig 1) in L blocks of $N_\mathrm{d}^{'}$ symbols each: $\mathbf{x}_\mathrm{d} = [\mathbf{x}_\mathrm{d}^\mathrm{T}(1) \cdots \mathbf{x}_\mathrm{d}^\mathrm{T}(L)]^\mathrm{T}$, with $\mathbf{x}_\mathrm{d}(\ell) = [x_\mathrm{d}(1;\ell) \cdots x_\mathrm{d}(N_\mathrm{d}^{'};\ell)]^\mathrm{T}$ being the ℓth data block for $\ell = 1, \ldots, L$. An uncoded training sequence $\mathbf{x}_\mathrm{t}(\ell) = [x_\mathrm{t}(1;\ell) \cdots x_\mathrm{t}(N_\mathrm{t}^{'};\ell)]^\mathrm{T}$ is added as preamble within each block yielding an overall block length of $N' = N_\mathrm{t}^{'} + N_\mathrm{d}^{'}$ symbols. The L blocks are then transmitted over a frequency-selective channel, here modelled as a linear filter $\mathbf{h}(\ell) \in \mathbb{C}^{W \times 1}$ (including transmitter/receiver filters and multipath effects) that is constant within the block interval but varying from block to block (block-fading channel). The equivalent complex baseband model is shown in Fig. 2.

Let $y(i;\ell)$ denote the ith sample measured at the receiver within the ℓth block (after matched filtering and symbol-rate sampling), the vectors $\mathbf{y}_\mathrm{t}(\ell) = [y(W;\ell) \; \cdots \; y(N_\mathrm{t}^{'};\ell)]^\mathrm{T} \in \mathbb{C}^{N_\mathrm{t} \times 1}$ and $\mathbf{y}_\mathrm{d}(\ell) = [y(N_\mathrm{t}^{'} + W;\ell) \; \cdots \; y(N';\ell)]^\mathrm{T} \in \mathbb{C}^{N_\mathrm{d} \times 1}$ gather, respectively, the $N_\mathrm{t} = N_\mathrm{t}^{'} - W + 1$ and the $N_\mathrm{d} = N_\mathrm{d}^{'} - W + 1$ signals received within the training and the data-transmission phases. These signals are modelled as

$$\begin{cases} \mathbf{y}_\mathrm{t}(\ell) = \mathbf{X}_\mathrm{t}(\ell)\,\mathbf{h}(\ell) + \mathbf{w}_\mathrm{t}(\ell), & \text{Training} \\ \mathbf{y}_\mathrm{d}(\ell) = \mathbf{X}_\mathrm{d}(\ell)\,\mathbf{h}(\ell) + \mathbf{w}_\mathrm{d}(\ell), & \text{Data} \end{cases} \quad (1)$$

where the Toeplitz matrices $\mathbf{X}_\mathrm{t}(\ell) \in \mathbb{C}^{N_\mathrm{t} \times W}$ and $\mathbf{X}_\mathrm{d}(\ell) \in \mathbb{C}^{N_\mathrm{d} \times W}$ denote the convolution of, respectively, the training

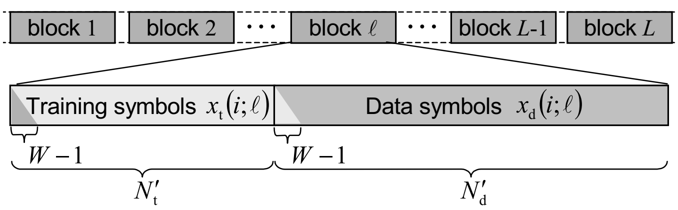

Fig. 1. Block-based transmission

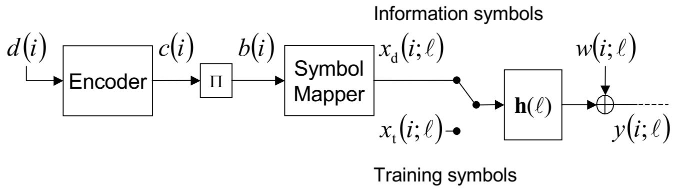

Fig. 2. Signal model for convolutionally coded system.

and data symbols with the channel: $[\mathbf{X}_{\mathrm{t}}(\ell)]_{m,n} = x_{\mathrm{t}}(W + n - m;\ell)$ and $[\mathbf{X}_{\mathrm{d}}(\ell)]_{m,n} = x_{\mathrm{d}}(W + n - m;\ell)$. Notice that the first $W-1$ samples at the beginning of each phase have been discarded to avoid overlapping between training and data and simplify the subsequent analysis (see Fig. 1). The vectors $\mathbf{w}_{\mathrm{t}}(\ell) = [w(W;\ell)\cdots w(N_{\mathrm{t}}^{'};\ell)]^{\mathrm{T}} \sim \mathcal{CN}(\mathbf{0},\sigma_w^2\mathbf{I}_{N_{\mathrm{t}}})$ and $\mathbf{w}_{\mathrm{d}}(\ell) = [y(N_{\mathrm{t}}^{'} + W;\ell)\cdots y(N^{'};\ell)]^{\mathrm{T}} \sim \mathcal{CN}(\mathbf{0},\sigma_w^2\mathbf{I}_{N_{\mathrm{d}}})$ collect uncorrelated complex Gaussian noise samples with zero mean and variance σ_w^2. The signal-to-noise ratio (SNR) is defined as $\rho = \mathrm{E}[||\mathbf{h}(\ell)||^2]/\sigma_w^2$.

The channel $\mathbf{h}(\ell) = \mathbf{G}(\boldsymbol{\tau})\boldsymbol{\alpha}(\ell)$ is the superposition of d paths having constant delays $\boldsymbol{\tau} = [\tau_1\cdots\tau_d]$ and block-fading amplitudes $\boldsymbol{\alpha}(\ell) = [\alpha_1(\ell)\cdots\alpha_d(\ell)]$. The kth column of the $W\times d$ matrix $\mathbf{G}(\boldsymbol{\tau}) = [\mathbf{g}(\tau_1)\cdots\mathbf{g}(\tau_d)]$ contains the system pulse waveform (convolution between the transmitter and the receiver filters), delayed by τ_k and sampled at the symbol rate. According to the Rayleigh fading and wide sense stationary uncorrelated scattering (WSSUS) assumptions, it is $\boldsymbol{\alpha}(\ell) \sim \mathcal{CN}(\mathbf{0},\mathbf{R}_\alpha)$ with power-delay-profile $\mathbf{R}_\alpha = \mathrm{diag}\{A_1,\ldots,A_d\}$. It follows that $\mathbf{h}(\ell) \sim \mathcal{CN}(\mathbf{0},\mathbf{R}_h)$, with covariance $\mathbf{R}_h = \mathbf{G}(\boldsymbol{\tau})\mathbf{R}_\alpha\mathbf{G}^{\mathrm{T}}(\boldsymbol{\tau})$. Notice that, since the columns of $\mathbf{G}(\boldsymbol{\tau})$ are not necessarily independent, it is $r = \mathrm{rank}[\mathbf{R}_h] \leq W$. The r-dimensional invariant subspace $\mathcal{R}(\mathbf{R}_h)$ is here referred to as the channel subspace. r represents the number of resolvable delays for the bandwidth of the transmitted signal [7]. Based on these assumptions, the channel vector can be rewritten in terms of the new parameters [7]

$$\mathbf{h}(\ell) = \mathbf{U}\mathbf{b}(\ell), \tag{2}$$

where $\mathbf{U} \in \mathbb{C}^{W\times r}$ is a constant full-column rank matrix having as column space the channel subspace $\mathcal{R}(\mathbf{U}) = \mathcal{R}(\mathbf{R}_h)$, while $\mathbf{b}(\ell) \in \mathbb{C}^{r\times 1}$ is a block-fading vector. A possible (but not unique) choice for $\mathbf{U}$ is the matrix containing the r eigenvectors (stationary channel modes) of $\mathbf{R}_h$, the corresponding modal amplitudes are $\mathbf{b}(\ell) = \mathbf{U}^{\mathrm{H}}\mathbf{h}(\ell) \sim \mathcal{CN}(\mathbf{0},\boldsymbol{\Lambda})$, where $\boldsymbol{\Lambda}$ is the $r\times r$ diagonal matrix containing the r eigenvalues of $\mathbf{R}_h$.

The iterative receiver structure is shown in Fig. 3. It consists of a soft-in channel estimator, a sliding-window soft-input-soft-output (SISO) minimum-mean-square-error (MMSE) linear equalizer [2][3] and a log-maximum-a-posteriori (log-MAP) SISO decoder [9], separated by an interleaver and a deinterleaver (of size $LN_{\mathrm{d}}^{'}$). At each iteration the soft channel estimator derives (as described later on in Sect. III) a new estimate for the channels $\{\mathbf{h}(\ell)\}_{\ell=1}^{L}$, by exploiting both the training symbols and the a-priori log-likelihood ratios (LLR) for the data-bearing bits $b(i)$: $\lambda_1[b(i)] = \log\{P[b(i) = +1]/P[b(i) = -1]\}$. The channel estimates are then used to perform equalization and decoding. Specifically, the SISO equalizer (or subsequently the decoder) computes the a-posteriori LLR $\Lambda_1[b(i)]$ (or $\Lambda_2[c(i)]$) about the code bits $b(i)$ (or $c(i)$). The extrinsic LLR is then provided at the output of the module as difference between the a-posteriori LLR and the a-priori LLR $\lambda_1[b(i)]$ (or $\lambda_2[c(i)]$). This refined soft information is de-interleaved (or interleaved) and it is passed to the decoder (or the equalizer) as new a-priori LLR $\lambda_1[c(i)]$ (or $\lambda_1[b(i)]$) for further processing.

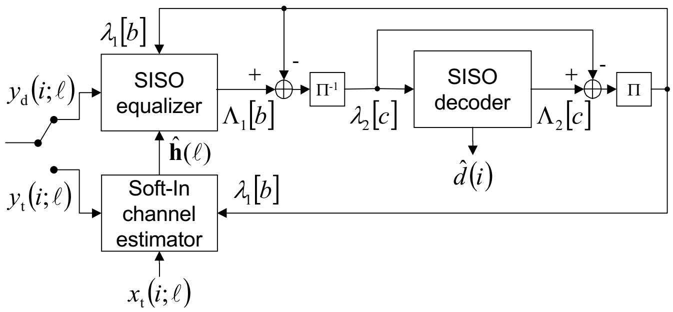

Fig. 3. Turbo receiver structure for convolutionally coded system.

III. Soft-based channel estimation

In the following we address the problem of estimation of the channel $\mathbf{h}(\ell)$ from the ensemble of L blocks $\{\mathbf{y}_{\mathrm{t}}(\ell),\mathbf{y}_{\mathrm{d}}(\ell)\}_{\ell=1}^{L}$. At first iteration the estimation is carried out from the training signals $\{\mathbf{y}_{\mathrm{t}}(\ell)\}_{\ell=1}^{L}$ only, using the knowledge of the pilot symbols $\mathbf{X}_{\mathrm{t}}$ (the a-priori information on $\mathbf{X}_{\mathrm{d}}(\ell)$ is missing). After the first channel estimation, equalization and decoding of the L blocks, the LLRs $\lambda_1[b(i)]$ can be exploited to refine the initial estimate. Namely, the a-priori LLRs are used to compute the mean value $\bar{x}_{\mathrm{d}}(i) = \mathrm{E}[x_{\mathrm{d}}(i)]$ and the variance $\sigma_{\mathrm{d}}^2(i) = \mathrm{E}[|\Delta x_{\mathrm{d}}(i)|^2]$, with $\Delta x_{\mathrm{d}}(i) = x_{\mathrm{d}}(i) - \bar{x}_{\mathrm{d}}(i)$, for each code symbol $x_{\mathrm{d}}(i)$, $i = 1,\ldots,LN_{\mathrm{d}}^{'}$. We recall that for QPSK modulation these statistics can be obtained as [3]

$$\bar{x}_{\mathrm{d}}(i) = \frac{1}{\sqrt{2}}\left(\tanh\frac{\lambda_1[b(2i)]}{2} + j\tanh\frac{\lambda_1[b(2i+1)]}{2}\right) \tag{3}$$

$$\sigma_{\mathrm{d}}^2(i) = 1 - |\bar{x}_{\mathrm{d}}(i)|^2. \tag{4}$$

Within the ℓth block, the quantities (3) and (4) will be indicated as $\bar{x}_{\mathrm{d}}(i;\ell)$ and $\sigma_{\mathrm{d}}^2(i;\ell)$, respectively. The convolution matrix built from the soft-valued data sequence $\{\bar{x}_{\mathrm{d}}(i;\ell)\}$ is $\bar{\mathbf{X}}_{\mathrm{d}}(\ell) = \mathrm{E}[\mathbf{X}_{\mathrm{d}}(\ell)] \in \mathbb{C}^{N_{\mathrm{d}}\times W}$, while $\Delta\mathbf{X}_{\mathrm{d}}(\ell) = \mathbf{X}_{\mathrm{d}}(\ell) - \bar{\mathbf{X}}_{\mathrm{d}}(\ell)$ is the matrix obtained from the data estimate errors $\{\Delta x_{\mathrm{d}}(i;\ell)\}$.

Some assumptions are herein made to perform channel estimation. The training sequence is assumed to be the same in all blocks so that $\mathbf{R}_{\mathrm{t}} = \mathbf{X}_{\mathrm{t}}^{\mathrm{H}}(\ell)\mathbf{X}_{\mathrm{t}}(\ell)$ is independent of the block index. The information-bearing symbols $\{x_{\mathrm{d}}(i;\ell)\}$ are independent and their number is large enough so that $\mathbf{R}_{\mathrm{d}} = \mathbf{X}_{\mathrm{d}}^{\mathrm{H}}(\ell)\mathbf{X}_{\mathrm{d}}(\ell) \approx N_{\mathrm{d}}\mathbf{I}_W$ and $\bar{\mathbf{R}}_{\mathrm{d}} = \bar{\mathbf{X}}_{\mathrm{d}}^{\mathrm{H}}(\ell)\bar{\mathbf{X}}_{\mathrm{d}}(\ell) \approx \tilde{N}_{\mathrm{d}}\mathbf{I}_W$. Here $\tilde{N}_{\mathrm{d}} = N_{\mathrm{d}}(1-\sigma_{\mathrm{d}}^2)$ represents the effective number of known data symbols depending on the average symbol variance $\sigma_{\mathrm{d}}^2 = \frac{1}{LN_{\mathrm{d}}^{'}}\sum_{i=1}^{LN_{\mathrm{d}}^{'}}\sigma_{\mathrm{d}}^2(i)$. We further assume $\Delta x_{\mathrm{d}}(i)$ as a stationary white process, independent from $w(i;\ell)$ and having variance σ_{d}^2.

Based on the assumptions above, the model (1) reduces to

$$\begin{cases} \mathbf{y}_{\mathrm{t}}(\ell) = \mathbf{X}_{\mathrm{t}}(\ell)\,\mathbf{h}(\ell) + \mathbf{w}_{\mathrm{t}}(\ell)\,, & \text{Training} \\ \mathbf{y}_{\mathrm{d}}(\ell) = \bar{\mathbf{X}}_{\mathrm{d}}(\ell)\,\mathbf{h}(\ell) + \Delta\mathbf{w}_{\mathrm{d}}(\ell) + \mathbf{w}_{\mathrm{d}}(\ell)\,, & \text{Data} \end{cases} \tag{5}$$

where the soft-valued data $\bar{\mathbf{X}}_{\mathrm{d}}(\ell)$ is known and can be treated as an extension of the training sequence, while $\Delta\mathbf{w}_{\mathrm{d}}(\ell) = \Delta\mathbf{X}_{\mathrm{d}}(\ell)\,\mathbf{h}(\ell)$ represents an additive noise term (independent from $\mathbf{w}_{\mathrm{d}}(\ell)$) having variance $\Delta\sigma_w^2 = \sigma_{\mathrm{d}}^2\sigma_w^2\rho$. With respect to the training-phase, in the data-transmission phase the input noise variance is virtually increased by a factor $1+\sigma_{\mathrm{d}}^2\rho$ to account for the unreliability of the soft values $\bar{x}_{\mathrm{d}}(i;\ell)$.

Below we consider soft-based channel estimation from (5), at first without imposing any specific constraint on the channel structure and then imposing the channel-subspace invariance constraint (2) (for known r). The unconstrained estimate of $\mathbf{h}(\ell)$ is obtained block by block from the single-block (SB) measurement $\{\mathbf{y}_{\mathrm{t}}(\ell), \mathbf{y}_{\mathrm{d}}(\ell)\}$. We briefly recall the following SB methods from the literature: the least squares (SB-LS) method [5]; the mixing (SB-M) and combining (SB-C) methods [4] (herein extended to the estimation of frequency-selective channels). Then we propose a new SB-ML method based on the white Gaussian assumption $\Delta\mathbf{w}_{\mathrm{d}}(\ell) \sim \mathcal{CN}(\mathbf{0}, \Delta\sigma_w^2\mathbf{I}_{N_{\mathrm{d}}})$. A structured soft-based approach is finally obtained by the ML estimation of $\mathbf{h}(\ell)$ parametrized as (2), from the multi-block (MB) ensemble $\{\mathbf{y}_{\mathrm{t}}(\ell), \mathbf{y}_{\mathrm{d}}(\ell)\}_{\ell=1}^{L}$.

A. SB channel estimation

1) SB-LS method [5]: By neglecting the term $\Delta\mathbf{w}_{\mathrm{d}}(\ell)$, the LS channel estimate from (5) yields:

$$\hat{\mathbf{h}}_{\mathrm{SBLS}}(\ell) = \left(\mathbf{R}_{\mathrm{t}} + \bar{\mathbf{R}}_{\mathrm{d}}\right)^{-1}[\mathbf{X}_{\mathrm{t}}^{\mathrm{H}}(\ell)\mathbf{y}_{\mathrm{t}}(\ell) + \bar{\mathbf{X}}_{\mathrm{d}}^{\mathrm{H}}(\ell)\mathbf{y}_{\mathrm{d}}(\ell)]. \tag{6}$$

2) SB-M method [4]: The mixing method (fully equivalent to [6]) estimates the channel vector by minimizing

$$\mathcal{F}_{\mathrm{M}} = ||\mathbf{y}_{\mathrm{t}}(\ell) - \mathbf{X}_{\mathrm{t}}(\ell)\mathbf{h}(\ell)||^2 + \mathrm{E}_{x_{\mathrm{d}}}[||\mathbf{y}_{\mathrm{d}}(\ell) - \mathbf{X}_{\mathrm{d}}(\ell)\mathbf{h}(\ell)||^2]. \tag{7}$$

The solution can be written as

$$\hat{\mathbf{h}}_{\mathrm{SBM}}(\ell) = \left(\mathbf{R}_{\mathrm{t}} + \overline{\overline{\mathbf{R}}}_{\mathrm{d}}\right)^{-1}[\mathbf{X}_{\mathrm{t}}^{\mathrm{H}}(\ell)\mathbf{y}_{\mathrm{t}}(\ell) + \bar{\mathbf{X}}_{\mathrm{d}}^{\mathrm{H}}(\ell)\mathbf{y}_{\mathrm{d}}(\ell)] \tag{8}$$

where $\overline{\overline{\mathbf{R}}}_{\mathrm{d}} = \mathrm{E}[\mathbf{X}_{\mathrm{d}}^{\mathrm{H}}(\ell)\mathbf{X}_{\mathrm{d}}(\ell)] = N_{\mathrm{d}}\mathbf{I}_W$.

3) SB-C method [4]: The combining method is a linear combination of the unbiased training-based LS estimate $\hat{\mathbf{h}}_{\mathrm{t}}(\ell) = \mathbf{R}_{\mathrm{t}}^{-1}\mathbf{X}_{\mathrm{t}}^{\mathrm{H}}(\ell)\mathbf{y}_{\mathrm{t}}(\ell)$ and the biased data-based estimate $\hat{\mathbf{h}}_{\mathrm{d}}(\ell) = \overline{\overline{\mathbf{R}}}_{\mathrm{d}}^{-1}\bar{\mathbf{X}}_{\mathrm{d}}^{\mathrm{H}}(\ell)\mathbf{y}_{\mathrm{d}}(\ell)$:

$$\hat{\mathbf{h}}_{\mathrm{SBC}}(\ell) = \mathbf{T}\hat{\mathbf{h}}_{\mathrm{t}}(\ell) + \mathbf{D}\hat{\mathbf{h}}_{\mathrm{d}}(\ell)\,. \tag{9}$$

The weighting matrices $\mathbf{D}$ and $\mathbf{T}$ are selected so as to minimize the estimate variance $\mathcal{F}_{\mathrm{C}} = \mathrm{E}_w[||\hat{\mathbf{h}}_{\mathrm{SBC}}(\ell) - \mathbf{h}(\ell)||^2]$ under the constraint $\mathrm{E}_w[\hat{\mathbf{h}}_{\mathrm{SBC}}(\ell)] = \mathbf{h}(\ell)$. Following the same procedure as in [4] it can be shown that the minimization yields

$$\mathbf{D} = N_{\mathrm{d}}\mathbf{R}_{\mathrm{d}}'^{\mathrm{H}}\left(\mathbf{R}_{\mathrm{d}}'\mathbf{R}_{\mathrm{d}}'^{\mathrm{H}} + \tilde{N}_{\mathrm{d}}\mathbf{R}_{\mathrm{t}}\right)^{-1} \tag{10}$$

$$\mathbf{T} = \mathbf{I}_W - \mathbf{R}_{\mathrm{d}}'^{\mathrm{H}}\left(\mathbf{R}_{\mathrm{d}}'\mathbf{R}_{\mathrm{d}}'^{\mathrm{H}} + \tilde{N}_{\mathrm{d}}\mathbf{R}_{\mathrm{t}}\right)^{-1}\mathbf{R}_{\mathrm{d}}' \tag{11}$$

where $\mathbf{R}_{\mathrm{d}}' = \bar{\mathbf{X}}_{\mathrm{d}}^{\mathrm{H}}(\ell)\mathbf{X}_{\mathrm{d}}(\ell)$. Similarly to [4], since the solution depends on the unknown symbols $\mathbf{X}_{\mathrm{d}}(\ell)$ the matrix $\mathbf{R}_{\mathrm{d}}'$ needs to be approximated, e.g. as $\mathbf{R}_{\mathrm{d}}' \approx N_{\mathrm{d}}(1-2\varepsilon)\bar{b}$ where ε is the bit error probability at the output of the SISO decoder and $\bar{b} = \frac{1}{2LN_{\mathrm{d}}'}\sum_{i=1}^{2LN_{\mathrm{d}}'}|\mathrm{E}[b(i)]|$. The disadvantage of this approach is the need of estimating ε. Notice also that for large N_{d} the SB-C estimate (9) reduces to the SB-LS estimate (6) as it is: $\mathbf{R}_{\mathrm{d}}' \approx \bar{\mathbf{R}}_{\mathrm{d}}$, $\mathbf{D} \approx N_{\mathrm{d}}\left(\mathbf{R}_{\mathrm{t}} + \bar{\mathbf{R}}_{\mathrm{d}}\right)^{-1}$ and $\mathbf{T} \approx \mathbf{I}_W - \tilde{N}_{\mathrm{d}}\left(\mathbf{R}_{\mathrm{t}} + \bar{\mathbf{R}}_{\mathrm{d}}\right)^{-1}$.

4) ML method: The SB-ML estimate, under the Gaussian approximation for $\Delta\mathbf{w}_{\mathrm{d}}(\ell)$, is the minimizer of the negative log-likelihood function

$$\mathcal{L}(\ell) = \|\mathbf{y}_{\mathrm{t}}(\ell) - \mathbf{X}_{\mathrm{t}}(\ell)\,\mathbf{h}(\ell)\|^2 + \gamma||\mathbf{y}_{\mathrm{d}}(\ell) - \bar{\mathbf{X}}_{\mathrm{d}}(\ell)\,\mathbf{h}(\ell)\,||^2 \tag{12}$$

for $\gamma = (1+\sigma_{\mathrm{d}}^2\rho)^{-1}$ (factor due to the noise non-stationarity). By setting $\bar{\mathbf{R}} = \mathbf{R}_{\mathrm{t}} + \gamma\bar{\mathbf{R}}_{\mathrm{d}}$, the minimization yields

$$\hat{\mathbf{h}}_{\mathrm{SB}}(\ell) = \bar{\mathbf{R}}^{-1}\left(\mathbf{X}_{\mathrm{t}}^{\mathrm{H}}(\ell)\,\mathbf{y}_{\mathrm{t}}(\ell) + \gamma\bar{\mathbf{X}}_{\mathrm{d}}^{\mathrm{H}}(\ell)\,\mathbf{y}_{\mathrm{d}}(\ell)\right). \tag{13}$$

B. MB channel estimation

By generalizing [7] to model (5), the MB-ML estimate can be obtained as the minimizer of $\mathcal{L} = \sum_{\ell=1}^{L}\mathcal{L}(\ell)$ constrained to the channel structure (2), yielding:

$$\hat{\mathbf{h}}_{\mathrm{MB}}(\ell) = \bar{\mathbf{R}}^{-1/2}\hat{\mathbf{P}}\bar{\mathbf{R}}^{1/2}\hat{\mathbf{h}}_{\mathrm{SB}}(\ell)\,. \tag{14}$$

$\hat{\mathbf{P}}$ represents the estimate of the projector onto the channel subspace and it is obtained from the r leading eigenvectors of

$$\mathbf{R}_{\mathrm{MB}}(L) = \frac{1}{L}\bar{\mathbf{R}}^{1/2}\left(\sum\nolimits_{\ell=1}^{L}\hat{\mathbf{h}}_{\mathrm{SB}}(\ell)\,\hat{\mathbf{h}}_{\mathrm{SB}}^{\mathrm{H}}(\ell)\right)\bar{\mathbf{R}}^{\mathrm{H}/2}. \tag{15}$$

Notice that the soft ML-MB estimate (14) reduces to the hard-based one proposed in [7] when the a-priori information is missing ($\sigma_{\mathrm{d}}^2 \approx 1$, $\tilde{N}_{\mathrm{d}} \approx 0$) or perfect ($\sigma_{\mathrm{d}}^2 \approx 0$, $\tilde{N}_{\mathrm{d}} \approx N_{\mathrm{d}}$).

IV. Performance analysis and comparison

We evaluate and compare the MSE for the SB methods (SB-LS, SB-M, SB-C, SB-ML) and the MB-ML method for $L \to \infty$ (i.e., perfect knowledge of the temporal subspace). The following definitions will be used: $\mathbf{n}_{\mathrm{t}}(\ell) = \mathbf{X}_{\mathrm{t}}^{\mathrm{H}}(\ell)\mathbf{w}_{\mathrm{t}}(\ell)$, $\mathbf{n}_{\mathrm{d}}(\ell) = \bar{\mathbf{X}}_{\mathrm{d}}^{\mathrm{H}}(\ell)\mathbf{w}_{\mathrm{d}}(\ell)$, $\Delta\mathbf{n}_{\mathrm{d}}(\ell) = \bar{\mathbf{X}}_{\mathrm{d}}^{\mathrm{H}}(\ell)\Delta\mathbf{w}_{\mathrm{d}}(\ell)$, $\Delta N_{\mathrm{d}} = N_{\mathrm{d}} - \tilde{N}_{\mathrm{d}}$. $\mathbf{P}$ denotes the true projector onto the temporal subspace $\mathcal{R}[\bar{\mathbf{R}}^{1/2}\mathbf{R}_h\bar{\mathbf{R}}^{\mathrm{H}/2}]$.

From the estimate definitions in Sec. III and model (5), the estimate error $\Delta\mathbf{h}(\ell) = \hat{\mathbf{h}}(\ell) - \mathbf{h}(\ell)$ for all methods can be written as indicated in Table I (rows 2-6). The MSE $\mathrm{E}[||\Delta\mathbf{h}(\ell)||^2]$ is evaluated by averaging over fading, noise and information-bearing data, and using the uncorrelation of $\mathbf{n}_{\mathrm{t}}(\ell)$, $\Delta\mathbf{n}_{\mathrm{d}}(\ell)$ and $\mathbf{n}_{\mathrm{d}}(\ell)$ (the proofs are omitted for lack of space). The resulting MSEs are listed in Table II (rows 8-12), together with the MSEs for uncorrelated training sequence, i.e. for $\mathbf{R}_{\mathrm{t}} = N_{\mathrm{t}}\mathbf{I}_W$ and $\bar{\mathbf{R}} = (N_{\mathrm{t}} + \gamma\tilde{N}_{\mathrm{d}})\mathbf{I}_W$ (rows 14-18). To simplify the performance comparison, in the following we focus on this second case (uncorrelated training sequence).

The comparison between the SB performances shows that the ML approach has some definite advantages with respect to the

TABLE I
MSE OF SOFT-BASED SB AND MB ESTIMATES.

Method	Estimate error
SB-LS	$(\mathbf{R}_\mathrm{t}+\bar{\mathbf{R}}_\mathrm{d})^{-1}[\mathbf{n}_\mathrm{t}(\ell)+\mathbf{n}_\mathrm{d}(\ell)+\Delta\mathbf{n}_\mathrm{d}(\ell)]$
SB-M	$(\mathbf{R}_\mathrm{t}+\bar{\bar{\mathbf{R}}}_\mathrm{d})^{-1}[\mathbf{n}_\mathrm{t}(\ell)+\mathbf{n}_\mathrm{d}(\ell)+\Delta\mathbf{n}_\mathrm{d}(\ell)-\Delta N_\mathrm{d}\mathbf{h}(\ell)]$
SB-C	$\mathbf{T}\mathbf{R}_\mathrm{t}^{-1}\mathbf{n}_\mathrm{t}(\ell)+\mathbf{D}\bar{\bar{\mathbf{R}}}_\mathrm{d}^{-1}\mathbf{n}_\mathrm{d}(\ell)$
SB-ML	$\bar{\mathbf{R}}^{-1}\{\mathbf{n}_\mathrm{t}(\ell)+\gamma[\mathbf{n}_\mathrm{d}(\ell)+\Delta\mathbf{n}_\mathrm{d}(\ell)]\}$
MB-ML	$\bar{\mathbf{R}}^{-1/2}\mathbf{P}\bar{\mathbf{R}}^{-\mathrm{H}/2}\{\mathbf{n}_\mathrm{t}(\ell)+\gamma[\mathbf{n}_\mathrm{d}(\ell)+\Delta\mathbf{n}_\mathrm{d}(\ell)]\}$
Method	**MSE for correlated training sequence**
SB-LS	$\sigma_w^2\,\mathrm{tr}[(\mathbf{R}_\mathrm{t}+\bar{\mathbf{R}}_\mathrm{d})^{-2}(\mathbf{R}_\mathrm{t}+\gamma^{-1}\bar{\mathbf{R}}_\mathrm{d})]$
SB-M	$\sigma_w^2\,\mathrm{tr}\left\{(\mathbf{R}_\mathrm{t}+\bar{\bar{\mathbf{R}}}_\mathrm{d})^{-2}\left[\mathbf{R}_\mathrm{t}+\gamma^{-1}\bar{\mathbf{R}}_\mathrm{d}+\frac{\Delta N_\mathrm{d}^2}{\sigma_w^2}\mathbf{R}_h\right]\right\}$
SB-C	$\sigma_w^2\,\mathrm{tr}\left\{\mathrm{E}_{x_\mathrm{d}}[\mathbf{T}\mathbf{R}_\mathrm{t}^{-1}\mathbf{T}^\mathrm{H}]+\mathrm{E}_{x_\mathrm{d}}[\mathbf{D}\mathbf{D}^\mathrm{H}]\tilde{N}_\mathrm{d}/N_\mathrm{d}^2\right\}$
SB-ML	$\sigma_w^2\,\mathrm{tr}[(\mathbf{R}_\mathrm{t}+\gamma\bar{\mathbf{R}}_\mathrm{d})^{-1}]$
MB-ML	$\sigma_w^2\,\mathrm{tr}[(\mathbf{R}_\mathrm{t}+\gamma\bar{\mathbf{R}}_\mathrm{d})^{-1/2}\mathbf{P}(\mathbf{R}_\mathrm{t}+\gamma\bar{\mathbf{R}}_\mathrm{d})^{-\mathrm{H}/2}]$
Method	**MSE for uncorrelated training sequence**
SB-LS	$\sigma_w^2\left[W\left(N_\mathrm{t}+\gamma^{-1}\tilde{N}_\mathrm{d}\right)\right]/(N_\mathrm{t}+\tilde{N}_\mathrm{d})^2$
SB-M	$\sigma_w^2\left[W\left(N_\mathrm{t}+\gamma^{-1}\tilde{N}_\mathrm{d}\right)+\rho\Delta N_\mathrm{d}^2\right]/(N_\mathrm{t}+N_\mathrm{d})^2$
SB-C	$\sigma_w^2\,\mathrm{tr}\,\mathrm{E}_{x_\mathrm{d}}[\mathbf{T}\mathbf{T}^\mathrm{H}]/N_\mathrm{t}+\sigma_w^2\,\mathrm{tr}\,\mathrm{E}_{x_\mathrm{d}}[\mathbf{D}\mathbf{D}^\mathrm{H}]\tilde{N}_\mathrm{d}/N_\mathrm{d}^2$
SB-ML	$\sigma_w^2 W/(N_\mathrm{t}+\gamma\tilde{N}_\mathrm{d})$
MB-ML	$\sigma_w^2 r/(N_\mathrm{t}+\gamma\tilde{N}_\mathrm{d})$

mixing, combining and LS techniques. First of all, let us consider the estimate errors in Table I. It can be seen that the SB-ML estimate is always unbiased, while the SB-M estimate is affected by the bias $\mathrm{E}[\hat{\mathbf{h}}_{\text{SB-M}}(\ell)]=[1+\sigma_\mathrm{d}^2/(N_\mathrm{t}/N_\mathrm{d}+1)]\mathbf{h}(\ell)\neq\mathbf{0}$ for $\sigma_\mathrm{d}^2\neq 0$. As highlighted in [4], for $\sigma_\mathrm{d}^2\approx 1$ (unreliable soft data) and $N_\mathrm{d}\gg N_\mathrm{t}$ this bias can be very large, preventing the turbo receiver to bootstrap.

The results in Table I show also that, for $\sigma_\mathrm{d}^2=1$, the MSEs for the SB-LS, SB-C and SB-ML methods reduce to $\mathrm{MSE}_\mathrm{t}=\sigma_w^2 W/N_\mathrm{t}$, i.e. the performance of the conventional training-based estimate $\hat{\mathbf{h}}_\mathrm{t}(\ell)$ (defined as in Sec. III-A). On the other hand, for large SNR ρ and $N_\mathrm{d}\gg N_\mathrm{t}$ the MSE of the SB-M method is $\mathrm{MSE}_{\text{SB-M}}/\mathrm{MSE}_\mathrm{t}\approx\rho N_\mathrm{t}/W\gg 1$. This implies that the SB-M receiver performance can even deteriorate for increasing number of iterations (as shown by the simulation results in Sec. V). Similarly, for $\sigma_\mathrm{d}^2=0$ (i.e., for $\tilde{N}_\mathrm{d}=N_\mathrm{d}$ and $\bar{\mathbf{X}}_\mathrm{d}(\ell)=\mathbf{X}_\mathrm{d}(\ell)$), all SB methods reduce to the LS estimate calculated from the overall sequence of $N_\mathrm{t}+N_\mathrm{d}$ known symbols, reaching the lowest MSE value $\mathrm{MSE}_{\mathrm{t+d}}=\sigma_w^2 W/(N_\mathrm{t}+N_\mathrm{d})$.

Let us now consider the intermediate values $\sigma_\mathrm{d}^2\in(0,1)$. By comparing row 17 in Table I with MSE_t it can be seen that the SB-ML estimate is always more accurate than the conventional training-based one. This improved accuracy, provided for any grade of reliability on the data estimate, is another advantage of the ML approach. On the other hand, the SB-LS and SB-M methods perform worse than $\hat{\mathbf{h}}_\mathrm{t}(\ell)$ for, respectively, $\sigma_\mathrm{d}^2\rho>(1+\tilde{N}_\mathrm{d}/N_\mathrm{t})$ and $\sigma_\mathrm{d}^2\rho>(N_\mathrm{d}/N_\mathrm{t}+\Delta N_\mathrm{d}/N_\mathrm{d}+1)/(\tilde{N}_\mathrm{d}/N_\mathrm{d}+N_\mathrm{d}\Delta N_\mathrm{d}/W)$ (i.e., for unreliable data and large SNR).

Let us finally compare the SB-ML and MB-ML methods (rows 17-18). For these estimators the MSE depends only on the ratio between the number of channel unknowns and the number of *effective* training symbols within each block $(N_\mathrm{t}+\gamma\tilde{N}_\mathrm{d})$. The latter number ranges from N_t (in case of missing prior information) to $N_\mathrm{t}+N_\mathrm{d}$ (for perfect prior information). On the other hand, the number of unknowns is W for the SB estimator, while for the MB estimator it is reduced to the number r of block-dependent amplitudes $\mathbf{b}(\ell)$ [7], as the projector $\mathbf{P}$ is perfectly estimated for $L\to\infty$.

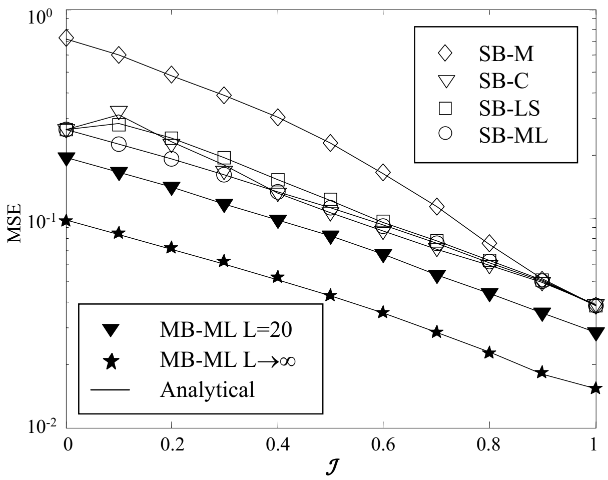

Fig. 4. MSE for soft-based SB/MB estimates vs. mutual information.

To conclude, the ML method is always unbiased and it outperforms the training-based estimate (and also all other SB methods as proved by simulations in Sec. V) for any $0\leq\sigma_\mathrm{d}^2\leq 1$. Furthermore, when integrated with MB processing, the ML approach allows a further MSE reduction by a factor W/r with respect to the SB approach, yielding the most accurate estimate among all other methods considered in this paper.

V. NUMERICAL RESULTS

A frame of 10000 randomly-chosen equiprobable information bits is coded by a 4-state convolutional code with generators $(7,5)_o$ and it is permuted by a random interleaver. The code bits are mapped into 10000 QPSK symbols and arranged into $L=20$ blocks with $N_\mathrm{d}^{'}=500$ symbols each. A training sequence of $N_t^{'}=46$ QPSK symbols is obtained by adding a cyclic prefix to the sequence of length $N_\mathrm{t}=31$ defined in [5]. The block-fading Rayleigh channel is simulated according to the following multipath models: $d=r=6$ resolvable paths, delays $\boldsymbol{\tau}=[0,1.2,2.2,8.2,9.2,10.2]\mu$s, mean powers $\mathbf{R}_\alpha(0)=\frac{2}{7}\mathrm{diag}\{1,\frac{1}{2},\frac{1}{4},1,\frac{1}{2},\frac{1}{4}\}$ (Model C1) ; $d=r=9$ resolvable paths, delays $\boldsymbol{\tau}=[0,1.1,2.2,5.8,6.2,7.2,10.2,11.2,12.6]\,\mu$s, mean powers $\mathbf{R}_\alpha(0)=\frac{4}{41}\mathrm{diag}\{1,\frac{1}{2},\frac{1}{4},3,\frac{3}{2},\frac{5}{4},2,\frac{1}{2},\frac{1}{4}\}$ (Model C2). The MB method is simulated using as $\hat{\mathbf{P}}$ the estimate drawn from $L=20$ blocks or the true projector onto the channel subspace (as for an estimate from $L\to\infty$ blocks).

Fig. 4 compares the analytical (lines) and simulated (markers) MSE for: all SB methods ($\mathrm{MSE}_{\text{SB-M}}$, $\mathrm{MSE}_{\text{SB-C}}$, $\mathrm{MSE}_{\text{SB-LS}}$, $\mathrm{MSE}_{\text{SB-ML}}$) and the ML-MB method ($\mathrm{MSE}_{\text{MB-ML}}$), $\rho=3$dB, channel model C1, mutual informa-

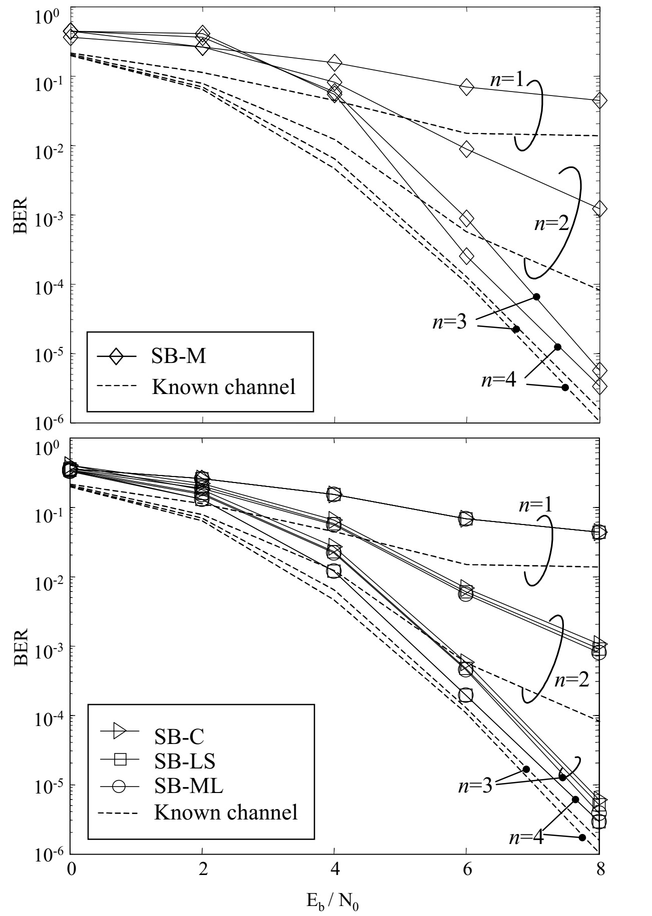

Fig. 5. BER performance for SB methods vs. E_b/N_0 for varying number of iteration (n): SB-M (top); SB-C, SB-LS and SB-ML (bottom).

tion $\mathcal{I} = 0 \div 1$. The latter is defined as the mutual information $\mathcal{I} = \mathcal{I}\{b(i), \lambda_1[b(i)]\}$ between the input code bits $b(i)$ and the output a-priori information $\lambda_1[b(i)]$ here approximated as Gaussian according to [10]. The number of data symbols used for channel estimation is $N_\mathrm{d} = 200$. Fig. 4 confirms that, for $\mathcal{I} = 0$ or $\mathcal{I} = 1$, all SB methods have the same performance (except SB-M that is affected by the estimate bias for low $\mathcal{I}$). The MSE reached for $\mathcal{I} = 0$ and $\mathcal{I} = 1$ equals the performance of the LS estimate evaluated from, respectively: the training sequence only ($\mathrm{MSE_t}$) and the whole training-data sequence when all symbols are known ($\mathrm{MSE_{t+d}} \approx \mathrm{MSE_t} \cdot 31/200$). As pointed out in Sec. IV, it is $\mathrm{MSE_{SB\text{-}ML}} \leq \mathrm{MSE_t}$ for all $\mathcal{I}$ values, while for low $\mathcal{I}$ the MSE of the estimate is above $\mathrm{MSE_t}$ for both the SB-C method (for $\mathcal{I} \leq 0.5$) and the SB-LS methods (for $\mathcal{I} \leq 0.2$). For instance, for $\mathcal{I} = 0.1$ ($\tilde{N}_\mathrm{d} \approx 15$), the simulation confirms the MSE gains/losses evaluated from Table I: $\mathrm{MSE_{SB\text{-}LS}}/\mathrm{MSE_t} \approx \mathrm{MSE_{SB\text{-}C}}/\mathrm{MSE_t} = 0.34$dB, $\mathrm{MSE_{SB\text{-}M}}/\mathrm{MSE_t} \approx 4$dB, $\mathrm{MSE_{ML}}/\mathrm{MSE_t} \approx -0.67$dB. It is also shown that $\mathrm{MSE_{SB\text{-}ML}}/\mathrm{MSE_{MB\text{-}ML}} \approx 4.2$dB for $L \to \infty$.

Fig. 5 and 6 show the BER performance vs. E_b/N_0 for the complete iterative receiver after $n = 1 \div 4$ iterations. The channel model is simulated according to model C1 (Fig. 5) and C2 (Fig. 6). The number of data symbols used for soft-based channel estimation is $N_\mathrm{d} = 500$. Fig. 5 compares the receiver for known channel with the receiver based on SB-M (top figure) and SB-C/LS/ML (bottom figure) channel estimation. It can

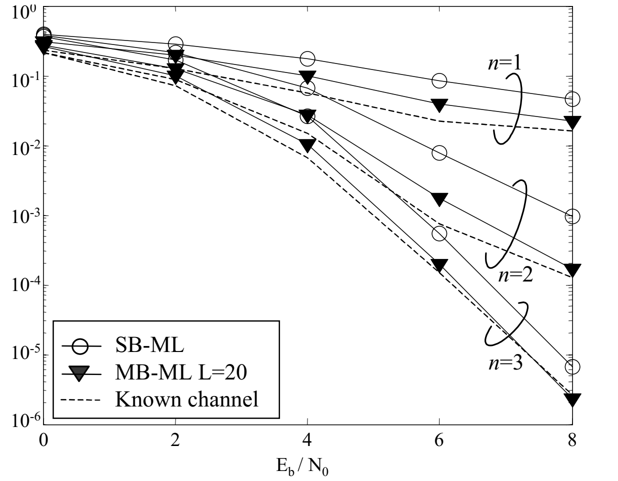

Fig. 6. BER performance for SB-ML and MB-ML methods vs. E_b/N_0 for varying number of iteration (n).

be seen that for large E_b/N_0 all SB methods have comparable performance, with SB-ML performing slightly better than the others. On the other hand, for $E_b/N_0 \leq 4$dB the bias of the SB–M estimate is shown to worsen the receiver performance for increasing n. Fig. 6 compares the ML-SB and ML-MB estimators highlighting the advantages of the MB approach: the performance of the MB receiver is far below the SB one and it is very close to the known-channel lower bound.

VI. Concluding remarks

This paper proposes the integration of the training-based MB estimate [7] for block-fading channels with soft iterative equalization. The soft ML-MB method exploits the invariance of the channel subspace across blocks and it estimates the channel using the soft statistics on the information-bearing data. The analytical performance evaluation and comparison with other soft methods in the literature, also validated by simulation results, show the advantages of the proposed method.

References

[1] C. Douillard, M. Jézéquel, C. Berrou, "Iterative correction of intersymbol interference, turbo-equalization," *Eur. Trans. Telecommun.*, vol. 6, pp. 507-511, Sept.-Oct. 1995.

[2] H.V. Poor, "Iterative Multiuser Detection," *IEEE Signal Processing Mag.,* vol. 21, No. 1, pp. 67-80, Jan. 2004, pp. 81-88, Jan. 2004.

[3] R. Koetter, A.C. Singer and M. Tüchler, "Turbo equalization", *IEEE Signal Processing Mag.,* vol. 21, no. 1, pp. 67-80, Jan. 2004.

[4] M. Kobayashi, J. Boutros, G. Caire, "Successive interference cancellation with SISO decoding and EM channel estimation," *IEEE J. Select. Areas Commun.*, vol. 19, no. 8, pp. 1450-1460, Aug. 2001.

[5] M. Tuchler, R. Otnes, A. Schmidbauer, "Performance of soft iterative channel estimation in turbo equalization," *Proc. IEEE ICC 2002*, pp. 1858-1862.

[6] S. Song, A. Singer, K-M. Sung, "Turbo equalization with an unknown channel," *Proc. IEEE ICASSP 2002*, Vol. 3, pp. 2805-2808.

[7] M. Nicoli, O. Simeone, U. Spagnolini, "Multi-slot estimation of frequency-selective fast-varying channels," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1337-1347, Aug. 2003.

[8] M. Nicoli and U. Spagnolini, "Subspace-methods for space-time processing," in *Smart Antennas State of the Art*, Chapter 1: Receiver Processing, EURASIP-Book series, Hindawi Publ. Corp., to be published, 2005.

[9] L. R. Bahl, J. Cocke, F. Jelinek, J. Ravin, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Infor. Theory*, vol. 20, pp. 284-287, Mar. 1974.

[10] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727-1737, Oct. 2001.