A SUBSPACE METHOD FOR CHANNEL ESTIMATION IN SOFT-ITERATIVE RECEIVERS

M. Nicoli and U. Spagnolini
Dip. di Elettronica e Informazione, Politecnico di Milano, Italy, e-mail {nicoli, spagnoli}@elet.polimi.it

ABSTRACT

In this paper we propose a new *soft* method for the estimation of block-fading channels based on multi-block (MB) processing. The MB estimator [1] 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 also *soft* information, which is available in iterative (turbo) equalizers. The mean square error (MSE) of the soft-based estimate is evaluated analytically and validated by simulations. The comparison with the conventional training-based block-by-block estimate shows the benefits of the proposed approach on the turbo equalizer convergence.

1. 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 (ISI) and/or co-channel interference. The equalization and decoding tasks are performed iteratively on the same block of received signals, with exchange of soft information, so as to refine the data estimate [2].

It is well known how the reliability of channel state information (CSI) can be crucial for the convergence of turbo equalization [4]. In block transmission systems the CSI is usually obtained block-by-block from the training symbols included in each block (*training-based single-block* or SB estimate). In this paper we propose to improve the CSI accuracy by means of a *soft-based multi-block* (MB) processing. The method is developed for a generic wireless communication system under the assumption of block-fading channel (the fading is constant within each block, but it varies from block to block due to the terminal mobility).

In the literature the use of soft information for channel estimation has been largely investigated to improve the performance of iterative receivers [3]-[5]. The basic idea is to repeat channel estimation at each iteration by exploiting the soft information fed back by the channel decoder. In this paper we derive a soft-iterative version of the *training-based MB* estimator [1]. Since turbo processing is usually performed on a set of $L > 1$ data blocks (L depending on the interleaver size), we propose to take advantage of this inherent latency to improve the estimate accuracy for the slowly varying channel parameters. The MB approach relies on the assumption that the multipath delays remain constant within the L blocks, while the fading amplitudes vary from block to block. The subspace spanned by the channel responses over the different paths (here referred to as the *channel subspace*) can be estimated by sample averaging from the signals received over the L blocks, while the fast varying parameters need to be calculated block-by-block. In this paper we propose a *soft-based MB* approach where the initial estimate is obtained from the pilot symbols as in [1] and it is then refined in the subsequent iterations by extending the training set with soft-valued information symbols. With respect to [1], here the use of the soft information allows to improve the accuracy for both the channel subspace and the fading amplitude estimates. In this paper the estimate is proposed for a single-input-single-output system, but the same method can be also applied to single-input-multiple-output (SIMO) or multiple-input-multiple-output (MIMO) systems [6].

The paper is organized as follows. Sec. 2 presents the signal model for a block-based transmission system and the receiver structure. Soft MB estimation is in Sec. 3, the analytical evaluation of the MSE is in Sec. 4. Sec. 5 shows by simulations the advantage of the proposed method and Sec. 6 gives the concluding remarks.

2. SYSTEM DESCRIPTION

2.1 Signal model

We consider the equivalent complex baseband model for the convolutionally coded system in Fig. 1. A sequence $\{d(i)\}$ of binary information symbols, $d(i) \in \mathcal{D} = \{+1, -1\}$, is convolutionally encoded with code rate R. The output code bits $\{c(i)\}$ are permuted by a random interleaver $\Pi[\cdot]$, $b(i) = c(\Pi[i])$, and mapped into quadrature phase-shift-keying (QPSK) symbols $x_\mathrm{d}(i) = (b(2i) + jb(2i+1))/\sqrt{2}$ of duration T_s (the analysis can be easily extended to larger constellations). After mapping, the overall sequence $\{x_\mathrm{d}(i)\}$ is split into L blocks of $N_\mathrm{d}^{'}$ symbols each: $\mathbf{x}_\mathrm{d}(\ell) = [x_\mathrm{d}(0;\ell), \ldots, x_\mathrm{d}(N_\mathrm{d}^{'} - 1;\ell)]$, with $x_\mathrm{d}(i;\ell) = x_\mathrm{d}((\ell-1)N_\mathrm{d} + i)$, for $\ell = 1, \ldots, L$. In order to allow channel estimation at the receiver, an uncoded training sequence $\mathbf{x}_\mathrm{t}(\ell) = [x_\mathrm{t}(0;\ell), \ldots, x_\mathrm{t}(N_\mathrm{t}^{'} - 1;\ell)]$ is added as preamble within each block, yielding the overall sequence $\mathbf{x}(\ell) = [\mathbf{x}_\mathrm{t}(\ell), \mathbf{x}_\mathrm{d}(\ell)] = [x(0;\ell), \ldots, x(N^{'} - 1;\ell)]$ of length $N' = N_\mathrm{t}^{'} + N_\mathrm{d}^{'}$. The L blocks are then transmitted over a block-faded frequency-selective channel.

At the receiver, after matched filtering and sampling at the symbol rate, the signal measured within the ℓth block is

$$y(i;\ell) = \mathbf{x}^\mathrm{T}(i;\ell)\mathbf{h}(\ell) + w(i;\ell), \qquad i = 0, \ldots, N^{'} + W - 2 \quad (1)$$

where $\mathbf{x}(i;\ell) = [x(i;\ell), \ldots, x(i-W+1;\ell)]^\mathrm{T} \in \mathbb{C}^{W\times 1}$ collects W (either training or information) symbols, the complex Gaussian noise $w(i;\ell)$ is white, with zero mean and known variance σ_w^2: $\mathrm{E}[w^*(i;\ell)w(i+k;\ell+m)] = \sigma_w^2 \delta_k \delta_m$ ($\delta_k = 1$ for $k = 0$, $\delta_k = 0$ elsewhere). The channel is 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.

2.2 Iterative receiver structure

The iterative receiver structure is shown in Fig. 2. It consists of a soft-in channel estimator, a sliding window soft-input-soft-output (SISO) minimum mean square error (MMSE) linear equalizer [7][8] and a log maximum a-posteriori (log-MAP) SISO decoder [9] separated by interleaver/de-interleaver.

At each iteration the soft channel estimator derives (as described later in Sect. 3) 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) $\lambda_1[b(i)] = \log[P[b(i) = +1]/P[b(i) = -1]]$ for the data-bearing bits $b(i)$. The channel estimates and the a-priori LLR $\lambda_1[b(i)]$ are used by the SISO equalizer [8] to compute the MMSE estimate $\hat{b}(i)$ and the corresponding extrinsic LLR $\lambda_1^\mathrm{E}[b(i)] = \log[P[\hat{b}(i)|b(i) = +1]/P[\hat{b}(i)|b(i) = -1]]$ (calculated under the Gaussian approximation), with $\lambda_1^\mathrm{E}[b(i)] = \Lambda_1[b(i)] - \lambda_1[b(i)]$, and $\Lambda_1[b(i)]$ denoting the a-posteriori LLR. After equalization, the soft information $\lambda_1^\mathrm{E}[b(i)]$ is reversed interleaved and it is passed to the decoder as a-priori LLR $\lambda_2[c(i)] = \log[P[c(i) = +1]/P[c(i) = -1]]$, for each code bit $c(i) = b(\Pi^{-1}[i])$. The decoder [9] computes the a-posteriori LLR $\Lambda_2[c(i)]$ and it delivers the new extrinsic LLR $\lambda_2^\mathrm{E}[c(i)] = \Lambda_2[c(i)] - \lambda_2[c(i)]$. The refined soft information is interleaved again and used as new a-priori LLR $\lambda_1[b(i)]$ for further iterations. At the last iteration, the a-posteriori LLR for the information bit $d(i)$ is computed as well to provide the final estimate $\hat{d}(i)$.

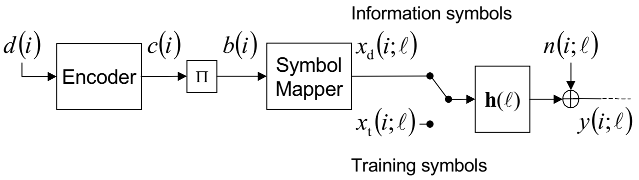

Figure 1: Transmitter structure.

In the following we focus on channel estimation. For further details on the equalization and the decoding tasks in addition to the review above we refer to the cited papers.

2.3 Subspace channel model

Within the L blocks the channel is modelled according to the multipath model $\mathbf{h}(\ell) = \mathbf{G}(\boldsymbol{\tau})\boldsymbol{\alpha}(\ell)$, superposition of d paths having constant delays $\boldsymbol{\tau} = [\tau_1, ..., \tau_d]$ and block-fading amplitudes $\boldsymbol{\alpha}(\ell) = [\alpha_1(\ell), ..., \alpha_d(\ell)]$. The kth column, $\mathbf{g}(\tau_k)$, of the matrix $\mathbf{G}(\boldsymbol{\tau}) \in \mathbb{R}^{W\times d}$ 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 $\mathbf{R}_\alpha = \mathrm{diag}\{A_1, ..., A_d\}$ describing the power-delay-profile. 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 $r = \mathrm{rank}[\mathbf{G}(\boldsymbol{\tau})] = \mathrm{rank}[\mathbf{R}_h] \leq W$. The r-dimensional subspace $\mathcal{R}(\mathbf{G}(\boldsymbol{\tau})) = \mathcal{R}(\mathbf{R}_h)$, defined by the multipath components $\{\mathbf{g}(\tau_k)\}_{k=1}^d$, will be referred to as the channel subspace. Its dimension r represents the number of resolvable delays for the bandwidth of the transmitted signal.

Based on the assumptions above, the channel vector can be rewritten in terms of the new parameters [1]

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

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. Notice that the parameterization (2) is not unique; for instance, we can select $\mathbf{U}$ as the matrix containing the r eigenvectors of $\mathbf{R}_h$ (*stationary channel modes*), the corresponding amplitudes (*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 $\mathbf{R}_h$.

3. CHANNEL ESTIMATION

For channel estimation it is convenient to rewrite the model (1) as

$$\begin{cases} \mathbf{y}_{\mathrm{t}}(\ell) = \mathbf{X}_{\mathrm{t}}(\ell)\mathbf{h}(\ell) + \mathbf{w}_{\mathrm{t}}(\ell), & \text{Training} \\ \mathbf{y}_{\mathrm{d}}(\ell) = \mathbf{X}_{\mathrm{d}}(\ell)\mathbf{h}(\ell) + \mathbf{w}_{\mathrm{d}}(\ell), & \text{Data} \end{cases} \tag{3}$$

where $\mathbf{y}_{\mathrm{t}}(\ell) = [y(W;\ell), \ldots, y(N_{\mathrm{t}}' - 1;\ell)]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{t}}\times 1}$ and $\mathbf{y}_{\mathrm{d}}(\ell) = [y(N_{\mathrm{t}}' + W - 1;\ell), \ldots, y(N' - 1;\ell)]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{d}}\times 1}$ 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 of the ℓth block (the first $W-1$ samples at the beginning of each phase are discarded to avoid the overlapping between training and data). Accordingly, $\mathbf{X}_{\mathrm{t}}(\ell) \in \mathbb{C}^{N_{\mathrm{t}}\times W}$ and $\mathbf{X}_{\mathrm{d}}(\ell) \in \mathbb{C}^{N_{\mathrm{d}}\times W}$ are Toeplitz matrices collecting the training and the data symbols, $\mathbf{R}_{\mathrm{t}} = \mathbf{X}_{\mathrm{t}}^{\mathrm{H}}(\ell)\mathbf{X}_{\mathrm{t}}(\ell)$ and $\mathbf{R}_{\mathrm{d}} = \mathbf{X}_{\mathrm{d}}^{\mathrm{H}}(\ell)\mathbf{X}_{\mathrm{d}}(\ell)$ are the corresponding correlation matrices (both assumed to be independent of the block index), the vectors $\mathbf{w}_{\mathrm{t}}(\ell) \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2\mathbf{I}_{N_{\mathrm{t}}})$ and $\mathbf{w}_{\mathrm{d}}(\ell) \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2\mathbf{I}_{N_{\mathrm{d}}})$ contain the noise samples.

In the following we address the problem of ML estimation of the channel vector $\mathbf{h}(\ell)$ from the ensemble of L blocks $\{\mathbf{y}_{\mathrm{t}}(\ell), \mathbf{y}_{\mathrm{d}}(\ell)\}_{\ell=1}^L$ under the constraint (2) and for known rank order r. At the first iteration, as no a-priori information is available on the information-bearing data $\mathbf{X}_{\mathrm{d}}(\ell)$ (the a-priori LLRs are $\lambda_1[b(i)] = 0$ for all code bits $b(i)$), channel estimation is carried out from the training signals $\{\mathbf{y}_{\mathrm{t}}(\ell)\}_{\ell=1}^L$ only, using the knowledge of the pilot symbols $\mathbf{X}_{\mathrm{t}}$ (*training-based channel estimation*). After the first channel estimation, equalization and decoding of the L blocks, the a-priori LLRs $\{\lambda_1[b(i)]\}$ can be exploited to refine the initial estimate (*soft channel estimation*). 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 every code symbol $x_{\mathrm{d}}(i)$, $i = 0, \ldots, LN_{\mathrm{d}}' - 1$. We recall that for QPSK modulation these statistics can be easily obtained as [8]

$$\begin{aligned} \bar{x}_{\mathrm{d}}(i) &= \frac{1}{\sqrt{2}}\left(\tanh\frac{\lambda_1[b(2i)]}{2} + j\tanh\frac{\lambda_1[b(2i+1)]}{2}\right) && (4) \\ \sigma_{\mathrm{d}}^2(i) &= 1 - |\bar{x}_{\mathrm{d}}(i)|^2 && (5) \end{aligned}$$

(within the ℓth block, the quantities (4) and (5) 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 approximations are needed to perform ML channel estimation. We first assume that the information-bearing symbols $\{x_{\mathrm{d}}(i;\ell)\}$ are independent and N_{d} is large enough so that $\mathbf{R}_{\mathrm{d}} \approx N_{\mathrm{d}}\mathbf{I}_W$ and $\bar{\mathbf{R}}_{\mathrm{d}} = \bar{\mathbf{X}}_{\mathrm{d}}^{\mathrm{H}}(\ell)\bar{\mathbf{X}}_{\mathrm{d}}(\ell) \approx \mathbf{I}_W\tilde{N}_{\mathrm{d}}$ (both matrices are constant over the blocks). The parameter $\tilde{N}_{\mathrm{d}} = N_{\mathrm{d}}(1 - \sigma_{\mathrm{d}}^2)$ depends on the average variance σ_{d}^2 of the information-bearing symbols:

$$\sigma_{\mathrm{d}}^2 = \frac{1}{LN_{\mathrm{d}}'}\sum_{i=1}^{LN_{\mathrm{d}}'}\left(1 - |\bar{x}_{\mathrm{d}}(i)|^2\right) = \frac{1}{LN_{\mathrm{d}}'}\sum_{i=1}^{LN_{\mathrm{d}}'}\sigma_{\mathrm{d}}^2(i). \tag{6}$$

We further assume $\{\Delta x_{\mathrm{d}}(i;\ell)\}$ as a stationary white process with variance σ_{d}^2, independent from the noise $\{w(i;\ell)\}$. It follows that $\mathrm{E}[\Delta\mathbf{X}_{\mathrm{d}}^{\mathrm{H}}(\ell)\Delta\mathbf{X}_{\mathrm{d}}(\ell)] = N_{\mathrm{d}}\sigma_{\mathrm{d}}^2\mathbf{I}_W$. It is worth noticing that $\tilde{N}_{\mathrm{d}}$ represents the *effective* number of known data symbols that can be used in each block for channel estimation. It is indeed $0 \leq \tilde{N}_{\mathrm{d}} \leq N_{\mathrm{d}}$, with $\tilde{N}_{\mathrm{d}} = 0$ for missing prior information ($\lambda_1[b(i)] = 0$) and $\tilde{N}_{\mathrm{d}} = N_{\mathrm{d}}$ for perfect prior information ($\lambda_1[b(i)] = \pm\infty$).

Based on the assumptions above, the model (3) reduces to

$$\begin{cases} \mathbf{y}_{\mathrm{t}}(\ell) = \mathbf{X}_{\mathrm{t}}(\ell)\mathbf{h}(\ell) + \mathbf{w}_{\mathrm{t}}(\ell), & \text{Training} \\ \mathbf{y}_{\mathrm{d}}(\ell) = \bar{\mathbf{X}}_{\mathrm{d}}(\ell)\mathbf{h}(\ell) + \Delta\mathbf{w}_{\mathrm{d}}(\ell) + \mathbf{w}_{\mathrm{d}}(\ell), & \text{Data} \end{cases} \tag{7}$$

where the soft-valued data $\bar{\mathbf{X}}_{\mathrm{d}}(\ell)$ are 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 $\mathbf{w}_{\mathrm{d}}(\ell)$ and having variance $\Delta\sigma_w^2 = \sigma_{\mathrm{d}}^2\mathrm{E}[||\mathbf{h}(\ell)||^2]$.

In the following channel estimation will be performed from (3) under the white Gaussian assumption $\Delta\mathbf{w}_{\mathrm{d}}(\ell) \sim$

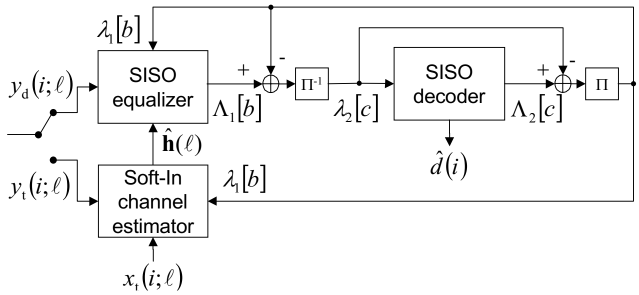

Figure 2: Receiver structure.

$\mathcal{CN}(\mathbf{0},\Delta\sigma_w^2\mathbf{I}_{N_\mathrm{d}})$ (whiteness holds for diagonal $\mathbf{R}_h$, e.g. for symbol-spaced delays and Nyquist impulse waveform). Notice however that the new signal model is not homogeneous, as, due to the unreliability of the soft-valued training data $\bar{x}_\mathrm{d}(i;\ell)$, the input noise variance σ_w^2 is increased by $\Delta\sigma_w^2$ in the signal $\mathbf{y}_\mathrm{d}(\ell)$.

3.1 Training-based channel estimation

The constrained ML estimate of the channel $\mathbf{h}(\ell)$ from the signals $\{\mathbf{y}_\mathrm{t}(\ell)\}_{\ell=1}^L$ and for known $\{\mathbf{X}_\mathrm{t}(\ell)\}_{\ell=1}^L$ is obtained by minimizing the negative log-likelihood function $\mathcal{L}_\mathrm{t}=\sum_{\ell=1}^L||\mathbf{y}_\mathrm{t}(\ell)-\mathbf{X}_\mathrm{t}(\ell)\mathbf{h}(\ell)||^2$ under the constraint (2). The optimization yields the training-based MB estimator [1] that requires the preliminary evaluation of L SB estimates obtained by performing an unconstrained ML estimation within each block.

3.2 Soft-based channel estimation

The constrained ML estimate of $\mathbf{h}(\ell)$ from the signals $\{\mathbf{y}_\mathrm{t}(\ell),\mathbf{y}_\mathrm{d}(\ell)\}_{\ell=1}^L$ and for known $\{\mathbf{X}_\mathrm{t}(\ell),\bar{\mathbf{X}}_\mathrm{d}(\ell)\}_{\ell=1}^L$ is obtained by minimizing $\mathcal{L}_\mathrm{s}=\mathcal{L}_\mathrm{t}+\gamma\sum_{\ell=1}^L||\mathbf{y}_\mathrm{d}(\ell)-\bar{\mathbf{X}}_\mathrm{d}(\ell)\mathbf{h}(\ell)||^2$ under the constraint (2) and for $\gamma=(1+\Delta\sigma_w^2/\sigma_w^2)^{-1}$. Similarly to [1], it can be shown that the minimizer is the soft MB estimate

$$\hat{\mathbf{h}}_\mathrm{MB}(\ell)=\bar{\mathbf{R}}^{-1/2}\hat{\mathbf{P}}\bar{\mathbf{R}}^{1/2}\hat{\mathbf{h}}_\mathrm{SB}(\ell) \quad (8)$$

that is based on the (soft SB) unconstrained estimate

$$\hat{\mathbf{h}}_\mathrm{SB}(\ell)=\bar{\mathbf{R}}^{-1}\left(\mathbf{X}_\mathrm{t}^\mathrm{H}(\ell)\mathbf{y}_\mathrm{t}(\ell)+\gamma\bar{\mathbf{X}}_\mathrm{d}^\mathrm{H}(\ell)\mathbf{y}_\mathrm{d}(\ell)\right) \quad (9)$$

where we set $\bar{\mathbf{R}}=\mathbf{R}_\mathrm{t}+\gamma\bar{\mathbf{R}}_\mathrm{d}$. The estimate $\hat{\mathbf{P}}$ for the projector onto the channel subspace is obtained from the r leading eigenvectors of the sample correlation matrix

$$\mathbf{R}_\mathrm{MB}(L)=\frac{1}{L}\bar{\mathbf{R}}^{1/2}\left(\sum_{\ell=1}^L\hat{\mathbf{h}}_\mathrm{SB}(\ell)\hat{\mathbf{h}}_\mathrm{SB}^\mathrm{H}(\ell)\right)\bar{\mathbf{R}}^{\mathrm{H}/2}. \quad (10)$$

Remark 1. Notice that if the data symbol estimates provided by the decoder are unreliable (i.e., at the first iterations of the iterative processing for moderate SNR), it is $\sigma_\mathrm{d}^2\approx 1$, $\tilde{N}_\mathrm{d}\approx 0$, $\bar{\mathbf{X}}_\mathrm{d}(\ell)\approx 0$, and the soft MB estimate (8) coincides with the training-based one [1]. On the other hand, for perfect a-priori information (i.e., after a large enough number of iterations, provided that the iterative approach converges) it is $\sigma_\mathrm{d}^2\approx 0$, $\tilde{N}_\mathrm{d}\approx N_\mathrm{d}$, $\bar{\mathbf{X}}_\mathrm{d}(\ell)=\mathbf{X}_\mathrm{d}(\ell)$ and therefore the soft estimate equals the training-based estimate that would be obtained from a training sequence of $N_\mathrm{t}+N_\mathrm{d}$ symbols.

Remark 2. The MB method is based on the soft SB estimate (9), which is suboptimal as it is derived under the Gaussian assumption for $\Delta\mathbf{w}_\mathrm{d}(\ell)$. Nevertheless, this approach has some definite advantages with respect to other channel estimation techniques combining training and soft-valued data. For instance, consider the "local" EM estimation (mixing method [4], also equivalent to [3]) applied to the incomplete data $\{\mathbf{y}_\mathrm{t}(\ell),\mathbf{y}_\mathrm{d}(\ell)\}$ with missing data $\mathbf{X}_\mathrm{d}(\ell)$ and known parameter $\mathbf{X}_\mathrm{t}$. The estimate is the minimizer of $\mathcal{L}_1=\mathcal{L}_\mathrm{t}+\mathrm{E}_{x_\mathrm{d}}[||\mathbf{y}_\mathrm{d}(\ell)-\mathbf{X}_\mathrm{d}(\ell)\mathbf{h}(\ell)||^2]$ and it can be obtained from (9) by setting $\gamma=1$ and replacing $\bar{\mathbf{R}}_\mathrm{d}$ with $\mathrm{E}[\mathbf{R}_\mathrm{d}]$. As highlighted in [4], for $N_\mathrm{d}\gg N_\mathrm{t}$ and unreliable soft information ($\bar{\mathbf{X}}_\mathrm{d}(\ell)\approx 0$) this estimate suffers from an evident bias that might prevent the iterative receiver to bootstrap. This is not the case of the method herein proposed, that provides always an unbiased estimate. A similar unbiased estimate is proposed in [5] as the minimizer of $\mathcal{L}_\mathrm{s}$ for $\gamma=1$. The solution is obtained from (9) by setting $\gamma=1$ (the variance of the information-bearing symbols is not taken into account). It can be shown that the performance of the two estimates are similar for $\Delta\sigma_w^2\ll\sigma_w^2$ ($\gamma\approx 1$), but when $\Delta\sigma_w^2$ and σ_w^2 are comparable (i.e., for large SNR and unreliable prior information) the soft estimate [5] performs worse than the conventional training-based SB estimate. This never occurs with the method (9), as it will be shown analytically in Sec. 4 and by simulation results in Sec. 5.

4. PERFORMANCE ANALYSIS

Let us assume $L\to\infty$ (i.e., perfect knowledge of the temporal subspace), the estimate error $\Delta\mathbf{h}(\ell)=\hat{\mathbf{h}}(\ell)-\mathbf{h}(\ell)$ for the SB ($\Delta\mathbf{h}_\mathrm{SB}(\ell)$) and MB ($\Delta\mathbf{h}_\mathrm{MB}(\ell)$) methods can be written as

$$\begin{aligned}\Delta\mathbf{h}_\mathrm{SB}(\ell) &= \bar{\mathbf{R}}^{-1}\{\mathbf{X}_\mathrm{t}^\mathrm{H}(\ell)\mathbf{w}_\mathrm{t}(\ell)+\gamma\bar{\mathbf{X}}_\mathrm{d}^\mathrm{H}(\ell)[\Delta\mathbf{w}_\mathrm{d}(\ell)+\mathbf{w}_\mathrm{d}(\ell)]\}\\ \Delta\mathbf{h}_\mathrm{MB}(\ell) &= \bar{\mathbf{R}}^{-1/2}\mathbf{P}\bar{\mathbf{R}}^{1/2}\Delta\mathbf{h}_\mathrm{SB}(\ell)\end{aligned}$$

where $\mathbf{P}$ is the true projector onto the temporal subspace $\mathcal{R}[\bar{\mathbf{R}}^{1/2}\mathbf{G}]$. Recalling that $\mathbf{w}_\mathrm{t}(\ell)$, $\Delta\mathbf{w}_\mathrm{d}(\ell)$ and $\mathbf{w}_\mathrm{d}(\ell)$ are uncorrelated, it can be shown that the covariance $\mathrm{Cov}(\hat{\mathbf{h}})=\mathrm{E}[\Delta\mathbf{h}(\ell)\Delta\mathbf{h}^\mathrm{H}(\ell)]$ (where averaging is performed over fading and noise) is

$$\mathrm{Cov}(\hat{\mathbf{h}}_\mathrm{SB}) = \sigma_w^2\bar{\mathbf{R}}^{-1} \quad (11)$$

$$\mathrm{Cov}(\hat{\mathbf{h}}_\mathrm{MB}) = \sigma_w^2\bar{\mathbf{R}}^{-1/2}\mathbf{P}\bar{\mathbf{R}}^{-\mathrm{H}/2}. \quad (12)$$

The MSE is obtained as MSE=$\mathrm{tr}(\mathrm{Cov}(\hat{\mathbf{h}}))$ from (11)-(12) yielding the results in Table 1. The following relationships hold between then MSE of the training-based estimate (superscript t), the soft-based estimate and the soft-based estimate for $\sigma_\mathrm{d}^2=0$ (superscript t+d): $\mathrm{MSE}_\mathrm{SB}^{(\mathrm{t+d})}\leq\mathrm{MSE}_\mathrm{SB}\leq\mathrm{MSE}_\mathrm{SB}^{(\mathrm{t})}$, $\mathrm{MSE}_\mathrm{MB}^{(\mathrm{t+d})}\leq\mathrm{MSE}_\mathrm{MB}\leq\mathrm{MSE}_\mathrm{MB}^{(\mathrm{t})}$. This can be proved by observing that $\mathbf{R}_\mathrm{t}\leq\bar{\mathbf{R}}\leq\mathbf{R}_\mathrm{t}+\mathbf{R}_\mathrm{d}$.

The MSE expressions simplify for uncorrelated training sequences, i.e. for $\mathbf{R}_\mathrm{t}=N_\mathrm{t}\mathbf{I}_W$ and thus $\bar{\mathbf{R}}=(N_\mathrm{t}+\gamma\tilde{N}_\mathrm{d})\mathbf{I}_W$, as shown in the third column of Table 1. As expected, in this case the performance 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 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 $\mathbf{b}(\ell)$ [1], as the projector $\mathbf{P}$ (as well as its basis $\mathbf{U}$) is perfectly estimated for $L\to\infty$. Table 1 also shows the performance for two extreme conditions: missing prior information (i.e., at the first iteration for $\sigma_\mathrm{d}^2=1$); perfect prior information (i.e., close to the convergence of the iterative approach for $\sigma_\mathrm{d}^2=0$).

5. SIMULATION RESULTS

The performance for the SB and MB methods are compared by simulating the following system. A frame of 4000 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 4000 QPSK symbols and arranged into $L=20$ blocks with $N_\mathrm{d}'=200$ symbols each. A training sequence of $N_\mathrm{t}=31$ QPSK symbols is added in each block (to avoid border effects a cyclic prefix of $W-1$ symbols is used yielding $N_\mathrm{t}'=46$). The L blocks are then transmitted over a block-faded Rayleigh channel having $r=6$ resolvable paths, $\mathbf{R}_\alpha=\mathrm{diag}\left\{1,\frac{1}{2},\frac{1}{4},1,\frac{1}{2},\frac{1}{4}\right\}$ and $\boldsymbol{\tau}=[0,1.5,2.8,10.5,11.8,13]$

Table 1: MSE of soft iterative SB and MB estimates.

Estimate	Correlated	Unc.
from training + soft-valued data ($\sigma_\mathrm{d}^2\in[0,1]$)		
SB	$\sigma_w^2\,\mathrm{tr}[\bar{\mathbf{R}}^{-1}]$	$\sigma_w^2\frac{W}{N_\mathrm{t}+\gamma\tilde{N}_\mathrm{d}}$
MB	$\sigma_w^2\,\mathrm{tr}[(\mathbf{R}_\mathrm{t}+\gamma\bar{\mathbf{R}}_\mathrm{d})^{-1/2}\mathbf{P}(\mathbf{R}_\mathrm{t}+\gamma\bar{\mathbf{R}}_\mathrm{d})^{-\mathrm{H}/2}]$	$\sigma_w^2\frac{r}{N_\mathrm{t}+\gamma\tilde{N}_\mathrm{d}}$
from training only ($\sigma_\mathrm{d}^2=1$)		
SB	$\sigma_w^2\,\mathrm{tr}[\mathbf{R}_\mathrm{t}^{-1}]$	$\sigma_w^2\frac{W}{N_\mathrm{t}}$
MB	$\sigma_w^2\,\mathrm{tr}[\mathbf{R}_\mathrm{t}^{-1/2}\mathbf{P}\mathbf{R}_\mathrm{t}^{-\mathrm{H}/2}]$	$\sigma_w^2\frac{r}{N_\mathrm{t}}$
from training + data ($\sigma_\mathrm{d}^2=0$)		
SB	$\sigma_w^2\,\mathrm{tr}[(\mathbf{R}_\mathrm{t}+\mathbf{R}_\mathrm{d})^{-1}]$	$\sigma_w^2\frac{W}{N_\mathrm{t}+N_\mathrm{d}}$
MB	$\sigma_w^2\,\mathrm{tr}[(\mathbf{R}_\mathrm{t}+\mathbf{R}_\mathrm{d})^{-1/2}\mathbf{P}(\mathbf{R}_\mathrm{t}+\mathbf{R}_\mathrm{d})^{-\mathrm{H}/2}]$	$\sigma_w^2\frac{r}{N_\mathrm{t}+N_\mathrm{d}}$

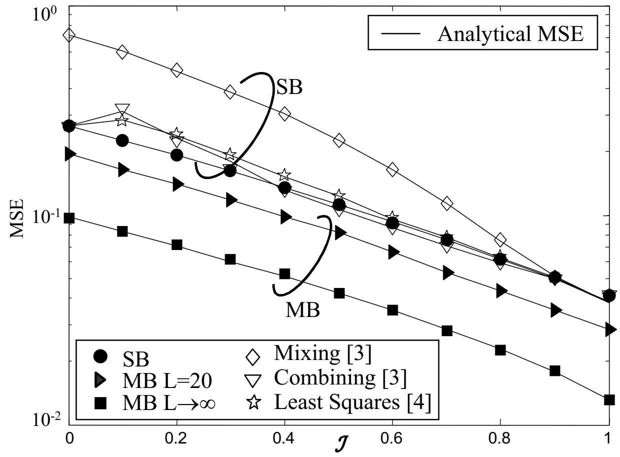

Figure 3: MSE for SB-MB soft estimate vs. mutual information.

[samples]. For the MB method the projector $\mathbf{P}$ onto the temporal subspace is estimated from either $L = 20$ or $L \to \infty$ blocks.

Fig. 3 compares the MSE of the soft SB and MB estimates for varying mutual information $\mathcal{I} = \mathcal{I}\{b(i), \lambda_1[b(i)]\}$ between the code bits $b(i)$ and the a-priori information $\lambda_1[b(i)]$. According to [10], the a-priori LLR $\lambda_1[b(i)]$ is modelled as Gaussian and $E_b/N_0 = 3$dB. The simulated MSE values (markers) are compared with the analytical results (continuous lines) of Table 1. It can be seen that all soft-based methods become more accurate for increasing $\mathcal{I}$ (or, equivalently, for decreasing σ_d^2), from $\mathcal{I} = 0$ (or $\sigma_d^2 = 1$, when only training symbols are used) to $\mathcal{I} = 1$ (or $\sigma_d^2 = 0$, when the whole block of N known symbols is used). The SB ML method proposed in this paper is also compared with the other soft-based estimators: mixing method [4], combining method [4], LS method [5]. All SB methods reach the same accuracy at $\mathcal{I} = 1$, while for moderate $\mathcal{I}$ the SB ML estimator outperforms all other methods. The effect of the EM estimate bias is evident for small $\mathcal{I}$. The soft MB estimate outperforms all method reaching a gain (from Table 1) equal to $\text{MSE}_{\text{SB}}/\text{MSE}_{\text{MB}} \approx W/r = 4.26$dB with respect to the SB ML method (here it is $\mathbf{R}_t \approx N_t \mathbf{I}_W$) and to $\text{MSE}_{\text{MB}}^{(t)}/\text{MSE}_{\text{MB}} \approx N/N_t = 7.8$dB (for $\mathcal{I} = 1$) with respect to the training-based MB approach.

Fig. 4 shows the BER performance for the complete iterative receiver. Fig. 4-a compares the receiver with MB soft channel estimation (for known projector $\mathbf{P}$ or $L \to \infty$) with the case of known channel. $n = 5$ iterations are enough for the MB iterative channel estimator to approach the performance of known channel. Fig. 4-b shows the performance of the receiver with SB and MB estimation after 5 iterations. Both training-based and soft-iterative approaches are used. We observe that the soft MB method outperforms both the training-based and the soft SB methods. Its performance at the 5th iteration is close to that obtained for known channel.

6. CONCLUDING REMARKS

This paper proposes the integration of MB channel estimation for block-fading channels with soft iterative equalization. The MB method exploits the invariance of the temporal subspace across blocks and it estimates the channel using the soft statistics fed back by the decoder. The analytical evaluation of the MSE for the channel estimate and the simulation results on the BER for the complete iterative receiver show the benefits of the proposed method.

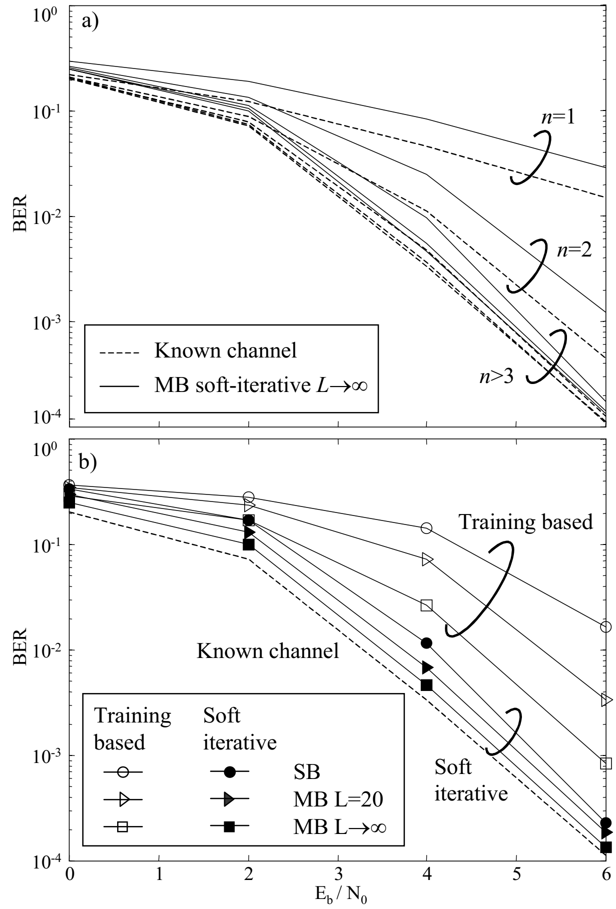

Figure 4: BER performance vs. E_b/N_0 for varying number n of iterations (top) and at the 5th iteration (bottom).

REFERENCES

[1] 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.

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

[3] M. Sandell, C. Luschi, P. Strauch, R. Yan, “Iterative channel estimation using soft decision feedback”, *Proc. IEEE GLOBECOM 1998*, Vol. 6, pp-3728 - 3733. Nov. 1998.

[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] 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., 2005.

[7] X. Wang, V. Poor, “Iterative (turbo) soft interference cancellation and decoding for coded CDMA,” *IEEE Trans. Commun.*, vol. 47, pp. 1046-1061, July 1999.

[8] M. Tüchler, A. C. Singer, “Minimum mean squared error equalization using *a priori* information,” *IEEE Trans. Signal Processing*, vol. 50, pp. 673-683, Mar. 2002.

[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.