# A PERFORMANCE OF BAYESIAN SEMI-BLIND FIR CHANNEL ESTIMATION ALGORITHMS IN SIMO SYSTEMS


*Samir-Mohamad Omar*†, *Dirk T.M. Slock*† *and Oussama Bazzi*‡

†EURECOM, Dept. of Mobile Communications,
2229 Route des Crêtes,
06904 Sophia Antipolis Cedex, France
email: omar@eurecom.fr, dirk.slock@eurecom.fr

‡Department of Physics and Electronics,
Faculty of Sciences I, Lebanese University
Hadath ,Beirut, Lebanon
email: obazzi@ul.edu.lb



## ABSTRACT

When the transmission scenario includes a training sequence or pilots, semi-blind channel estimation techniques have shown a tendency to fully exploit the information available from the received signal if they are correctly implemented. This feature leads semi-blind channel estimation performance to exceed that of the schemes based on the blind part or the training sequence only. Moreover, in some situations they can estimate the channel when the other techniques fail. Semi-blind channel estimation techniques were developed and usually evaluated for a given channel realization, i.e. with a deterministic channel model. On the other hand, in wireless communications the channel is typically modeled as Rayleigh fading, i.e. with a Gaussian (prior) distribution expressing variances of and correlations between channel coefficients. In recent years, such prior information on the channel has started to get exploited in pilot-based channel estimation, since often the pure pilot-based (deterministic) channel estimate is of limited quality due to limited pilots. In this paper we provide a performance comparison between ML/MAP algorithms that use Bayesian and deterministic approaches in semi-blind channel estimation.


## 1. INTRODUCTION

Traditionally, the transmitter sends some known information to the receiver to aid the latter in estimating the channel. However, in wireless communication the channel varies rapidly with time and as a consequence, more training sequence/pilots are required. This process wastes a lot of bandwidth as a result of augmenting the transmission rate to maintain the throughput. In the last two decades a new branch of channel estimation has emerged focusing on accomplishing this task blindly i.e. without the need for a training sequence. Nevertheless, most wireless standards that have evolved during that period are still relying on the training sequence/pilots to estimate the channel. This is due probably to the unsatisfactory results of the blind channel estimation algorithms. On the other hand, there are some powerful channel estimation algorithms that take advantage of both aforementioned techniques have been also developed during the same era. These are known as semi-blind techniques where a superior performance is achieved although few training sequence/pilots are transmitted [1], [2], [3].

In [4] and [5] we introduced some Bayesian (semi-)blind channel estimation algorithms that exploit perfectly the knowledge of the channel a priori information to enhance the channel estimation quality. In this paper, we are exploring an approach that exploits partially the knowledge of the Power Delay Profile (PDP) to enhance the estimation of a part of the channel while neglecting totally the remaining part. It is worth noting that this approach is not restricted to Bayesian algorithms but can rather be implemented to any existing deterministic algorithm. By doing so, we are extending those deterministic algorithms to a point in the middle between deterministic and Bayesian, hence we can classify them as quasi-Bayesian algorithms. The question that may raise here, is there still a room to enhance the estimation quality of the Bayesian algorithms? Moreover, sometimes the estimation of the channel is required by itself, to be used in the beamforming for instance, while in some other cases it constitutes only one step toward another ultimate goal, the detection of symbols. Hence, one may wonder what are the consequences of neglecting a part of the channel on the detection of the symbols. In the following sections, we will try to elaborate our approach and answer these questions.

## 2. SIMO FIR TX SYSTEM MODEL

In (semi-)blind channel identification, a multichannel framework can be obtained from oversampling a received signal and leads to a Single Input Multiple Output (SIMO) vector channel representation. The multiple FIR channels we obtain in this representation can also be obtained from multiple received signals from an array of antennas (in the context of mobile digital communications [6]) or from a combination of both. To further develop the case of oversampling, consider a linear digital modulation over a linear channel with additive noise so that the received signal $y(t)$ has the following form:

$$y(t) \;=\; \sum_k h(t-kT)a(k) \;+\; v(t). \tag{1}$$

In (1) $a(k)$ are the transmitted symbols, $T$ is the symbol period, $h(t)$ is the channel impulse response and $v(t)$ designates noise. The channel is assumed to be FIR with length $NT$. If the received signal is oversampled at the rate $\frac{m}{T}$ (or if $m$ different samples of the received signal are captured by $m$ sensors every $T$ seconds, or a combination of both), the discrete input-output relationship can be written as:

$$\boldsymbol{y}(k) = \sum_{i=0}^{N-1} \boldsymbol{h}(i)a(k-i) + \boldsymbol{v}(k) = \boldsymbol{H}A_N(k) + \boldsymbol{v}(k) \tag{2}$$

where $\boldsymbol{y}(k) = [y_1^H(k)\cdots y_m^H(k)]^H, \boldsymbol{h}(i) = \left[h_1^H(i)\cdots h_m^H(i)\right]^H$, $\boldsymbol{v}(k) = [v_1^H(k)\cdots v_m^H(k)]^H$, $\boldsymbol{H} = [\boldsymbol{h}(N{-}1)\cdots \boldsymbol{h}(0)]$, $A_N(k) = \left[a(k{-}N{+}1)^H\cdots a(k)^H\right]^H$ and superscript $^H$ denotes Hermitian transpose. Let $\mathbf{H}(z) = \sum_{i=0}^{N-1}\boldsymbol{h}(i)z^{-i} = [\mathrm{H}_1^H(z)\cdots \mathrm{H}_m^H(z)]^H$ be the SIMO channel transfer function, and $\boldsymbol{h} = \left[\boldsymbol{h}^H(N{-}1)\cdots \boldsymbol{h}^H(0)\right]^H$. Consider additive independent white Gaussian circular noise $\boldsymbol{v}(k)$ with $r_{\boldsymbol{vv}}(k{-}i) = \mathrm{E}\left[\boldsymbol{v}(k)\boldsymbol{v}(i)^H\right] = \sigma_v^2 I_m\,\delta_{ki}$. Assume we receive $M$ samples:

$$\boldsymbol{Y}_M(k) \;=\; \mathcal{T}_M(\boldsymbol{h})\,A_{M+N-1}(k) + \boldsymbol{V}_M(k) \tag{3}$$

where $\boldsymbol{Y}_M(k) = [\boldsymbol{y}^H(k{-}M{+}1)\cdots \boldsymbol{y}^H(k)]^H$ and similarly for $\boldsymbol{V}_M(k)$, and $\mathcal{T}_M(\boldsymbol{h})$ is a block Toepltiz matrix with $M$ block rows and $[\boldsymbol{H}\;\; 0_{m\times(M-1)}]$ as first block row. We shall simplify the notation in (3) with $k = M{-}1$ to

$$\begin{aligned}\boldsymbol{Y} \;=\; \mathcal{T}(\boldsymbol{h})\,A + \boldsymbol{V} \;\;&=\;\; \mathcal{T}_K(\boldsymbol{h})\,A_K + \mathcal{T}_U(\boldsymbol{h})\,A_U + \boldsymbol{V}\\ &=\;\; \mathcal{A}_K\boldsymbol{h} + \mathcal{A}_U\boldsymbol{h} + \boldsymbol{V}\;.\end{aligned} \tag{4}$$

Where $\mathcal{T}_K(\boldsymbol{h})$ and $\mathcal{T}_U(\boldsymbol{h})$ represent respectively the portions of $\mathcal{T}(\boldsymbol{h})$ that correspond to $A_k$ ($M_K$ known symbols) and $A_U$ ($M_U$ unknown symbols).

$$\mathcal{T}(\boldsymbol{h}) = \left[ \begin{array}{c|c} \mathcal{T}_K(\boldsymbol{h}) & \mathcal{T}_U(\boldsymbol{h}) \end{array} \right] \quad (5)$$

Here we assume for simplicity that the known symbols are gathered at the beginning of the block. On the other hand, $\mathcal{A}$ is a block Toeplitz matrix filled with the elements of A while $\mathcal{A}_K$ and $\mathcal{A}_U$ are block Toeplitz matrices filled with the elements of $A_K$ and $A_U$ respectively.

## 3. CHANNEL APPROXIMATION

In this section we introduce the concept of approximating the channel by neglecting some taps at the tail during the estimation process. The neglect is justified by the fact that the estimation of those taps will introduce an estimation error that exceeds the approximation error. In fact, this is true if the power of the channel approximation error times the input is below the noise power at some finite SNR. In this case, the approximation error does not count and we have an approximated channel whose length varies with SNR. However, in order to make this channel approximation, we need to have a certain finite (and small) covariance of the part of the channel that we are going to neglect (approximation error). Hence, in a way or another this looks like a Bayesian approach. In a deterministic model, we don't indeed have any prior information about the channel that would allow to make such an approximation.

Assume that $\boldsymbol{h}$ can be split into two parts, the approximated part $\boldsymbol{h}_a$ $(mN_a \times 1)$ and the neglected part $\boldsymbol{h}_n$ $(mN_n \times 1)$ where $N = N_a + N_n$. Hence, we can write $\boldsymbol{h} = [\boldsymbol{h}_a^H \quad \boldsymbol{h}_n^H]^H$. The number of the neglected taps $N_n$ should be upper bounded by $\min(N-1, M_K)$. When the length of the training sequence is greater than the number of the channel taps, the interpretation of this bound is that the approximated channel should be composed of at least one tap. Hence, the maximum number of taps that can be neglected is $N-1$. However, when the number of the channel taps is greater than the training sequence, and apart from the identifiability issues that may raise here, the number of the neglected channel taps should not exceed $M_K$. This is due to the fact that every further neglected tap will lead to a one symbol loss. Actually, the size of $\mathcal{T}(\widehat{\boldsymbol{h}})$ is $mM \times (M+N-1)$, however, when some taps are neglected, the size of the estimated channel matrix is reduced to $mM \times (M+N_a-1)$. This reduction in the number of columns leads in general to a reduction in the number of the detected symbols. On the other hand, we are treating a semi-blind scenario where we assume that the training sequence (which fortunately there is no need to be detected) is gathered at the beginning of the block. This permits a margin of $M_K$ symbols, at the beginning of the block, to be skipped in the detection process. On the contrary, in the blind scenario there is no allowable margin and consequently, every neglected tap will lead to a one symbol loss. This puts a severe limitation for implementing this approach in the blind scenario. Fortunately, this is no more true in the cyclic prefix case. Taking a close look at the structure of the FIR cyclic prefix channel matrix, shows that there is obviously no symbol loss due to the neglected taps. This is true because the estimated channel matrix has a size $mM \times M$ which is independent of the number of taps. In fact this feature makes our approach more attractive in the context of cyclic prefix systems. To illustrate the procedure by which the neglected channel length is determined, we start with the description of the channel model used throughout this paper. In fact we consider a Rayleigh fading channel with exponentially decaying PDP for the channel between each transmitting and receiving antenna pair as follows: $e^{-wn}$ where $n = 0 : N-1$ and $w$ is a constant that controls how fast the decaying is. Hence, if we denote by $C^o_{\mathbf{h}}$ the channel covariance matrix, which is diagonal in this case because the taps are independent, then $C^o_{\mathbf{h}} = I_m \otimes C$ where $C = \text{diag}\{e^{-wn}, n = 0 : N-1\}$. Assume that the PDP and the variance of the noise are known (in practice they are estimated from the received signal), we start searching from the tail, for the maximum number of taps whose power times the power of the symbols is less than the variance of the noise. Mathematically, this can be written as:

$$\max_i \quad \sigma_a^2 \sum_i C(N-i, N-i) \qquad 0 \le i \le min(N-2, M_K-1) \quad (6)$$

The above maximization is done subject to the following constraint: $\sigma_a^2 \sum_i C(N-i, N-i) \le \sigma_v^2$. If we cannot find $i$ that fulfills the above constraint, this means that we can't neglect any part of the channel. Otherwise, the last $i+1$ taps in the tail of the channel can be neglected and consequently the length of the neglected part is $(i+1) \times m$. Now, we may reformulate the model in (4) as follows:

$$\begin{array}{rcl} \boldsymbol{Y} & = & \mathcal{T}(\boldsymbol{h})\, A + \boldsymbol{V} \\ & = & \underbrace{\mathcal{T}(\boldsymbol{h}_a)}_{Mm\times(M+N_a-1)} \quad A_a + \underbrace{\mathcal{T}(\boldsymbol{h}_n)}_{Mm\times(M+N_n-1)} \quad A_n + V \\ & = & \mathcal{T}(\boldsymbol{h}_a)\, A_a + Z \\ & = & \mathcal{A}_a \boldsymbol{h}_a + \boldsymbol{Z}. \end{array} \quad (7)$$

where $\mathcal{T}(\boldsymbol{h}_a)$ and $\mathcal{T}(\boldsymbol{h}_n)$ are Toeplitz matrices containing the elements of $\boldsymbol{h}_a$ and $\boldsymbol{h}_n$ respectively. On the other hand, $A_a$ constitutes the last $(M+N_a-1)$ elements of $A$ while $A_n$ constitutes the first $(M+N_n-1)$ elements of $A$. Finally, $Z = \mathcal{T}(\boldsymbol{h}_n)\, A_n + V$ is in general a spatially and temporally colored Gaussian noise with covariance $R_{ZZ}$. It should be noted that $R_{ZZ}$ varies from one estimator to another, depending on how we treat $A_n$ as we will see later. However, $h_n$ is going to be treated always as random with Gaussian distribution.

To treat the semi-blind case correctly, we have to split the approximated channel in its turn into two parts. These two parts correspond respectively to the known and the unknown symbols in analogy to what we have done in (4). Hence we can write:

$$\begin{array}{rcl} Y & = & \mathcal{T}_K(\boldsymbol{h}_a)\, A_{K,a} + \mathcal{T}_U(\boldsymbol{h}_a)\, A_U + Z \\ & = & \mathcal{A}_{K,a}\boldsymbol{h}_a + \mathcal{A}_{U,a}\boldsymbol{h}_a + Z \end{array} \quad (8)$$

where $\mathcal{T}_K(\boldsymbol{h}_a)$ and $\mathcal{T}_U(\boldsymbol{h}_a)$ contain the first $(M_K - N_n)$ and the last $(M+N-1-M_K)$ columns of $\mathcal{T}(\boldsymbol{h}_a)$ respectively. Similarly, $A_{K,a}$ and $A_U$ contain the first $(M_K - N_n)$ and the last $(M+N-1-M_K)$ elements of $A_a$. Finally, $\mathcal{A}_{K,a}$ and $\mathcal{A}_{U,a}$ are Toeplitz matrices filled with the elements of $A_{K,a}$ and $A_U$ respectively. It is worth noting that only $A_{K,a}$ undergoes a change compared to $A_K$ in (4) while $A_U$ remains unchanged. This is true thanks to the upper bound on the length of the neglected channel imposed in (6).

## 4. ENHANCED ESTIMATORS

In [4] we have introduced a general framework that permitted the derivation of three Bayesian semi-blind channel estimators and another three deterministic ones. Among those estimators, there were four that jointly estimate the channel and the symbols while the remaining two were based on estimating the channel and marginalizing the symbols. In the following sections, we will show a slight variation of those estimators relying on the channel approximation approach that was introduced in the previous section. On the other hand, there is an important difference between the model we stated in (4) and that used in [4] namely, in the latter we neglected a part of the received signal that contains both known and unknown symbols, whereas in this paper we are using an optimal model that allows a proper exploitation of the training sequence and the blind part of the received data.

### 4.1 SB-ML-ML (SB-DML)

We start with SB-ML-ML or what is called SB-DML in the literature [7]. In this case, both the unknown symbols and the approximated channel are considered as deterministic unknowns to be estimated. Thus, the cost function is given by:

$$\min_{A_U, \boldsymbol{h}_a} \quad ||Y - \mathcal{T}(h_a)A||^2_{R_{ZZ}} =$$
$$\min_{A_U, \boldsymbol{h}_a} \quad ||Y - \mathcal{T}_K(h_a)A_K - \mathcal{T}_U(h_a)A_U||^2_{R_{ZZ}} \quad (9)$$

The nonlinear LS optimization can be done by iterating between minimization with respect to $A_U$ and $\boldsymbol{h}$. By doing so, we get the following estimates:

$$\widehat{\boldsymbol{h}_a} = (\mathcal{A}_a^H R_{ZZ}^{-1}\mathcal{A}_a)^{-1}\mathcal{A}_a^H R_{ZZ}^{-1}\boldsymbol{Y} \tag{10}$$

$$\widehat{A_U} = (\mathcal{T}_U^H(\boldsymbol{h}_a)R_{ZZ}^{-1}\mathcal{T}_U(\boldsymbol{h}_a))^{-1}\mathcal{T}_U^H(\boldsymbol{h}_a)R_{ZZ}^{-1}(\boldsymbol{Y}-\mathcal{T}_K(h_a)A_{K,a}) \tag{11}$$

where $R_{ZZ} = \mathcal{A}_n C_{h_n}^o \mathcal{A}_n^H + \sigma_v^2 I$.

On the other hand, we can derive the deterministic CRB that represents a lower bound for this estimator as shown in [8]. Doing so we get:

$$DCRB_{det,joint}^{app} = \Big\{\mathcal{A}_a^H R_{ZZ}^{-1}\Big[R_{ZZ} - \mathcal{T}_U(\boldsymbol{h}_a)\Big(\mathcal{T}_U(\boldsymbol{h}_a)^H R_{ZZ}^{-1}\mathcal{T}_U(\boldsymbol{h}_a)\Big)^{-1}\mathcal{T}_u(\boldsymbol{h}_a)^H\Big]R_{ZZ}^{-1}\mathcal{A}_a\Big\}^{-1}$$

At high SNR, there is no room to neglect any tap and we have $N_a = N$ hence $\boldsymbol{h}_a = \boldsymbol{h}$ and consequently $\boldsymbol{h}_n$ disappears completely. As a result $R_{ZZ} = \sigma_v^2 I$. Substituting this result in (12), we get the same formula derived in [8] for the $DCRB_{det,joint}$.

$$DCRB_{det,joint} = \sigma_v^2\ \left(\mathcal{A}^H P_{\mathcal{T}_U(\mathbf{h})}^{\perp}\mathcal{A}\right)^{-1} \tag{12}$$

Where $P_{\mathcal{T}_U(\boldsymbol{h})}^{\perp} = I - P_{\mathcal{T}_U(\boldsymbol{h})}$ and $P_{\mathcal{T}_U(\boldsymbol{h})} = \mathcal{T}_U(\boldsymbol{h})(\mathcal{T}_U^H(\boldsymbol{h})\mathcal{T}_U(\boldsymbol{h}))^{-1}\mathcal{T}_U^H(\boldsymbol{h})$ is the projection matrix on $\mathcal{T}_U(\boldsymbol{h})$. However, at low SNR there are usually many taps at the tail of the channel that are immersed in the noise. As a consequence, they can be neglected without having any negative effect on the detection of the symbols at the receiver. On the contrary, as we will see in the simulation section, neglecting these taps enhances the detection quality at the receiver. Hence, $\boldsymbol{h}$ is approximated by $\boldsymbol{h}_a$ and there is a term that depends on $\boldsymbol{h}_n$, and that appears in $R_{ZZ}$. At a sufficient low SNR, $\sigma_v^2 I$ dominates $R_{ZZ}$ so we can neglect the term that depends on $\boldsymbol{h}_n$. Substitute this result in (12) we get:

$$DCRB_{det,joint}^{app} \cong \sigma_v^2\ \left(\mathcal{A}_a^H P_{\mathcal{T}_U(\boldsymbol{h}_a)}^{\perp}\mathcal{A}_a\right)^{-1} \tag{13}$$

To prove that the approximation approach, we propose in this paper, enhances the channel estimation quality at the receiver, we compare the CRB in (13) with a part of the CRB matrix stated in [8] namely, the part that corresponds to the approximated channel. Let's call this part $\widetilde{DCRB}_{det,joint}$. It is composed of the first $mN_a$ rows and columns of $DCRB_{det,joint}$.

Knowing that CRB is the inverse of the Fisher Information Matrix (FIM), let $\widetilde{FIM}_{\mathbf{h}_a\mathbf{h}_a}(\boldsymbol{h})$ denotes the FIM of the first $N_a$ taps of the channel where we estimate not only those taps but also the remaining $N_n$ taps and $FIM_{\mathbf{h}_a\mathbf{h}_a}$ denotes the FIM of the approximated channel where we are interested only in estimating the first $N_a$ taps. Now, we can write $\widetilde{DCRB}_{det,joint} = \widetilde{FIM}_{\mathbf{h}_a\mathbf{h}_a}^{-1}(\boldsymbol{h})$ and $DCRB_{det,joint}^{app} = FIM_{\mathbf{h}_a\mathbf{h}_a}^{-1}$. On the other hand, it is well known that the FIM of $\boldsymbol{h}$ can be decomposed into four parts corresponding to different combinations of $\boldsymbol{h}_a$ and $\boldsymbol{h}_n$. In order to extract the $\widetilde{FIM}_{\boldsymbol{h}_a\boldsymbol{h}_a}(\boldsymbol{h})$ from these FIMs, we apply the Schur's complement so we get: $\widetilde{FIM}_{\mathbf{h}_a\mathbf{h}_a}(\boldsymbol{h}) = FIM_{\mathbf{h}_a\mathbf{h}_a} - FIM_{\mathbf{h}_a\mathbf{h}_n}FIM_{\mathbf{h}_n\mathbf{h}_n}^{-1}FIM_{\mathbf{h}_n\mathbf{h}_a}$ Since $FIM_{\mathbf{h}_a\mathbf{h}_n}FIM_{\mathbf{h}_n\mathbf{h}_n}^{-1}FIM_{\mathbf{h}_n\mathbf{h}_a} \geq 0$ i.e. a positive semi-definite matrix, we infer that $\widetilde{FIM}_{\mathbf{h}_a\mathbf{h}_a}(\boldsymbol{h}) \leq FIM_{\mathbf{h}_a\mathbf{h}_a}$ and consequently $\widetilde{DCRB}_{det,joint} \geq DCRB_{det,joint}^{app}$. This result shows that our approximation approach leads to an enhancement in the channel estimation quality. This result has been confirmed also by numerical simulations as will show later.

### 4.2 SB-GMAP-ML

This estimator is considered as an extension of the corresponding blind one proposed in [9], [10]. In this estimator we treat the unknown symbols as random with Gaussian distribution, while the approximated channel is considered deterministic to be jointly estimated with the unknown symbols. Hence, the cost function is given by:

$$\min_{A_U,\boldsymbol{h}_a}\ ||Y - \mathcal{T}_K(h_a)A_K - \mathcal{T}_U(h_a)A_U||_{R_{ZZ}}^2 + \frac{||A_U||^2}{\sigma_a^2}$$

Following the same methodology used in SB-ML-ML estimator we get:

$$\widehat{\boldsymbol{h}_a} = (\mathcal{A}_a^H R_{ZZ}^{-1}\mathcal{A}_a)^{-1}\mathcal{A}_a^H R_{ZZ}^{-1}\boldsymbol{Y} \tag{14}$$

$$\widehat{A_U} = (\mathcal{T}_U^H(\boldsymbol{h}_a)R_{ZZ}^{-1}\mathcal{T}_U(\boldsymbol{h}_a)+\frac{1}{\sigma_a^2}I)^{-1}\mathcal{T}_U^H(\boldsymbol{h}_a)R_{ZZ}^{-1}(\boldsymbol{Y}-\mathcal{T}_K(h_a)A_{K,a}) \tag{15}$$

where $R_{ZZ} = \sigma_a^2\mathcal{T}(\boldsymbol{h}_n)\mathcal{T}(\boldsymbol{h}_n)^H + \sigma_v^2 I$. It is worth noting that we treat $A_n$ as random with Gaussian distribution although it contains some known symbols. This approximation is justified by the fact that usually the number of the known symbols is small compared to the unknown symbols.

### 4.3 SB-GMAP-Elm-ML (SB-GML)

In this estimator we are only interested in estimating the approximated channel and the variance of the noise, while the unknown symbols are supposed to be eliminated during the estimation process. Hence, $\theta = [\boldsymbol{h}_a^H, \sigma_v^2]^H$. Furthermore, we consider the channel and the noise variance to be deterministic while the unknown symbols have a Gaussian distribution. Hence, the cost function is given by:

$$\min_{\boldsymbol{h}_a,\sigma_v^2}\ \ln|C_{YY}| + (Y - \mathcal{T}_K(\boldsymbol{h}_a)A_a)^H C_{YY}^{-1}(Y - \mathcal{T}_K(\boldsymbol{h}_a)A_a) \tag{16}$$

where $C_{YY} = E\ \ (\boldsymbol{Y} - \mathcal{T}_K(\boldsymbol{h}_a)A_a)(\boldsymbol{Y} - \mathcal{T}_K(\boldsymbol{h}_a)A_a)^H = \sigma_a^2\mathcal{T}_U(\boldsymbol{h}_a)\mathcal{T}_U(\boldsymbol{h}_a)^H + R_{ZZ}$ and $R_{ZZ} = \left(\sigma_a^2\,\mathrm{tr}\,(C_{h_n}^o) + \sigma_v^2\right) I$. This cost function can be minimized by resorting to the method of scoring ([11] see also [4]).

As for deriving the CRB that corresponds to this estimator, we can follow the same methodology used in [8]. Doing so we get these formulas:

$$J_{\theta\theta}^{sto}(i,j) = \mathrm{tr}\left\{C_{YY}^{-1}\frac{\partial C_{YY}}{\partial\theta_i^*}C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial\theta_j^*}\right)^H\right\} + \left[\mathcal{A}_{K,a}^H C_{YY}^{-1}\mathcal{A}_{K,a}\right]_{i,j}$$
$$J_{\theta\theta^*}^{sto}(i,j) = \mathrm{tr}\left\{C_{YY}^{-1}\frac{\partial C_{YY}}{\partial\theta_i^*}C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial\theta_j^*}\right)\right\} \tag{17}$$

where $\frac{\partial C_{YY}}{\partial\boldsymbol{h}_{a,i}^*} = \sigma_a^2\mathcal{T}_U(\boldsymbol{h}_a)\mathcal{T}_U(\frac{\partial\boldsymbol{h}_a}{\partial\boldsymbol{h}_{a,i}^*})^H$ and $\frac{\partial C_{YY}}{\partial\sigma_v^2} = \frac{1}{2}$. Once we compute both $J_{\theta\theta}$ and $J_{\theta\theta^*}$ from (17), we substitute them in ([8], eq(13)) to compute $J_{\theta_R\theta_R}$ where $\theta_R = [Re(\theta)^T\ Im(\theta)^T]^T$, $Re$ and $Im$ denotes Real and Imaginary respectively. Consequently, by using Schur's complement we can extract easily $J_{\boldsymbol{h}_a\boldsymbol{h}_a}$ from $J_{\theta_R\theta_R}$ then $DCRB_{sto,marg} = J_{\boldsymbol{h}\boldsymbol{h}}^{-1}$ follows directly. Following the same discussion elaborated in the SB-ML-ML section, we can show that $DCRB_{sto,marg}^{app}$ is lower than $\widetilde{DCRB}_{sto,marg}$ that can be drawn from $DCRB_{sto,marg}$ ([8],eq 23) by taking the first $mN_a$ rows and columns.

### 4.4 SB-ML-GMAP

This estimator is Bayesian since we treat the approximated channel as random with Gaussian distribution. However, the unknown symbols are considered as deterministic to be jointly estimated with the channel. Therefore, the cost function is given by:

$$\min_{A_U, \boldsymbol{h}_a} \quad ||Y - \mathcal{T}_K(h_a)A_K - \mathcal{T}_U(h_a)A_U||^2_{R_{ZZ}} + \boldsymbol{h}_a^H C_{h_a}^{o^{-1}} \boldsymbol{h}_a \quad (18)$$

$$\widehat{\boldsymbol{h}_a} = (\mathcal{A}_a^H R_{ZZ}^{-1} \mathcal{A}_a + C_{h_a}^{o^{-1}} \mathcal{A}_a^H R_{ZZ}^{-1} \boldsymbol{Y} \quad (19)$$

$$\widehat{A_U} = (\mathcal{T}_U^H(\boldsymbol{h}_a) R_{ZZ}^{-1} \mathcal{T}_U(\boldsymbol{h}_a))^{-1} \mathcal{T}_U^H(\boldsymbol{h}_a) R_{ZZ}^{-1} (\boldsymbol{Y} - \mathcal{T}_K(\boldsymbol{h}_a) A_{K,a}) \quad (20)$$

where $R_{ZZ} = \mathcal{A}_n C_{h_n}^o \mathcal{A}_n^H + \sigma_v^2 I$ and $C_{h_n}^o$ is the part of $C_h^o$ that corresponds to $\boldsymbol{h}_n$.

### 4.5 SB-GMAP-GMAP

In this estimator both the approximated channel and the unknown symbols are assumed random with Gaussian distribution. Moreover, they are supposed to be estimated jointly. The cost function is given by:

$$\min_{A_U, \boldsymbol{h}_a} \quad ||Y - \mathcal{T}_K(h_a)A_K - \mathcal{T}_U(h_a)A_U||^2_{R_{ZZ}} + \boldsymbol{h}_a^H C_{h_a}^{o^{-1}} \boldsymbol{h}_a + \frac{1}{\sigma_a^2} ||A_U||^2$$

Also here, following the same methodology used in SB-ML-ML estimator we get:

$$\widehat{\boldsymbol{h}_a} = (\mathcal{A}_a^H R_{ZZ}^{-1} \mathcal{A}_a + C_{h_a}^{o^{-1}}) \mathcal{A}_a^H R_{ZZ}^{-1} \boldsymbol{Y} \quad (21)$$

$$\widehat{A_U} = (\mathcal{T}_U^H(\boldsymbol{h}_a) R_{ZZ}^{-1} \mathcal{T}_U(\boldsymbol{h}_a) + \frac{1}{\sigma_a^2} I)^{-1} \mathcal{T}_U^H(\boldsymbol{h}_a) R_{ZZ}^{-1} (\boldsymbol{Y} - \mathcal{T}_K(h_a) A_{K,a}) \quad (22)$$

where $R_{ZZ} = \left(\sigma_a^2 \, \mathrm{tr}\,(C_{h_n}^o) + \sigma_v^2\right) \, I$.

### 4.6 SB-GMAP-Elm-GMAP

As in the case of SB-GMAP-Elm-ML, in this estimator the symbols are supposed to be eliminated. Hence, it can be considered as an extension of SB-GMAP-Elm-ML by exploiting the prior information that exists about the channel. The corresponding cost function is given by:

$$\min_{\boldsymbol{h}_a, \sigma_v^2} \quad \ln |C_{YY}| + (Y - \mathcal{T}_K(\boldsymbol{h}_a)A_a)^H C_{YY}^{-1} (Y - \mathcal{T}_K(\boldsymbol{h}_a)A_a) + \boldsymbol{h}_a^H C_{h_a}^{o^{-1}} \boldsymbol{h}_a \quad (23)$$

This cost function can be minimized using also the scoring method.

## 5. SIMULATIONS

In this section, we show, by means of MonteCarlo simulations, how our approach for approximating the channel leads to a superior performance compared to classical techniques. In each MonteCarlo simulation we generate a Rayleigh fading channel as discussed previously while for the symbols, we generate random 8PSK symbols to reflect the real world case. The performance of the different channel estimators is evaluated by means of the Normalized MSE (NMSE) vs. SNR. The SNR is defined as: SNR $= \frac{||\mathcal{T}(h)A||^2}{mM\,\sigma_v^2}$. The NMSE is defined as $\frac{avg\,||\boldsymbol{h}_a - \hat{\boldsymbol{h}}_a||^2}{avg\,||\boldsymbol{h}_a||^2}$. All the simulations are initialized by the semi-blind Subchannel Response Matching (SRM) estimate [12]. In Figure 1 we compare the performance of the SB-SRM and the SB-ML-ML estimators with their enhanced counterparts proposed in this paper. We can notice how the SB-SRM based on our approach, (SB-SRM-Approx), exceeds its counterpart (SB-SRM) by more than 7 dB at low SNR and by couple of dBs at moderate SNR. However, no more enhancement is possible at very high SNR because, as explained before, no taps can be neglected at this SNR. As for our SB-ML-ML-Approx, we can notice from the same figure that only an enhancement of 2.5 dB is possible at low SNR compared to its counterpart SB-ML-ML, while this advantage diminishes as SNR increases. On the other hand, we plot on the same figure both the $\widetilde{DCRB}_{det,joint}$ and $DCRB^{app}_{det,joint}$. We notice that the latter exceeds the former by around 5 dB at low SNR which means that there is a considerable room to enhance the estimators that treat the channel and the symbols as deterministic. However, we notice that SB-SRM-Approx, and not SB-ML-ML-Approx succeeds well in taking advantage of our approach and fills the gap between both CRBs. The result is somehow surprising because SB-SRM-Approx is considered as a non-weighted version of SB-ML-ML-Approx. Finally, it is obvious that our approach leads the SB-ML-ML-Approx to almost attain the $\widetilde{DCRB}_{det,joint}$. It is well known that the latter is only attainable by SB-ML-ML asymptotically in SNR while it is not attainable asymptotically in the number of data. In Figure 2 we compare SB-GMAP-Elm-ML with our proposed counterpart. It is clear that the gain offered by our approach ( 6 dB) at low to moderate SNR is tremendous. Also on the same figure, we plot both $DCRB_{sto,joint}$ and $DCRB^{app}_{sto,joint}$. Once again, we can notice that our approach leads to a lower bound ( 2 dB). It is interesting to note here also that our approach leads SB-GMAP-Elm-Ml-Approx to attain $\widetilde{DCRB}_{det,joint}$ which is not attainable by SB-GMAP-Elm-ML. In Figure 3 we can observe once again the great enhancement (5 dB) obtained by our approximation approach at low SNR, specially in the SB-ML-GMAP case whereas in the SB-GMAP-ML case the gain is around 2 dB. In Figure 4 we show numerically that SB-GMAP-GMAP (which jointly estimate the channel and the symbols) and SB-GMAP-Elm-GMAP (which estimates the channel and marginalizes the symbols) are perfect in the sense that our approach for approximating the channel is not capable of enhancing their performance.

In all the simulations we have conducted up till now the PDP is assumed to be known perfectly. However, in Figure 5 we estimate the PDP from the received data and we apply our approach using the SB-SRM algorithm. We compare the enhancement obtained by our approach relying on the estimated PDP against the perfect PDP. We observe that although the improvement degrades when we use the estimated PDP but the reduction in the NMSE compared to the traditional SB-SRM is still interesting. At last, to prove that our approach leads also to an enhancement in the probability of error (Pe), we plot in Figure 6 the Pe for SB-SRM and an enhanced version of it based on our channel approximation approach. We can readily observe the considerable gain (2 dB) offered by our approach at medium and low SNR.

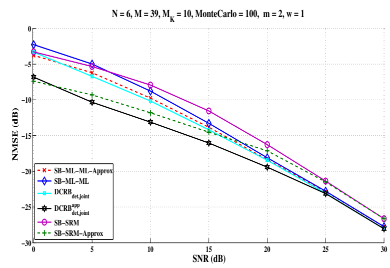


Figure 1: NMSE vs. SNR for SB-SRM, SB-ML-ML, $\mathbf{DCRB_{det,joint}}$ and $\mathbf{DCRB^{app}_{det,joint}}$.

## 6. CONCLUSION

We have introduced in the context of semi-blind channel estimation a new approach that relies on the partial exploitation of the PDP of the channel (assumed known or estimated from the received data) to reduce the channel estimation error. Based on this approach, we have shown that, by neglecting some taps at the tail of the channel that are immersed in noise, the quality of the channel estimation has been improved considerably. The proposed approach has been implemented to a series of deterministic and Bayesian estimators

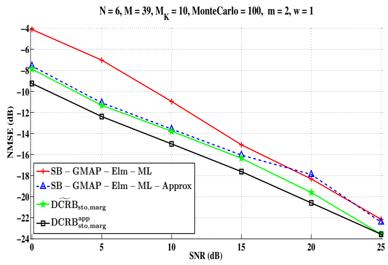


Figure 2: NMSE vs. SNR for SB-GMAP-Elm-ML, $\mathbf{DCRB_{sto,joint}}$ and $\mathbf{DCRB^{app}_{sto,marg}}$.

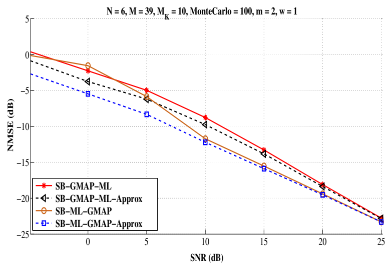


Figure 3: NMSE vs. SNR for SB-ML-GMAP and SB-GMAP-ML.

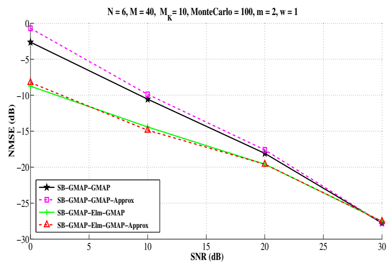


Figure 4: NMSE vs. SNR for SB-GMAP-GMAP and SB-GMAP-Elm-GMAP.

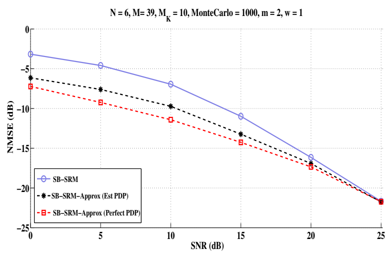


Figure 5: NMSE vs. SNR for SB-SRM with perfect and estimated PDP.

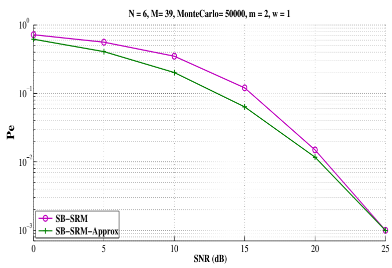


Figure 6: Probability of error vs. SNR for SB-SRM and its enhanced counterpart.

introduced previously. We have shown by numerical simulations that there is a great enhancement in the NMSE over a wide range of SNR. Moreover, we have shown analytically that the CRBs of two of the proposed estimators are lower than their corresponding ones that exist in the literature. Finally, we have shown also numerically that not only the NMSE of the channel has been improved but also the probability of error of the detected symbols. On the other hand, our simulations show that there is no room left to enhance the estimators that take full advantage of the prior information about the channel and the symbols. This fact has been reflected in both SB-GMAP-GMAP and SB-GMAP-Elm-GMAP performance where our approach has not succeeded to reduce their NMSE.

## 7. ACKNOWLEDGMENT

EURECOM's research is partially supported by its industrial members: BMW Group, Swisscom, Cisco, ORANGE, SFR, ST Ericsson, Thales, Symantec, SAP, Monaco Telecom. The research reported herein was also partially supported by the French ANR project SESAME, the EU FET project CROWN, the EU NoE Newcom++ and a PACA regional scholarship BDO550.

## REFERENCES

[1] A.K. Jagannatham and B.D. Rao, "Constrained ML Algorithms For Semi-Blind MIMO Channel Estimation," in *IEEE Globecom 2004*, Dallas, Texas, USA, Dec 2004.

[2] W. Yang; Y. Cai; Y. Xun;, "Semi-blind Channel Estimation for OFDM Systems," in *IEEE Proc. VTC*, Seattle, USA, May 2006.

[3] Wei-Chieh Huang, Chun-Hsien Pan, Chih-Peng Li, and Hsueh-Jyh Li, "Subspace-Based Semi-Blind Channel Estimation in Uplink OFDMA Systems," *IEEE Transactions on Broadcasting*, vol. 56, no. 1, pp. 58–65, Jan. 2010.

[4] S.-M. Omar, D.T.M. Slock, and O. Bazzi, "Bayesian Semi-Blind FIR Channel Estimation Algorithms in SIMO Systems," in *Proc. IEEE Int'l Workshop on Signal Processing Advances in Wireless Comm's (SPAWC)*, San Francisco, USA, June 2011.

[5] S.-M. Omar, D.T.M. Slock, and O. Bazzi, "Bayesian Blind FIR Channel Estimation Algorithms in SIMO Systems," in *Proc. IEEE Int'l Workshop on Statistical Signal Processing (SSP)*, Nice, France, June 2011.

[6] D.T.M. Slock, "Blind Fractionally-Spaced Equalization, Perfect-Reconstruction Filter Banks and Multichannel Linear Prediction," in *Proc. ICASSP 94 Conf.*, Adelaide, Australia, April 1994.

[7] J. Ayadi, E. de Carvalho, and D.T.M. Slock, "Blind and Semi-Blind Maximum Likelihood Methods for FIR Multichannel Identification," in *Proc. ICASSP 98 Conf.*, Seattle, USA, May 1998.

[8] S.-M. Omar, D.T.M. Slock, and O. Bazzi, "Bayesian and Deterministic CRBs for Semi-Blind Channel Estimation in SIMO Single Carrier Cyclic Prefix Systems," Submitted, IEEE PIMRC, Toronto, Canada, Sep. 2011.

[9] D.T.M. Slock and C.B. Papadias, "Further Results on Blind Identification and Equalization of Multiple FIR Channels," in *Proc. ICASSP 95 Conf.*, Detroit, Michigan, May 1995.

[10] E. de Carvalho and D.T.M. Slock, "Maximum-Likelihood FIR Multi-Channel Estimation with Gaussian Prior for the Symbols," in *Proc. ICASSP 97 Conf.*, Munich, Germany, April 1997.

[11] P.E. Kay, W. Murray, and M.H. Wright, *Practical Optimization*, Academic press, London, 1981.

[12] E. Carvalho and D.T.M. Slock, "Deterministic quadratic semi-blind FIR multichannel estimation : algorithms and performance," in *IEEE International conference on acoustic, speech and signal (ICASSP)*, Istanbul, Turkey, June 2000.