# Hybrid Supervised-Unsupervised Channel Estimation Scheme with Dynamic Transmission of Pilots

**Adriana Dapena** · **Paula M. Castro**
· **José A. García-Naya**

**Abstract** Multiple-Input Multiple-Output (MIMO) digital communications standards typically include pilot symbols in the definition of the transmit signals with the purpose of acquiring the Channel State Information (CSI) using supervised algorithms at the receiver side. Such pilot symbols convey no information and, therefore, system throughput, spectral efficiency and transmit energy consumption are all penalized. In this article, we propose to acquire the CSI combining supervised and unsupervised algorithms. Our strategy avoids the periodical transmission of unnecessary pilots by using a simple decision criterion to determine the time instants when the performance obtained with an unsupervised algorithm degrades or, equivalently, the time instants when pilots are required. We show the performance of this scheme for MIMO systems with Decision Feedback Equalizers (DFE) at the receiver.

**Keywords** Semi-blind approach · Blind Source Separation · MIMO Systems · Decision Feedback Equalizers

Department of Electronics and Systems. University of A Coruña.
Campus de Elviña, 15071 A Coruña, Spain.
{adriana,pcastro,jagarcia}@udc.es
http://gtec.des.udc.es

## 1 Introduction

The Least Squares (LS) criterion is the best known and most widely used approach to acquire the Channel State Information (CSI) at the receiver side of a digital communications system. The LS-based channel estimation consists in minimizing the Mean Square Error (MSE) between the transmitted pilot symbols and the observed ones. Such pilot symbols are periodically sent by the transmitter without considering whether the receiver actually needs such pilots to track the variations of the channel. Since pilot symbols do not convey information, system throughput, spectral efficiency and transmit energy consumption are all penalized.

On the contrary, the so-called unsupervised techniques are able to estimate the channel coefficients directly from the observations, without requiring pilot symbols. The only assumption is that the transmitted signals have to be statistically independent. Unfortunately, such unsupervised approaches —also known as Blind Source Separation (BSS) techniques [5]— have some major drawbacks:

- They present a poor performance when the channel suffers a significant variation.
- They present two ambiguities: permutation and gain. This means that the signals can be recovered in an incorrect order and with a different scale with respect to the ones used at the transmitter side.
- Adaptive BSS algorithms require a considerable number of observations (received symbols) until they converge to a channel estimation and, most of them, can converge to undesirable solutions.

The combination of supervised and unsupervised techniques can avoid their respective drawbacks while preserving their advantages [5, 11]. On the one hand, thanks to the utilization of an unsupervised criterion, the length of the pilot sequence needed to estimate the channel is reduced and the spectral efficiency is thus increased. On the other hand, the incorporation of pilot symbols suppresses undesirable solutions, accelerates the convergence and eliminates the permutation ambiguity typical in unsupervised approaches. Unfortunately, both supervised and unsupervised approaches need a perfect synchronization between transmitter and receiver in order to ensure a perfect knowledge of the time instants when the pilot sequence starts. Nowadays, most of the wireless communications standards require time synchronization and dedicate preamble sequences for that purpose. However, such standards also make use of pilot symbols with the objective of estimating the wireless channel, but without taking into account how many pilot symbols are actually necessary for a given scenario. This limitation of today's standards constitutes the main motivation of our work.

We propose a strategy to determine, in frame-based wireless communications systems, when pilot symbols have to be transmitted. Additionally, we also propose a so-called hybrid channel estimation scheme employing an unsupervised technique when no pilot symbols are being transmitted together with a supervised algorithm when pilot symbols are available. The main goal is to track the variations of the wireless channel. If such variations can be assumed by the unsupervised approach, then no pilot symbols have to be transmitted. However, when a significant variation is detected, the transmitter is notified and pilot symbols are transmitted. At the receiver side, a supervised channel estimation algorithm utilizes such pilots to update the channel estimation and to start again the unsupervised algorithm until a new significant channel variation is observed.
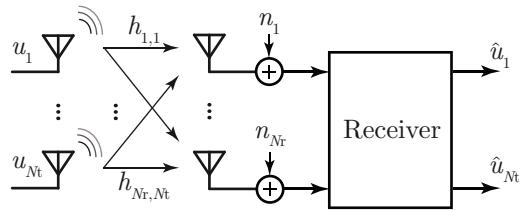
**Fig. 1** MIMO channel model.

The immense majority of current wireless communications standards make use of feedback channels (usually limited in terms of throughput) between the receiver and the transmitter sides of the link to periodically send the CSI (and information about the received power). However, to the knowledge of the authors, none of the current standards —even those under development— make use of such information to decide whether it is really necessary to send pilot symbols. This feedback channel will be used in this work to inform the transmitter when the channel has suffered a significant variation and pilot symbols are thus required. Consequently, the penalties caused by the continuous transmission of pilot symbols are mitigated. Although our method could be combined with different types of receiver designs, we will show the performance when the Decision-Feedback Equalizer (DFE) is utilized at the receiver.

DFE was initially proposed to reduce the effect of multiple-delayed copies of a signal transmitted in Single-Input Single-Output (SISO) systems [1], i.e. to equalize the channel. It consists of two linear filters: the feedforward filter, whose input is the received sequence, and the feedback filter, whose input is the detected sequence. The basic idea is to use feedback from a past decision to cancel the interference of the symbols that have already been detected. Extensions of this idea to Multiple-Input Multiple-Output (MIMO) systems have been proposed by several authors (see, for instance, [9]). In fact, DFE has been proposed for the implementation of several wireless communications standards such as Digital Terrestrial Multimedia Broadcast (DTMB) [8].

All derivations are based on the assumption of zero-mean and stationary random variables. Vectors and matrices are denoted by lower case bold and capital bold letters, respectively. A scalar function $f$ and its first and second derivatives are respectively denoted by $f(\cdot)$, $f'(\cdot)$, and $f''(\cdot)$. We use $E[\cdot]$, $tr(\cdot)$, $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, $det(\cdot)$, $adj(\cdot)$, and $\|\cdot\|_2$, for expectation, trace of a matrix, complex conjugation, transposition, conjugate transposition, determinant of a matrix, adjunct matrix, and Euclidean norm, respectively. $\ln(\cdot)$ denotes the natural logarithm, and $\mathbf{I}_N$ and $\mathbf{O}_N$ denote the $N \times N$ identity and zero matrix, respectively.

## 2 System Model

We consider a MIMO system with $N_t$ transmit antennas and $N_r$ receive antennas, as plotted in Figure 1. For simplicity, we will assume $N_t = N_r = N$. The data symbols $\mathbf{u} = [u_1, ..., u_N]^T$ are transmitted to the different receive antennas such that, for flat
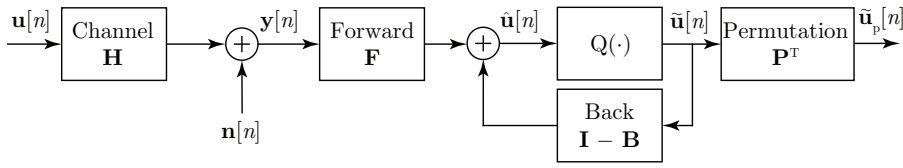
**Fig. 2** MIMO system with DFE.

fading channels, the received signals have the form

$$y_j[n] = \sum_{i=1}^{N} h_{ji}[q]u_i[n] + n_j[n],$$

where $n = 0, 1, 2, \ldots$ corresponds to samples periodically spaced every $T_s$ seconds and $q$ denotes the time slot. For brevity, we henceforth omit the slot index $q$.

In a compact form, we can express the vector of received signals, $\mathbf{y}[n]$, as follows

$$\mathbf{y}[n] = \mathbf{H}\,\mathbf{u}[n] + \mathbf{n}[n], \tag{1}$$

where $\mathbf{u}[n]$ is the data symbol vector (sources), $\mathbf{n}[n]$ is the noise vector, and $\mathbf{H}$ contains the channel coefficients, $h_{ij}$, from the $i$-th transmit antenna to the $j$-th receive antenna,

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N} \\ h_{21} & h_{22} & \cdots & h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N1} & h_{N2} & \cdots & h_{NN} \end{bmatrix}.$$

The channel remains unchanged during the transmission of a block of $N_B$ symbols per antenna[1], i.e. over the whole data frame. We also assume that the transmitted sources are independent and identically distributed with unit power, i.e. $\mathbf{C_u} = E[\mathbf{u}[n]\mathbf{u}^H[n]] = \mathbf{I}_N$. This channel model is known as time-varying flat block fading channel and this assumption is made in the following steps.

DFE is a powerful technique to combat the distortion introduced by the wireless channel. The block diagram of a MIMO system employing DFE is plotted in Figure 2. As said before, the DFE equalizer uses feedback from past decisions to cancel the interference of the symbols that have already been detected. It consists of two linear filters: the feedforward filter $\mathbf{F}$, whose input is the received sequence, and the feedback filter $\mathbf{I}_N - \mathbf{B}$, whose input is the sequence detected in the previous step. The feedforward filter provides spatial causality and ensures that the error is white. The feedback filter, however, exploits causality for the feedback loop and Inter-Symbol Interference (ISI) cancellation due to its strictly lower triangular structure [9]. For achieving optimum performance, the symbols have to be detected according to a specific ordering, which is performed by means of the permutation matrix $\mathbf{P}$.

As shown in Figure 2, the estimated signal $\hat{\mathbf{u}}[n]$ can be expressed as

$$\hat{\mathbf{u}}[n] = \mathbf{F}\mathbf{y}[n] + (\mathbf{I}_N - \mathbf{B})\,\tilde{\mathbf{u}}[n]. \tag{2}$$

The quantized symbols $\tilde{\mathbf{u}}[n]$ are reordered (permuted) by $\mathbf{P}^T$ to obtain the detected symbols $\tilde{\mathbf{u}}_p$. Given that $\mathbf{P}\mathbf{P}^T = \mathbf{I}$, we obtain the symbols $\tilde{\mathbf{u}}[n] = \mathbf{P}\tilde{\mathbf{u}}_p[n]$. Assuming

---

[1] From now on we always consider symbols transmitted or received per antenna.

Compute $\boldsymbol{\Phi} = (\mathbf{H}^H \mathbf{C_n}^{-1} \mathbf{H} + \mathbf{C_u}^{-1})^{-1}$
Initialize $\mathbf{P} = \mathbf{I}_N$ and $\mathbf{D} = \mathbf{O}_N$
**for** $i = 1, \ldots, N$
    Find $q = \arg\min_{p=1,\ldots,N} \boldsymbol{\Phi}(p, p)$
    Exchange the $i$-th and $q$-th rows of $\mathbf{P}$ to obtain $\mathbf{P}_i$
    Compute $\mathbf{P} = \mathbf{P}_i \mathbf{P}$ and $\boldsymbol{\Phi} = \mathbf{P}_i \boldsymbol{\Phi} \mathbf{P}_i^T$
    Set $\mathbf{D}(i, i) = \boldsymbol{\Phi}(i, i)$
    Compute $\boldsymbol{\Phi}(i : N, i) = \boldsymbol{\Phi}(i : N, i)/\mathbf{D}(i, i)$
    Compute $\boldsymbol{\Phi}(i + 1 : N, i + 1 : N) = \boldsymbol{\Phi}(i + 1 : N, i + 1 : N)$
       $-\boldsymbol{\Phi}(i + 1 : N, i)\boldsymbol{\Phi}(i + 1 : N, i)^H \mathbf{D}(i, i)$
Let $\mathbf{L}$ be the lower triangular part of $\boldsymbol{\Phi}$
Compute $\mathbf{B} = \mathbf{L}^{-1}$ and $\mathbf{F} = \mathbf{D}\mathbf{L}^H \mathbf{P}\mathbf{H}^H \mathbf{C_n}^{-1}$

**Table 1** Calculation of MSE-DFE filters with ordering.

that decisions made prior to each detection are correct (i.e. $\tilde{\mathbf{u}}_p[n] = \mathbf{u}[n]$) and using Equation (2), we have the error vector defined as follows

$$\mathbf{e}_p[n] = \mathbf{P}\mathbf{u}[n] - \hat{\mathbf{u}}[n] = \mathbf{P}\mathbf{u}[n] - (\mathbf{F}\mathbf{y}[n] + (\mathbf{I}_N - \mathbf{B})\,\mathbf{P}\mathbf{u}[n])$$
$$= \mathbf{B}\mathbf{P}\mathbf{u}[n] - \mathbf{F}\mathbf{y}[n]. \tag{3}$$

The DFE feedforward and feedback filters are found by minimizing this error restricting the feedback filter $\mathbf{B}$ to being lower triangular [9], i.e.

$$\left\{ \mathbf{P}_{MSE}^{DFE}, \mathbf{F}_{MSE}^{DFE}, \mathbf{B}_{MSE}^{DFE} \right\} = E\left[ \left\| \mathbf{P}\mathbf{u}[n] - \hat{\mathbf{u}}[n] \right\|_2^2 \right] = E\left[ \left\| \mathbf{B}\mathbf{P}\mathbf{u}[n] - \mathbf{F}\mathbf{y}[n] \right\|_2^2 \right]$$
$$\text{s.t.: } \mathbf{B} \text{ is unit lower triangular.} \tag{4}$$

Table 1 summarizes the procedure proposed in [9] to find the matrices $\mathbf{P}$, $\mathbf{B}$, and $\mathbf{F}$. Some remarks must be made about this implementation. First, the estimation of channel matrix is used in the first step to compute $\boldsymbol{\Phi} = (\mathbf{H}^H \mathbf{C_n}^{-1} \mathbf{H} + \mathbf{C_u}^{-1})^{-1}$ and, in the last one, to compute $\mathbf{F}$. Second, with the aim of minimizing the effect derived from feeding back erroneous decisions, the iterative procedure allows the signals to be extracted in a descending order according to the diagonal elements in the MSE matrix $\boldsymbol{\Phi}$. Compared to previous ordering proposals, as shown in [10], its computational complexity is less than that exhibited by the Bell Laboratories Layered Space-Time (BLAST) proposal [6], with no penalization with respect to the Bit Error Ratio (BER) performance. Finally, note that $\mathbf{C_n} = \sigma_n^2 \mathbf{I}_N$ and $\mathbf{C_u} = \mathbf{I}_N$.

## 3 Channel Estimation Approaches

In the literature, different strategies for recovering the transmit signals (and estimate the channel matrix) can be found. In particular, since the transmission model in Equation (1) corresponds to a linear combination of the transmit signals, we consider a linear recovering system whose outputs are computed as follows

$$\mathbf{z}[n] = \mathbf{W}^H[n]\mathbf{y}[n], \tag{5}$$

where $\mathbf{W}[n]$ is an $N \times N$ matrix that can be found using different supervised and unsupervised algorithms. The model in Equation (5) is only utilized to estimate the channel

matrix $\mathbf{H}$ required by the DFE procedure described in Table 1. Although unsupervised methods could be used to directly estimate the transmit signals, the resulting performance is significantly lower than that obtained when the DFE is employed. Note that combining both Equations (1) and (5), the output $\mathbf{z}[n]$ can be rewritten as a linear combination of the sources

$$\mathbf{z}[n] = \mathbf{\Gamma}[n]\mathbf{u}[n] + \mathbf{W}^H[n]\mathbf{n}[n], \tag{6}$$

where $\mathbf{\Gamma}[n] = \mathbf{W}^H[n]\mathbf{H}$ represents the overall separating/mixing system (or gain matrix). The matrix $\mathbf{W}[n]$ must be selected in order to recover the transmitted sources while the noise power is reduced. In situations with a moderate noise power, this occurs when $\mathbf{W}[n]$ is an estimation of the inverse of the channel matrix. However, a scaled and permuted version of the channel matrix is considered as a valid estimate. As a consequence, the recovering algorithms can be used to solely estimate the channel matrix $\mathbf{H}$ needed by the DFE.

In this section, we explain several methods to obtain the channel matrix estimate, denoted by $\hat{\mathbf{H}}$, which is required by the DFE. To avoid confusion, we will denote by $\mathbf{W}_s[n]$, $\mathbf{W}_u[n]$, and $\mathbf{W}_{SB}[n]$ the recovering matrices corresponding to supervised, unsupervised (blind), or semi-blind algorithms, respectively. From now on, we assume a system utilizing two frame types: pilot frames and data frames, both composed of $N_B$ symbols per transmit antenna. Consequently, pilot symbols are only available at the receiver when a pilot frame has been transmitted. For example, when a supervised approach is employed, the channel is only estimated when a pilot frame is received.

### 3.1 Supervised Approach

The classical method to estimate the channel matrix consists in using a supervised strategy. The best known algorithm arises from the minimization of the MSE between the outputs $\mathbf{z}[n]$ and the desired signals $\mathbf{u}[n]$ [7]. Mathematically, the cost function is written as

$$J_{\mathrm{MSE}} = \sum_{i=1}^{N} E\left[|z_i[n] - u_i[n]|^2\right] = E\left[\mathrm{tr}\left((\mathbf{W}_s^H[n]\mathbf{y}[n] - \mathbf{u}[n])(\mathbf{W}_s^H[n]\mathbf{y}[n] - \mathbf{u}[n])^H\right)\right]. \tag{7}$$

The gradient of this cost function is obtained as

$$\nabla_{\mathbf{W}} J_{\mathrm{MSE}} = E\left[\mathbf{y}[n](\mathbf{W}_s^H[n]\mathbf{y}[n] - \mathbf{u}[n])^H\right]. \tag{8}$$

The direct form of determining the optimum matrix $\mathbf{W}_s[n]$ consists in finding the point where the gradient vanishes. For that purpose, we utilize the definition of the Hessian matrix containing the second derivatives of $J$ [3]. In particular, the optimum separating matrix corresponding to Equation (8) is given by the following expression

$$\nabla_{\mathbf{W_s}} J_{\mathrm{MSE}} = 0 \Rightarrow \mathbf{W}_s = \mathbf{C_y^{-1}}\mathbf{C_{yu}}, \tag{9}$$

where $\mathbf{C_y} = E[\mathbf{y}[n]\mathbf{y}^H[n]]$ is the autocorrelation of the observations and $\mathbf{C_{yu}} = E[\mathbf{y}[n]\mathbf{u}^H[n]]$ is the cross-correlation between the observations and the desired signals. Equation (9) is referred to as the Widrow-Hoff solution.

In practice, the desired signal is only known when a pilot frame (solely constituted by pilot symbols) is transmitted and the above expectations $\mathbf{C_y}$ and $\mathbf{C_{yu}}$ are estimated by sample averaging over the block of $N_B$ pilot symbols that constitute the frame, i.e.

$$\mathbf{C_y} = \mathbf{y}\mathbf{y}^H, \ \mathbf{C_{yu}} = \mathbf{y}\mathbf{u}^H, \text{where } \mathbf{y} \in \mathbb{C}^{N \times N_B}.$$

Table 2 presents the computational complexity of the implementation of Equation (9) in terms of the number of operations needed to process a frame of $N_B$ pilots in a system with $N$ antennas. It can be seen that the dominant factor is $N_B \times N^2$, which results in a cubic computational complexity.

### Widrow-Hoff Solution

| | |
|---|---|
| Compute $\mathbf{C_y}$ (or $\mathbf{C_{yu}}$) | $N^2 \times N_B$ complex multiplications |
| | $N^2 \times (N_B - 1)$ complex summations |
| Compute $\mathbf{C_y}^{-1}\mathbf{C_{yu}}$ | $N^3$ complex multiplications |
| | $N^2 \times (N - 1)$ complex summations |
| Matrix inversion | $O(N^3)$ for the Gauss-Jordan method |
| **Total** | $O(N^2 \times N_B)$ |

### Unsupervised approach Infomax-CM

| For each iteration: | |
|---|---|
| Compute $g(z) = z_i^*(1 - |z_i|^2)$ for the $N$ outputs | $N$ complex multiplications |
| | $N$ real summations |
| | $N$ real-complex multiplications |
| Compute $\mathbf{P}_1 = \mathbf{z}\mathbf{g}^H(\mathbf{z}[n]) - \mathbf{I}_N$ | $N^2$ complex multiplications |
| | $N$ complex summations |
| Compute $\mathbf{P}_2 = \mathbf{W}_u\mathbf{P}_1$ | $N^3$ complex multiplications |
| | $N^2 \times (N - 1)$ complex summations |
| Update $\mathbf{W}_u[n+1] = \mathbf{W}_u[n+1] + \mu\mathbf{P}_2$ | $N^2$ real-complex multiplications |
| | $N^2$ complex summations |
| | $N(N - 1)$ real summations |
| **Total** (for $N_B$ symbols) | $O(N^3 \times N_B)$ |

### Decision criteria

| For each frame: | |
|---|---|
| Compute $\mathbf{\Gamma} = \mathbf{W}_u^H\hat{\mathbf{H}}$ | $N^3$ complex multiplications |
| | $N^2 \times (N - 1)$ complex summations |
| Compute $\text{SI}_i = \sum_{\substack{j=1 \\ i \neq j}}^{N} \frac{|\gamma_{ij}|^2}{|\gamma_{ii}|^2}$ | $N^2$ complex multiplications |
| | $N(N - 1)$ real divisions |
| Compute $\text{SI} = \sum_{i=1}^{N} \text{SI}_i$ | $N$ real summations |
| **Total** (once per frame) | $O(N^3)$ |

**Table 2** Computational cost of the supervised and unsupervised algorithms as well as the decision criteria. $N_B$ is the number of symbols (when a data frame is transmitted) and also the number of pilots (when a pilot frame is sent), $N$ is the number of antennas at each side of the link, and SI is the signal interference term defined in Equation (15)

.

### 3.2 Unsupervised Approach

Many unsupervised algorithms have been proposed to estimate the channel matrix without using pilot symbols (see the overview in [5]). Among them, we will consider the Infomax algorithm proposed by Bell and Sejnowski in [2]. Given an activation function $h(\cdot)$, the idea proposed by these authors is to obtain the weighted coefficients of a neural network, $\mathbf{W}_u[n]$, in order to maximize a cost function $J_{MI}$ defined as the Mutual Information (MI) between the outputs before the activation function, $\mathbf{h}(\mathbf{z}[n]) = \mathbf{h}(\mathbf{W}_u^H[n]\mathbf{y}[n])$, and its inputs $\mathbf{y}[n]$:

$$J_{MI}(\mathbf{W}_u[n]) = \ln(\det(\mathbf{W}^h[n])) + \sum_{i=1}^{N} E[\ln(h_i'(z_i[n]))], \qquad (10)$$

where $h_i$ is the $i$–th element of the vector $\mathbf{h}(\mathbf{z}[n])$. Considering this criterion, Yang and Amari have proposed in [12] the following relative (or natural) algorithm

$$\begin{aligned}
\mathbf{W}_u[n+1] &= \mathbf{W}_u[n] + \mu\mathbf{W}_u[n]\mathbf{W}_u^H[n]\nabla_{\mathbf{W}_u}J_{\mathrm{MI}} \\
&= \mathbf{W}_u[n] + \mathbf{W}_u[n]\mathbf{W}_u^H[n]\mu\left(\mathbf{y}[n]\mathbf{g}^H(\mathbf{z}[n]) - \mathbf{W}_u^{-H}[n]\right) \\
&= \mathbf{W}_u[n] + \mu\mathbf{W}_u[n]\left(\mathbf{z}[n]\mathbf{g}^H(\mathbf{z}[n]) - \mathbf{I}_N\right),
\end{aligned} \qquad (11)$$

where $\mathbf{g}(\mathbf{z}[n]) = [-h_1''(z_1[n])/h_1'(z_1[n]), \cdots, -h_N''(z_N[n])/h_N'(z_N[n])]^T$ depends on the activation function. This criterion can be interpreted as an extension of the Constant Modulus (CM) algorithm when the non-linear activation function is $h(z) = e^{j\ z^*(1-|z|^2)}$, which implies to use $g(z) = z^*(1-|z|^2)$ in Equation (11) (see [4]). We denote this algorithm as Infomax-CM. Table 2 shows the computational complexity of the Infomax-CM algorithm. We can see that the dominant term is $N_B \times N^3$, which results in a complexity of $O(N^3 \times N_B)$.

### 3.3 Semi-blind Approach

In the literature, different ways of combining supervised and unsupervised approaches can be found [5,11]. In particular, for a SISO system, Zarzoso and Comon have proposed a semi-blind equalization technique considering the MSE and the CM criterion [13]. Taking this idea into account, the following algorithm can be proposed for the case of MIMO systems

$$\mathbf{W}_{SB}[n+1] = \mathbf{W}_{SB}[n+1] - \mu\left(\lambda\nabla_{\mathbf{W}_{SB}}J_{\mathrm{MSE}} + (1-\lambda)\nabla_{\mathbf{W}_{SB}}J_{\mathrm{MI}}\right), \qquad (12)$$

where $\lambda \in [0,1]$ can be considered as the relative degree of confidence between the unsupervised and the supervised parts. Note that $J_{\mathrm{MSE}}$ and $J_{\mathrm{MI}}$ correspond to the gradients shown in Equations (8) and (10), respectively. Without loss of generality, $N_P$ pilot symbols are assumed to appear at the beginning of the frame, which implies a perfect synchronization between transmitter and receiver. As in the case of the unsupervised algorithm, the complexity of the recursion in Equation (12) depends on the number of symbols processed by the unsupervised part, i.e. the computational complexity can be approximated by $(N_B - N_P) \times N^3$, resulting in $O(N^3 \times (N_B - N_P))$.

It is important to note that, although the recovering matrix of unsupervised and semi-blind approaches is updated for each symbol, this matrix is transmitted to the DFE after processing all the symbols.
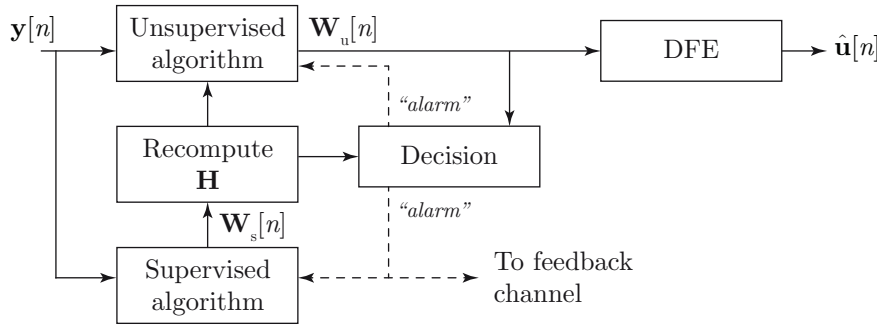
**Fig. 3** Block diagram for the hybrid approach.

## 4 Hybrid approach

Figure 3 shows the block diagram of the hybrid system proposed by us to estimate the channel matrix. If $\mathbf{W}_u[n]$ and $\mathbf{W}_s[n]$ are respectively the matrices of coefficients for the unsupervised and supervised modules as they were defined above, we start with an initial estimation of the channel matrix obtained using the supervised method, $\hat{\mathbf{H}} = \mathbf{W}_s^{-H}[n]$. This estimation is used to initialize the unsupervised algorithm to $\mathbf{W}_u[n] = \hat{\mathbf{H}}^{-H}$.

Each time a new frame is received, the unsupervised algorithm updates the separating matrix sample by sample using the unsupervised adaptive algorithm given in Equation (11) and the channel matrix needed by the DFE is estimated as $\hat{\mathbf{H}} = \mathbf{W}_u^{-H}[n]$. A "decision module" determines if the estimation obtained with the unsupervised algorithm is poor due to, for example, a large variation in the channel. When this occurs, the receiver sends an "alarm" to the transmitter. At this point, the next frame sent by the transmitter has to be a pilot frame. Afterwards, at the receiver, the supervised algorithm estimates the channel from the observed pilot symbols. Finally, such an estimation is employed to initialize the unsupervised algorithm.

The important question is how to determine when the unsupervised algorithm presents a poor performance and, consequently, a pilot frame has to be transmitted. Equation (6) indicates that each output contains a term corresponding to the desired source and another one due to the Signal Interference (SI). It is interesting to note that the initialization of the unsupervised algorithm removes the permutation ambiguity inherent to this class of learning rules. This implies that each output has the form

$$z_i[n] = \gamma_{ii}[n]u_i[n] + \sum_{j=1,i\neq j}^{N} \gamma_{ij}[n]u_j[n] + \mathbf{w}_i[n]^H \mathbf{n}[n], \tag{13}$$

where $\gamma_{ii}[n]$ are the diagonal elements of $\mathbf{\Gamma}[n]$ (i.e. the gains for each desired output). Dividing this equation by $\gamma_{ii}[n]$ and considering that the noise term is small with respect to the Signal Interference (SI) term, we obtain that the power of each output is given by

$$\frac{E[|z_i[n]|^2]}{|\gamma_{ii}[n]|^2} = E[|u_i[n]|^2] + \sum_{j=1,i\neq j}^{N} \frac{|\gamma_{ij}[n]|^2}{|\gamma_{ii}[n]|^2}E[|u_j[n]|^2] = E[|u_i[n]|^2] + \text{SI}_i[n]. \tag{14}$$

---

**Pilot frame**
Estimate the channel matrix, $\hat{\mathbf{H}}$, with the Widrow-Hoff solution in Equation (9).
Initialize the blind algorithm $\mathbf{W}_u[n] = \hat{\mathbf{H}}^{-H}$.

---

**Data frame**
For each symbol:
  Update $\mathbf{W}_u[n]$ using the recursive rule in Equation (11).
After the whole frame has been processed:
  Utilize $\mathbf{W}_u^{-1}[n]$ in the DFE to recover the user data.
  Compute $\mathbf{\Gamma}[n] = \mathbf{W}_u^H[n]\hat{\mathbf{H}}$.
  Compute SI according to Equation (15).
  If SI $> t$ then
    send an "alarm" to the transmitter indicating that a pilot frame has to be transmitted.

---

**Table 3** Description of the hybrid approach. The threshold $t$ is used to decide when a pilot frame has to be transmitted.

A large value of $\mathrm{SI}_i$ implies that the estimation of the channel matrix is poor. In this situation, pilot symbols have to be transmitted, i.e.

$$\mathrm{SI} = \sum_{i=1}^{N} \mathrm{SI}_i[n] = \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} \frac{|\gamma_{ij}[n]|^2}{|\gamma_{ii}[n]|^2} > t \rightarrow \text{Send an "alarm"}, \tag{15}$$

where $t$ is a real positive number (threshold). Assuming a frame-based communications systems, the gains $\gamma_{ij}[n]$ can be computed using $\mathbf{\Gamma}[n] = \mathbf{W}_u^H[n]\hat{\mathbf{H}}$, where $\hat{\mathbf{H}}$ is the estimated channel matrix generated by the supervised approach, and $\mathbf{W}_u^H[n]$ is the estimation obtained by the unsupervised approach after processing the current frame. If the signal interference value is greater than the threshold $t$, then the transmitter is informed to transmit a pilot frame. Obviously, the smaller the value of $t$, the smaller the error, but a larger number of pilot frames (and thus a larger number of pilots) will be needed.

Table 3 summarizes the steps performed by the proposed hybrid scheme depending on the frame type (with or without pilot symbols). The computational complexity is determined by means of two factors: the complexity of each supervised or unsupervised part, and the complexity associated to the decision criterion. As shown in Table 2, the complexity exhibited by the unsupervised algorithm is $O(N^3 \times N_B)$, which is greater than that exhibited by the supervised ($O(N^2 \times N_B)$) and the decision criteria ($O(N^3)$). Notice also that the computational complexity of the decision criterion does not depend on the number of symbols $N_B$, only on the number of antennas $N$, thus resulting in a significantly lower computational complexity compared to that of the channel estimation algorithms.

## 5 Simulation Results

In order to show the performance achieved with the proposed hybrid scheme, we present the results for several computer simulations performed considering that 4 900 symbols are transmitted in frames of size $N_B=100$ symbols mapped to a QPSK constellation. The system is formed by four transmit and four receive antennas, thus $N=4$. The

channel matrix is updated using the following linear model

$$\mathbf{H} = (1 - \alpha)\mathbf{H} + \alpha\mathbf{H}_r, \tag{16}$$

where $\mathbf{H}_r$ is a $4 \times 4$ matrix randomly generated according to a Gaussian distribution. We have evaluated the performance of the following schemes:

- Perfect CSI at the receiver, i.e. the DFE uses the generated channel matrix.
- The supervised approach in Equation (9) computed using a frame of $N_B$ pilot symbols transmitted each 10 frames.
- The unsupervised Infomax-CM algorithm initialized with the matrix obtained from the supervised approach. The step-size parameter was set to $\mu$=0.001.
- The semi-blind algorithm (see Equation (12)) employing frames formed by $N_B$=100 symbols in total, with $N_P$=5 pilots, $\mu$=0.001, and $\lambda$=0.8. Since there is no closed-form solution to select $\lambda$, we obtain the best value after carrying out a large number of simulations.
- The hybrid approach using different thresholds. $N_B$ pilots are transmitted in a pilot frame when the error exceeds the threshold valure. The step-size parameter for the Informax-CM is set to $\mu$=0.001.

All unsupervised algorithms have been initialized with the solution obtained using the supervised approach. The results have been obtained by averaging 500 independent realizations. Note that pilots have not been considered when BER is computed.
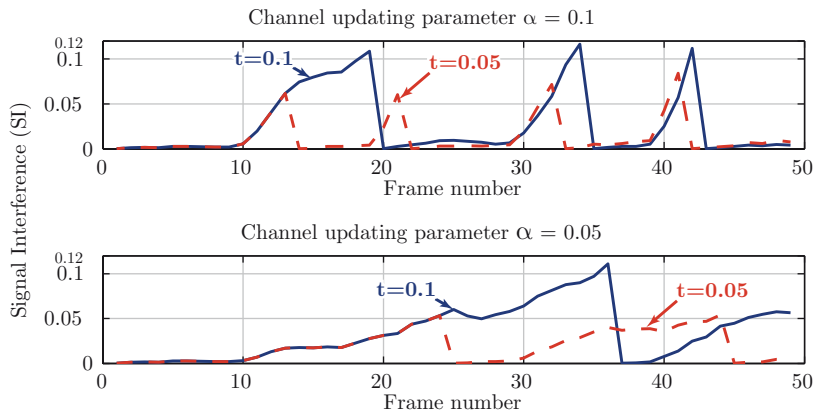


**Fig. 4** Experiment 1: SI for different values of the threshold ($t$) and the updating channel parameter ($\alpha$).

5.1 Experiment 1: Fixed Channel Updating Parameter

In the first experiment we have considered that the channel remains constant during 10 frames (i.e. the channel changes each 1 000 symbols). This is an ideal situation for the supervised approach because it exactly corresponds to the instants when pilot frames are transmitted.
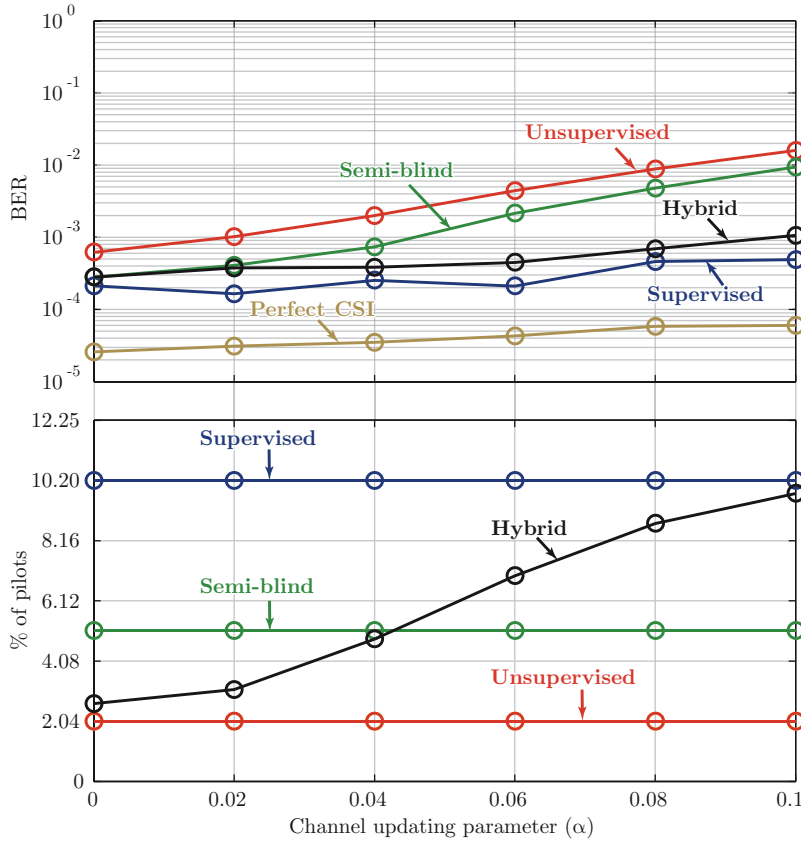
**Fig. 5** Experiment 1: BER and percentage of pilots versus the channel updating parameter for an SNR of 15 dB.

In order to show the effect obtained when the criterion in Equation (15) is employed, Figure 4 shows the SI obtained with two threshold values: $t=0.05$ and $t=0.1$. The top graph corresponds to an updating channel parameter $\alpha=0.1$, and the bottom one to $\alpha=0.05$. We can see that each time channel is updated, the SI increases in a considerable amount for $\alpha=0.1$, while this quantity becomes smaller for $\alpha=0.05$. In the figure, we can also see that a small value of the threshold reduces the time delay between channel variation detection and the instant when the unsupervised algorithm starts up correctly. For instance, note that two frames with pilots have to be transmitted for $\alpha=0.05$ and $t=0.05$, while only one is required for $t=0.1$. In fact, for the remaining simulations we consider $t=0.1$.

Figure 5 shows the BER and the percentage of transmitted pilots in terms of the channel updating parameter $\alpha$ for an SNR of 15 dB. Note the considerable improvement, in terms of BER, obtained with the hybrid approach compared to the Infomax-CM algorithm (unsupervised) and the semi-blind approach. Also note that the BER value is close to that obtained with the supervised approach, although the number of pilots is considerably lower.
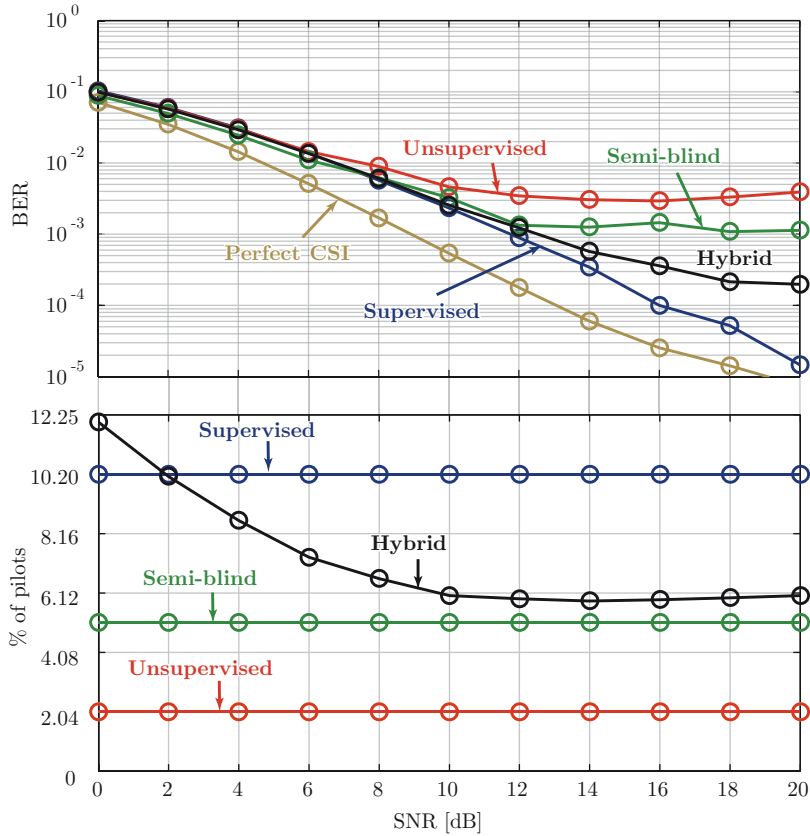
**Fig. 6** Experiment 1: BER and percentage of pilots in terms of SNR for $\alpha = 0.05$.

Figure 6 shows the BER and the percentage of pilots as a function of SNR when $\alpha$=0.05. Again, the hybrid approach improves the unsupervised and semi-blind algorithms in terms of the SNR needed to achieve the same target BER, reaching a BER performance close to that exhibited by the supervised approach. Note that the Infomax-CM algorithm presents a floor effect at an SNR of 10 dB, while for the hybrid approach this effect appears at about 14 dB, at which the BER value is quite low.

### 5.2 Experiment 2: Variable Channel Updating Parameter

In the second experiment, we have considered that the channel remains constant during a number of frames between 10 and 15 and randomly generated for each realization. Figure 7 shows the performance for an SNR of 15 dB with respect to the updating channel parameter $\alpha$. It can be seen that the BER exhibited by the hybrid approach matches that offered by the supervised algorithm. When compared to the Experiment 1, we observe that the percentage of required pilot symbols has decreased. The same conclusion is obtained from Figure 8, which illustrates the BER performance versus SNR considering $\alpha$=0.05. Note that the floor effect of the hybrid approach appears
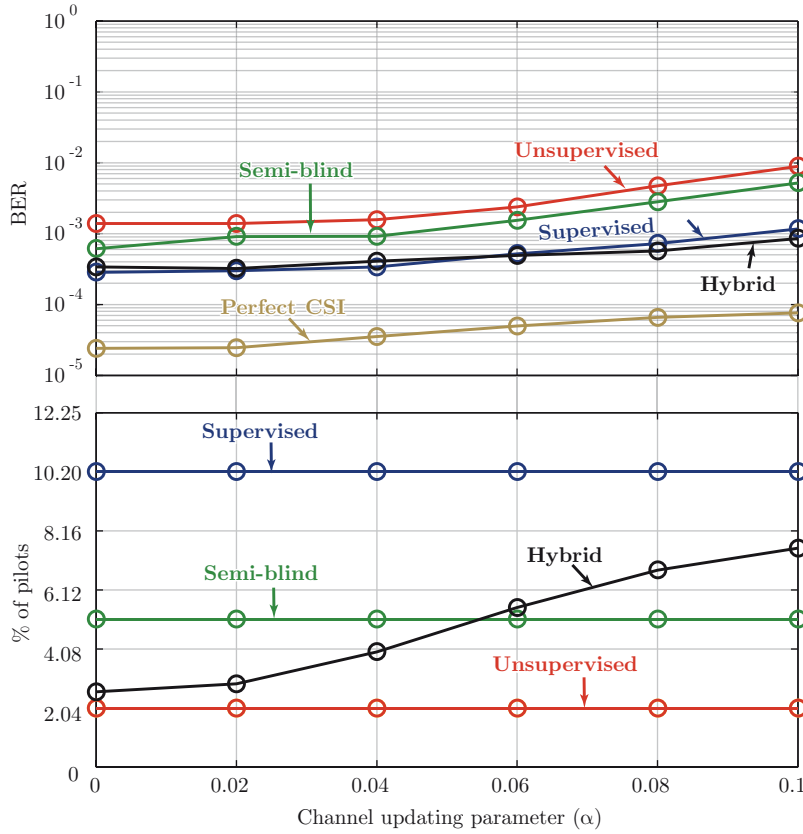
**Fig. 7** Experiment 2: BER and percentage of pilots in terms of the channel updating parameter for a SNR of 15 dB.

at an SNR of about 16 dB, with a corresponding BER value lower than $10^{-3}$. For this BER value, only 230 pilots (4.69 %) have been transmitted, which represents a considerable reduction compared to the 500 pilots (10.20 %) needed by the supervised approach.

## 6 Conclusions

This paper deals with the utilization of supervised and unsupervised algorithms for estimating the channel matrix in MIMO systems. Considering a communication model where the channel is block flat fading, we have proposed a simple way to dynamically determine the time instants when pilots must be transmitted and, therefore, when a supervised algorithm must be used. During the remaining time, the CSI is acquired employing an unsupervised approach, without requiring pilots. Simulation results show that the novel approach provides an adequate BER performance with a low overhead produced by the transmission of pilot symbols. Although in this paper we have considered two frame types (pilot frames and user data frames), similar ideas could be
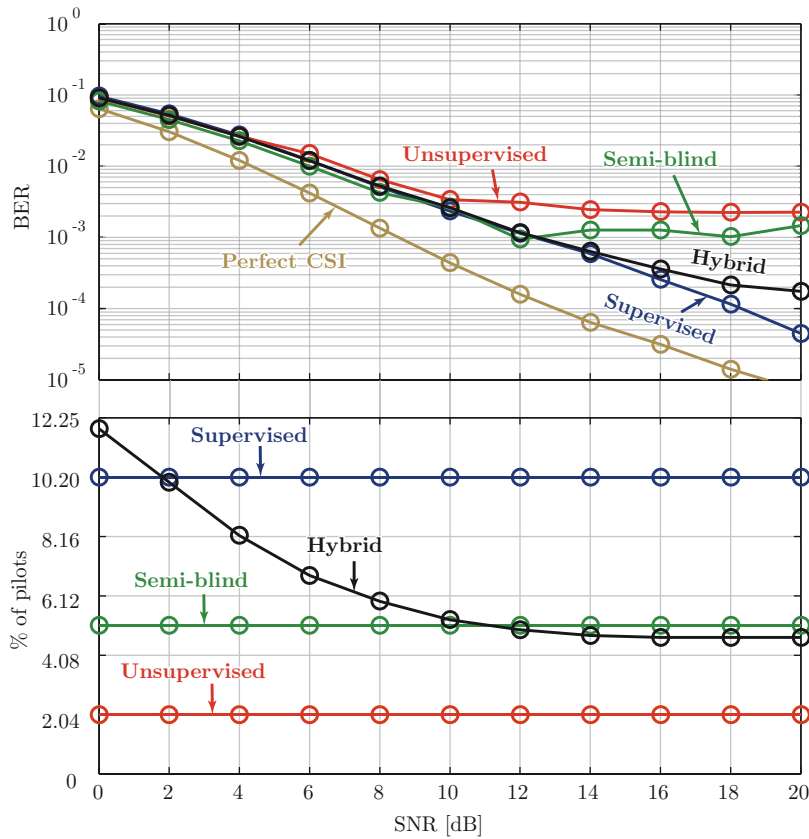
**Fig. 8** Experiment 2: BER and percentage of pilots in terms of SNR for $\alpha = 0.05$.

used when a single frame type is employed. In this case, the decision criterion would indicate whether pilots must be included or not in the frame.

## References

1. Austin, M.E.: Decision feedback equalization for digial communication over dispersive channels. Tecnhical report 437. Lincoln Laboratory (1967)
2. Bell, A., Sejnowski, T.: An information-maximization approach to blind separation and blind deconvolution. Neural Computation **7**(6), 1129–1159 (1995)
3. Benveniste, A., Métivier, M., Priourent, P.: Adaptive Algorithms and Stochastic Approximations. Springer-Verlag, New York (1990)
4. Castedo, L., Macchi, O.: Maximizing the information transfer for adaptive unsupervised source separation. In: Proc. SPAWC'97, pp. 65–68 (1997)
5. Comon, P., Jutten, C.: Handbook of Blind Source Separation, Independent Component Analysis and Applications. Academic Press (2010)

6. Golden, G., Foschini, J., Valenzuela, R., Wolniansky, P.: Detection algorithm and initial laboratory results using V-BlAST space-time communication architecture. Electronics Letters **35**(1), 14–15 (1999)
7. Haykin, S.: Neural Networks A Comprehensive Foundation. Macmillan College Publishing Company, New York (1994)
8. Huang, Q., Xu, Y., Long, B.: A novel Block-DFE algorithm for DTMB standard. In: International Conference on Communications and Mobile Computing (CMC 2009), vol. 1, pp. 433–436 (2009). DOI 10.1109/CMC.2009.98
9. Kusume, K., Joham, M., Utschick, W.: MMSE block decision-feedback equalizer for spatial multiplexing with reduced complexity. In: Proc. IEEE Global Telecommunications Conference, pp. 2540–2544, vol. 4 (2004). Dallas, Texas, USA
10. Kusume, K., Joham, M., Utschick, W., Bauch, G.: Cholesky factorization with symmetric permutation applied to detecting and precoding spatially multiplexed data streams. IEEE Transactions on Signal Processing **55**(6), 3089–3103 (2007)
11. Medles, A., Slock, D.: Space-Time Wireless Systems: From Array Procesing to MIMO Communications, chap. Blind and Semiblind MIMO channel estimation. Cambridge University Press (2006)
12. Yang, H., Amari, S.-I.: Adaptive on-line algorithms for blind separation: Maximum entropy and minimum mutual information. Neural Computation **9**(7), 1457–1482 (1997)
13. Zarzoso, V., Comon, P.: Semi-blind constant modulus equalization with optimal size. In: Proc. International Conference on Acoustics, Speech, and Signal Processing (2005). Philadelphia, USA