

# A Novel Hybrid Approach to Improve Performance of Frequency Division Duplex Systems with Linear Precoding

Paula M. Castro, José A. García-Naya, Daniel Iglesia, and Adriana Dapena

**Abstract**—*Linear precoding (LP) is an attractive technique to combat interference in Multiple-Input/Multiple-Output (MIMO) communication systems because it reduces cost and power consumption in the receive equipment. In most Frequency Division Duplex systems with LP, the Channel State Information (CSI) is acquired at the receiver by using supervised algorithms which work with pilot symbols periodically sent by the transmitter. Subsequently, the CSI is sent to the transmit side through a low cost feedback channel. In order to reduce the overhead inherent to the periodical transmission of training data, we propose to acquire the CSI by combining supervised and unsupervised algorithms. The simulation results show that the performance achieved with the proposed scheme is clearly better than that with standard algorithms.*

## I. INTRODUCTION

The increased demand of multimedia contents has produced a continuous development of new techniques to improve the capacity of digital communication systems. For instance, current transmission standards for *Multiple-Inputs/Multiple-Outputs* (MIMO) systems include precoders in order to guarantee that the link throughput be maximized [1], [2]. Precoding algorithms for MIMO can be sub-divided into linear and nonlinear precoding types. In this work, we will consider *Linear Precoding* (LP) approaches because they achieve reasonable throughput performance with lower complexity than nonlinear precoding approaches.

When implementing precoding the base station should know the *Channel State Information* (CSI). In most *Frequency Division Duplex* (FDD) systems, the transmitter cannot obtain the CSI from the received signals, even under the assumption of perfect calibration, because the channels are not reciprocal. Instead, the CSI is estimated at the receiver side and it is transmitted back by means of a feedback channel. In current standards, the channel estimation is performed by using supervised algorithms that work with pilot symbols periodically sent. Pilot symbols do not convey information and, therefore, the system throughput or, equivalently, the spectral efficiency.

In this paper, we propose to combine two important paradigm of Neural Networks: supervised and unsupervised learning. The kind of learning that must be used is decided by using simple criterion that determines the time instant when the channel has suffered a considerable variation. In these instants, a supervised algorithm is used to estimate the channel from pilot symbols. On the contrary, the unsupervised

The authors are with the Department of Electronics and Systems, University of A Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain (phone: +34 981 167000; email: {pcastro,jagarcia,dani,adriana}@udc.es).

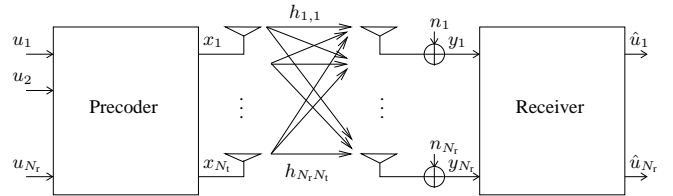


Fig. 1. System with Precoding over Flat MIMO Channel.

algorithm known as Infomax [3] is used when the variation is small.

This work is organized as follows. Sections II describe the signal and system model. Section III presents our method to combine supervised and unsupervised channel estimation and detection. Illustrative computer simulations are presented in Section V and some concluding remarks are made in Section VI.

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. The  $K \times K$  identity matrix is denoted by  $\mathbf{I}_K$  and  $\mathbf{0}_K$  is a  $K$ -dimensional zero vector. We use  $E[\bullet]$ ,  $\text{tr}(\bullet)$ ,  $(\bullet)^*$ ,  $(\bullet)^T$ ,  $(\bullet)^H$ ,  $\det(\bullet)$ , and  $\|\bullet\|_2$ , for expectation, trace of a matrix, complex conjugation, transposition, conjugate transposition, determinant of a matrix, and Euclidean norm, respectively. The  $i$ -th element of a vector  $\mathbf{x}$  is  $x_i$ .

## II. SYSTEM MODEL

We consider a MIMO system with  $N_t$  transmit antennas and  $N_r$  receive antennas, as plotted in Figure 1. The precoder generates the transmit signal  $\mathbf{x}$  from all data symbols  $\mathbf{u} = [u_1, \dots, u_{N_r}]$  belonging to the different receive antennas  $1, \dots, N_r$ . We denote the equivalent lowpass channel impulse response between the  $j$ -th transmit antenna and the  $i$ -th receive antenna as  $h_{i,j}(\tau, t)$ . Thus, the randomly time-varying channel is characterized by the  $N_r \times N_t$  matrix  $\mathbf{H}(\tau, t)$  defined as

$$\mathbf{H}(\tau, t) = \begin{pmatrix} h_{1,1}(\tau, t) & h_{1,2}(\tau, t) & \cdots & h_{1,N_t}(\tau, t) \\ h_{2,1}(\tau, t) & h_{2,2}(\tau, t) & \cdots & h_{2,N_t}(\tau, t) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_r,1}(\tau, t) & h_{N_r,2}(\tau, t) & \cdots & h_{N_r,N_t}(\tau, t) \end{pmatrix}.$$

Suppose that the transmitted signal from the  $i$ -th transmit antenna is  $x_i(t)$ . Then, the receive signal at the  $j$ -th receive

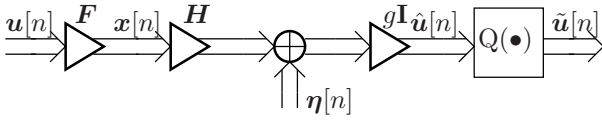


Fig. 2. MIMO System with Linear Transmit Filter (Linear Precoding).

antenna is given by

$$y_j(t) = \sum_{i=1}^{N_t} h_{j,i}(\tau, t) * x_i(t) + \eta_j(t)$$

where  $\eta_j(t)$  is the additive noise. In matrix notation, this equation can be rewritten as

$$\mathbf{y}(t) = \mathbf{H}(\tau, t) * \mathbf{x}(t) + \boldsymbol{\eta}(t)$$

where  $\mathbf{x}(t) = [x_1(t), \dots, x_{N_t}(t)]^T \in \mathbb{C}^{N_t}$ ,  $\mathbf{y}(t) = [y_1(t), \dots, y_{N_r}(t)]^T \in \mathbb{C}^{N_r}$ , and  $\boldsymbol{\eta}(t) = [\eta_1(t), \dots, \eta_{N_r}(t)]^T \in \mathbb{C}^{N_r}$ . For flat fading channels, the channel matrix  $\mathbf{H}(\tau, t)$  is transformed into the matrix  $\mathbf{H}(t)$  given by

$$\mathbf{H}(t) = \begin{pmatrix} h_{1,1}(t) & h_{1,2}(t) & \cdots & h_{1,N_t}(t) \\ h_{2,1}(t) & h_{2,2}(t) & \cdots & h_{2,N_t}(t) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_r,1}(t) & h_{N_r,2}(t) & \cdots & h_{N_r,N_t}(t) \end{pmatrix}$$

and the received signal is now

$$y_j(t) = \sum_{i=1}^{N_t} h_{ji}(t)x_i(t) + \eta_j(t)$$

which can be expressed in matrix form as

$$\mathbf{y}(t) = \mathbf{H}(t)\mathbf{x}(t) + \boldsymbol{\eta}(t). \quad (1)$$

In general, if we let  $f[n] = f(nT_s + \Delta)$  denote samples of  $f(t)$  every  $T_s$  seconds with  $\Delta$  being the sampling delay and  $T_s$  the symbol time, then sampling  $\mathbf{y}(t)$  every  $T_s$  seconds yields the discrete time signal  $\mathbf{y}[n] = \mathbf{y}(nT_s + \Delta)$  given by

$$\mathbf{y}[n] = \mathbf{H}[q]\mathbf{x}[n] + \boldsymbol{\eta}[n] \quad (2)$$

where  $n = 0, 1, 2, \dots$  corresponds to samples spaced with  $T_s$  and  $q$  denotes the slot time. The channel remains unchanged during a block of  $N_B$  symbols, i.e., over the data frame. Note that this discrete time model is equivalent to the continuous time model in Equation (1) only if ISI between samples is avoided, i.e. if the *Nyquist criterion* is satisfied. In that case, we will be able to reconstruct the original continuous signal from the samples by means of interpolation. This channel model is known as *time-varying flat block fading channel* and this assumption is made in the following.

For brevity, we omit the slot index  $q$  in the sequel.

#### A. Linear Precoding

In this section, we will obtain a form of performing the pre-equalizer (or precoder) step at the transmitter. Since this operation is performed prior to transmission, it is only

possible for a centralized transmitter as in the downlink of a cellular system.

We assume that the receive filter is an identity matrix (multiplied by a scalar  $g$ , with  $g \in \mathbb{C}$ ) allowing for decentralized receivers). The goal is to find the optimum transmit filter  $\mathbf{F}$ . Therefore, the transmit and receive filter are given by the matrices  $\mathbf{F} \in \mathbb{C}^{N_t \times N_t}$  and  $\mathbf{G} = g\mathbf{I} \in \mathbb{C}^{N_r \times N_r}$ , respectively. In other words, the number of scalar data streams is  $N_r$ . The resulting communications system is shown in Figure 2. It can be seen from the figure how the data symbols  $\mathbf{u}[n]$  are passed through the transmit filter  $\mathbf{F}$  to form the transmit signal  $\mathbf{x}[n] = \mathbf{F}\mathbf{u}[n] \in \mathbb{C}^{N_t}$ . Note that the constraint for the transmit energy must be fulfilled, i.e.

$$\mathbb{E} [\|\mathbf{x}[n]\|_2^2] = \text{tr}(\mathbf{F}\mathbf{C}_u\mathbf{F}^H) \leq E_{tx}.$$

The received signal is given by

$$\mathbf{y}[n] = \mathbf{H}\mathbf{F}\mathbf{u}[n] + \boldsymbol{\eta}[n] \in \mathbb{C}^{N_r}$$

where  $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$  and  $\boldsymbol{\eta}[n] \in \mathbb{C}^{N_r}$  is the *Additive White Gaussian Noise* (AWGN).

After multiplying by the receive gain  $g$ , we get the estimated symbols

$$\hat{\mathbf{u}}[n] = g\mathbf{H}\mathbf{F}\mathbf{u}[n] + g\boldsymbol{\eta}[n] \in \mathbb{C}^{N_r}. \quad (3)$$

Clearly, the restriction that all the receivers apply the same scalar weight  $g$  is not necessary for decentralized receivers. Replacing  $\mathbf{G}$  by a diagonal matrix suffices (e.g. [4]). However, usually no closed form can be obtained for the precoder if  $\mathbf{G}$  is diagonal. Fortunately,  $\mathbf{F}$  can be found in closed form for  $\mathbf{G} = g\mathbf{I}$ . Thus, we use  $\mathbf{G} = g\mathbf{I}$  in the following.

Although Wiener filtering for precoding has been dealt with by only a few authors [5] in comparison with other criteria for precoding, it is a very powerful transmit optimization that minimizes the *Mean Square Error* (MSE) with a transmit energy constraint [2], [6]–[8], i.e.

$$\begin{aligned} \{\mathbf{F}_{WF}, g_{WF}\} &= \underset{\{\mathbf{F}, g\}}{\text{argmin}} \mathbb{E} [\|\mathbf{u}[n] - \hat{\mathbf{u}}[n]\|_2^2] \\ \text{s.t.} &: \text{tr}(\mathbf{F}\mathbf{C}_u\mathbf{F}^H) \leq E_{tx}. \end{aligned} \quad (4)$$

In [5], it has been demonstrated that, which leads to a unique solution if we restrict  $g$  to being positive real, the solution for the Wiener filter is given by

$$\begin{aligned} \mathbf{F}_{WF} &= g_{WF}^{-1} (\mathbf{H}^H \mathbf{H} + \xi \mathbf{I})^{-1} \mathbf{H}^H \\ g_{WF} &= \sqrt{\frac{\text{tr}((\mathbf{H}^H \mathbf{H} + \xi \mathbf{I})^{-2} \mathbf{H}^H \mathbf{C}_u \mathbf{H})}{E_{tx}}}. \end{aligned} \quad (5)$$

### III. ADAPTIVE ALGORITHMS

The model explained in Section II states that the observations are linear and instantaneous mixtures of the transmitted signals  $\mathbf{x}[n]$  of Equation (2), i.e.

$$\mathbf{y}[n] = \mathbf{H}\mathbf{x}[n] + \boldsymbol{\eta}[n]. \quad (6)$$

For the case of the linear precoder described in previous section, this equation can be rewritten as follows

$$\mathbf{y}[n] = \mathbf{H}\mathbf{F}\mathbf{u}[n] + \boldsymbol{\eta}[n]. \quad (7)$$

This means that the observations  $\mathbf{y}[n]$  are instantaneous mixtures of the data symbols  $\mathbf{u}[n]$ , where the mixing matrix is given by  $\mathbf{H}\mathbf{F}$ . For brevity, we will denote this mixing matrix in the sequel as  $\mathbf{A}$ , so the observations  $\mathbf{y}[n]$  can be obtained in this way

$$\mathbf{y}[n] = \mathbf{A}\mathbf{d}[n] + \boldsymbol{\eta}[n]. \quad (8)$$

In accordance with our target, matrix  $\mathbf{A}$  may represent the channel matrix [cf. Equation (6)], or the whole coding-channel matrix,  $\mathbf{H}\mathbf{F}$  [cf. Equation (7)]. In the first case,  $\mathbf{d}[n]$  represents the code signal  $\mathbf{x}[n] = \mathbf{F}\mathbf{u}[n]$  and, in the second case, the data one,  $\mathbf{u}[n]$ . We assume that the mixing matrix is unknown but full rank nevertheless. Without any loss of generality we can suppose that the source data have a normalized power equal to one since possible differences in power may be included into the mixing matrix  $\mathbf{A}$ .

In order to recover the source data, we will use a linear system whose output is a combination of the observations, expressed as

$$\mathbf{z}[n] = \mathbf{W}^H[n]\mathbf{y}[n]. \quad (9)$$

By combining both Equations (8) and (9), the output  $\mathbf{z}[n]$  can be rewritten as a linear combination of the desired signal

$$\mathbf{z}[n] = \boldsymbol{\Gamma}[n]\mathbf{d}[n] \quad (10)$$

where  $\boldsymbol{\Gamma}[n] = \mathbf{W}^H[n]\mathbf{A}$  represents the overall mixing/separating system. Sources are optimally recovered when the matrix  $\mathbf{W}[n]$  is selected such as every output extracts a different single source. This occurs when the matrix  $\mathbf{G}[n]$  has the form

$$\boldsymbol{\Gamma}[n] = \mathbf{D}[n]\mathbf{P}[n] \quad (11)$$

where  $\mathbf{D}[n]$  is a diagonal invertible matrix and  $\mathbf{P}$  is a permutation matrix.

### A. Supervised Approach

A way to estimate the channel matrix,  $\mathbf{H}$ , consists on minimizing the MSE between the outputs  $\mathbf{y}[n]$  and the code signals  $\mathbf{x}[n]$ . Mathematically, using equation (6), the cost function is written as

$$\begin{aligned} J_{\text{MSE}} &= \sum_{i=1}^{N_B} \text{E} [|z_i[n] - d_i[n]|^2] \\ &= \text{E} [\text{tr} ((\mathbf{W}^H[n]\mathbf{y}[n] - \mathbf{d}[n])(\mathbf{W}^H[n]\mathbf{y}[n] - \mathbf{d}[n])^H)]. \end{aligned} \quad (12)$$

where the desired signals is obtained from the pilot symbols  $\mathbf{u}[n]$  by using  $\mathbf{d}(n) = \mathbf{F}\mathbf{u}(n)$ . A way to find the minima of this cost function consists in using a gradient algorithm that adapt the separating coefficients according with the gradient of this  $J_{\text{MSE}}$ , which is given by

$$\nabla_{\mathbf{W}} J_{\text{MSE}} = \text{E} [\mathbf{z}[n](\mathbf{W}^H[n]\mathbf{y}[n] - \mathbf{d}[n])^H]. \quad (13)$$

In general, the expectation included in  $\nabla_{\mathbf{W}} J_{\text{MSE}}[n]$  is unknown so it must be estimated from the available data. In particular, by considering only one sample, we obtain the *Least Mean Squares* (LMS) algorithm,

$$\mathbf{W}[n+1] = \mathbf{W}[n] - \mu \mathbf{y}[n](\mathbf{W}^H[n]\mathbf{y}[n] - \mathbf{d}[n])^H. \quad (14)$$

This algorithm is also called *delta rule* of *Widrow-Hopf* [9] in the context of Artificial Neural Networks [9]. It is easy to prove that the stationary points of this rule are

$$\mathbf{W}[n] = \mathbf{R}_z^{-1} \mathbf{R}_{z\mathbf{d}} \quad (15)$$

where  $\mathbf{R}_z = \text{E}[\mathbf{z}[n]\mathbf{z}^H[n]]$  is the autocorrelation of the observations and  $\mathbf{R}_{z\mathbf{d}} = \text{E}[\mathbf{z}[n]\mathbf{d}^H[n]]$  is the cross-correlation between the observations and the desired signals. In practice, the desired signal is considered as known only during a finite number of instants (pilot symbols) and the expectations are estimated by means of using samples averaging.

### B. Unsupervised Approach

The inclusion of pilot symbols reduces the system throughput (or equivalently, it reduces the system spectral efficiency) and wastes transmission energy because pilot sequences do not convey information. This limitation can be avoided by using Blind Source Separation (BSS) algorithms which simultaneously estimate the mixing matrix  $\mathbf{A}$  and the realizations of the source vector  $\mathbf{u}[n]$  from the corresponding realizations of the observed vector  $\mathbf{y}[n]$ .

One of the best known BSS algorithms has been approached by Bell and Sejnowski in [3]. Given an activation function  $h(\bullet)$ , the idea proposed by these authors is to obtain the weighted coefficients of a Neural Network,  $\mathbf{W}[n]$ , in order to maximize the mutual information between the outputs before the activation function,  $\mathbf{h}(\mathbf{z}[n]) = \mathbf{h}(\mathbf{W}^H[n]\mathbf{y}[n])$ , and its inputs  $\mathbf{y}[n]$ , which is given by

$$J_{\text{MI}}(\mathbf{W}[n]) = \ln(\det(\mathbf{W}^H[n])) + \sum_{i=1}^{N_B} \text{E}[\ln(h'_i(z_i[n]))] \quad (16)$$

where  $h_i$  is the  $i$ -th element of the vector  $\mathbf{h}(\mathbf{z}[n])$  and  $'$  denotes the first derivative. The maximum of this cost function can be obtained by means of using a gradient algorithm [3] or a relative gradient algorithm [10], [11]. Both approaches use the gradient of Equation (16) which is obtained as follows

$$\begin{aligned} \nabla_{\mathbf{W}} J_{\text{MI}} &= \nabla_{\mathbf{W}} (\ln(\det(\mathbf{W}^H[n]))) \\ &\quad + \nabla_{\mathbf{W}} \left( \sum_{i=1}^{N_B} \text{E}[\ln(h'_i(z_i[n]))] \right) \\ &= \frac{\text{adj}(\mathbf{W}^H[n])}{\det(\mathbf{W}^H[n])} - \text{E}[\mathbf{y}[n]\mathbf{g}^H(\mathbf{z}[n])] \\ &= \mathbf{W}^{-H}[n] - \text{E}[\mathbf{y}[n]\mathbf{g}^H(\mathbf{z}[n])] \end{aligned} \quad (17)$$

where  $\text{adj}(\bullet)$  is the adjunct of a matrix and  $\mathbf{g}(\mathbf{z}[n]) = [-h''_1(z_1[n])/h'_1(z_1[n]), \dots, -h''_N(z_N[n])/h'_N(z_N[n])]^T$  depends on the activation function. Finally, when the expectation is estimated by means of using only one sample

as before, we obtain the learning rules named *gradient algorithm* and *relative gradient algorithm* and given by

- *Gradient Algorithm:*

$$\begin{aligned}\mathbf{W}[n+1] &= \mathbf{W}[n] + \mu (\mathbf{W}^{-\text{H}}[n] - \mathbf{y}[n] \mathbf{g}^{\text{H}}(\mathbf{z}[n])) \\ &= \mathbf{W}[n] - \mu (\mathbf{y}[n] \mathbf{g}^{\text{H}}(\mathbf{z}[n]) - \mathbf{W}^{-\text{H}}[n])\end{aligned}\quad (18)$$

- *Relative Gradient Algorithm (Infomax):*

$$\begin{aligned}\mathbf{W}[n+1] &= \mathbf{W}[n] + \mu \mathbf{W}[n] \mathbf{W}^{\text{H}}[n] \\ &\quad \cdot (\mathbf{y}[n] \mathbf{g}^{\text{H}}(\mathbf{z}[n]) - \mathbf{W}^{-\text{H}}[n]) \\ &= \mathbf{W}[n] + \mu \mathbf{W}[n] (\mathbf{z}[n] \mathbf{g}^{\text{H}}(\mathbf{z}[n]) - \mathbf{I}).\end{aligned}\quad (19)$$

The expression in Equation (16) admits an interesting interpretation by means of the use of the non-linear function  $g(z) = z^*(1 - |z|^2)$ . In this case, Castedo et al. [12] have shown that the Bell and Sejnowski rules are equivalent to the *Constant Modulus Algorithm* (CMA) proposed by Godard in [13].

#### IV. HYBRID APPROACH

One of advantage of adaptive unsupervised (or blind) algorithms is their capacity of tracking low variations in the channel. On the contrary, supervised solutions provides a fast channel estimation for low or high variations at the cost of using pilot symbols. In this section, we combine this two parading in order to obtain an performance near to supervised approaches, but using lower number of pilot symbols. Figure 3 shows a simplified block diagram for this hybrid approach.

We will denote by  $\mathbf{W}_u[n]$  and  $\mathbf{W}_s[n]$  the matrices of coefficients for the unsupervised and the supervised module, respectively. We start with an initial estimation of the channel matrix obtained using the Widrow-Hopf solution given by Equation (15). This estimation is used at the transmitter in order to obtain the optimum coding matrix  $\mathbf{F}$  and at the receiver with the goal of initializing the unsupervised algorithm to  $\mathbf{W}_u[n] = (\mathbf{F}\mathbf{H})^{-\text{H}}$ .

When the “decision module” determines that the channel has not suffered from an important variation, the matrix  $\mathbf{W}_u[n]$  is adapted and the data symbols  $\mathbf{u}[n]$  are recovered by means of using  $\mathbf{z}[n] = \mathbf{W}_u^{\text{H}}[n] \mathbf{y}[n]$ . On the contrary, when a considerable variation has occurred, the receiver sends an “alarm” to the transmitter by means of the feedback channel. At this instant, a pilot sequence must be sent by the transmitter. At the receiver, an supervised algorithm estimates the channel from the pilot symbols. In particular, we consider Widrow-Hopf solution of Equation (15) by considering that  $\mathbf{u}[n]$  are the coded signals at the linear precoder output. This solution provides us the channel matrix estimate. This estimation is sent to the transmitter in order to adapt the coding matrix. The receiver also computes the coding matrix  $\mathbf{F}[n]$  and the reference matrix  $\hat{\mathbf{H}}\mathbf{F}$ , and initializes the unsupervised algorithms such as  $\mathbf{W}_u[n] = \hat{\mathbf{H}}\mathbf{F}^{-1}$

#### A. Decision rule

The question now is how to determine the instant where the channel has suffered a considerable change. An interesting consequence of using a linear precoder is that the permutation indeterminacy (see Equation (11)) associated to unsupervised algorithms is avoided because of the initialization to  $\mathbf{W}_u[n] = (\mathbf{F}\mathbf{H})^{-\text{H}}$ . This means that the sources are recovered in the same order as were transmitted. Taking into account, Equation (11) implies that optimum separation matrix produces a diagonal matrix  $\mathbf{\Gamma}[n]$  and, therefore, the mismatch of  $\mathbf{\Gamma}[n]$  with respect to a diagonal matrix allows us to measure the variations in the channel.

Although the channel matrix is unknown, we can use the estimation  $\hat{\mathbf{H}}\mathbf{F}$  computed by the supervised approach as a reference. This means that at each iteration we can compute  $\mathbf{\Gamma}[n] = \mathbf{W}_u^{\text{H}}[n] \hat{\mathbf{H}}\mathbf{F}$ . Subsequently, the difference with respect to a diagonal matrix can be obtained using the following “error” index

$$\text{Error}(n) = \sum_{i=1}^{N_B} \sum_{j=1, j \neq i}^{N_B} \left( \frac{|\gamma_{ij}[n]|^2}{|\gamma_{ii}[n]|^2} + \frac{|\gamma_{ji}[n]|^2}{|\gamma_{ii}[n]|^2} \right) \quad (20)$$

where  $\gamma_{ii}[n]$  denotes the  $i$ -th element of its diagonal. A way of decide when the channel has changed consists in comparing with some threshold  $t$ , i.e.

$$\text{Error}(n) > t \rightarrow \text{Use supervised approach} \quad (21)$$

The next section shows that the inclusion of this simple rule considerably improves the performance of both supervised and unsupervised approaches. Further work deals with designing a more “intelligent” decision criterion taking into account information about the environment.

#### V. EXPERIMENTAL RESULTS

In order to show the performance achieved with the proposed combined schemes, we present the results obtained for several computer were performed considering the transmission of 8,000 pixels of the image “cameraman” (in tif format with 256 gray levels) using a QPSK and an MIMO system with four transmit and receive antennas. The channel matrix is updated each 2,000 symbols using the following model

$$\mathbf{H} = (1 - \alpha)\mathbf{H} + \alpha\mathbf{H}_{\text{new}} \quad (22)$$

where  $\mathbf{H}_{\text{new}}$  is a  $4 \times 4$  matrix randomly generated according to a Gaussian distribution. The SNR has been stated to 20 dB.

In order to illustrate the form in which our system works using the rule (21), Figure 4 shows the result of evaluating the error measure given in equation (20) given a channel updating parameter  $\alpha = 0.1$  and two values of the threshold:  $t = 0.2$  and  $t = 0.5$ . We can see that the method detects the changes produced in the channel at instants 2,000, 4,000 and 6,000. The difference in using the two parameters is the delay needed to detect the variation.

We have compared the performance of the following schemes:



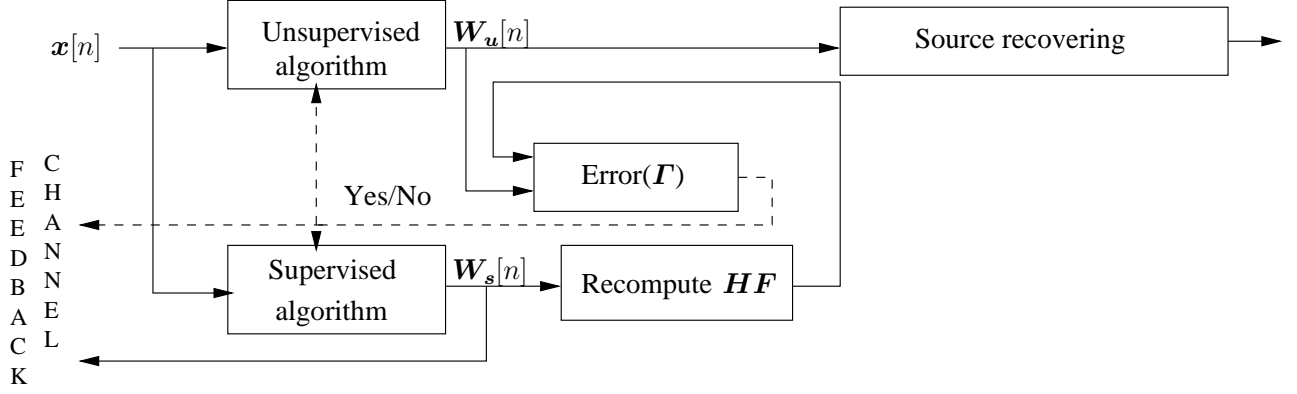


Fig. 3. Block Diagram for the Combined Approach.

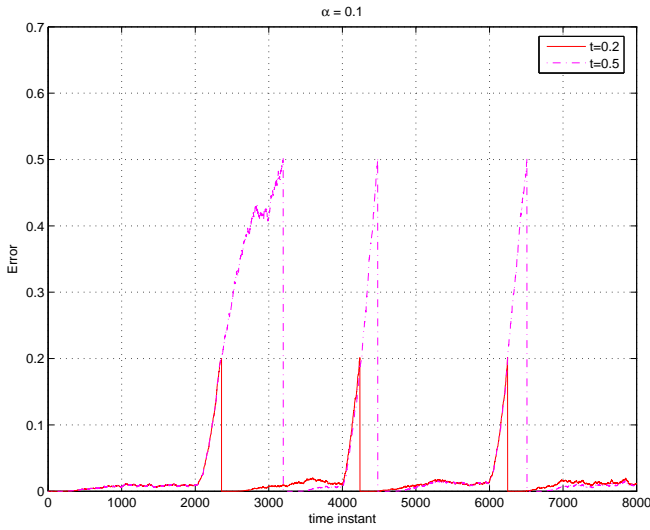


Fig. 4. Evolution of the error index.

- A supervised scheme where the Widrow-Hopf solution (14) is computed using 200 pilot symbols transmitted each 2,000 symbols.
- The unsupervised algorithm (Infomax) initialized to the Widrow-Hopf solution. The step size has been fixed to  $\mu = 0.001$ .
- The hybrid approach using the decision rule in equation (21) with  $t = 0.7$ . The Widrow-Hopf solution with 200 pilot symbols is used to estimate the channel matrix when the error is bigger than this threshold.

In the first experiment, we have considered that the channel is updated each  $N_B = 2,000$  symbols. This is the best situation for the supervised approach because implies a perfect synchronization. Figure 5 shows the performance obtained for the three approaches. The results have been obtained by averaging 100 independent realizations. Note that the considerable improvement in the BER obtained for the hybrid approach respect to the unsupervised approach. It is also apparent that when  $\alpha < 0.2$ , the hybrid approach achieves the same BER than the Wiener-Hopf solution with

less pilot symbols (or, equivalently, number of updating).

In the second experiment, the number of symbols in which the channel remains constant is a value between 2,000 and 3,000, which is randomly generated for each realizations. Figure 6 plots the BER and the number of updating. Note that the BER of the hybrid approach overcomes to the BER obtained with the other approaches, even the supervised solution, with a reduced number of updating.

Comparing Figure 5 and Figure 6, we observe that the BER of the hybrid system is invariant to the number of symbols in which the channel remains constant and it is obtained a considerable reduction in the number of of times in which the supervised approach is needed. This reduction is due to the fact that the channel remains constant more than 2,000 symbols. Remark also the important loss in quality of the supervised approach due to the outmatching between the channel updating instant and the instant when the pilot symbols are transmitted.

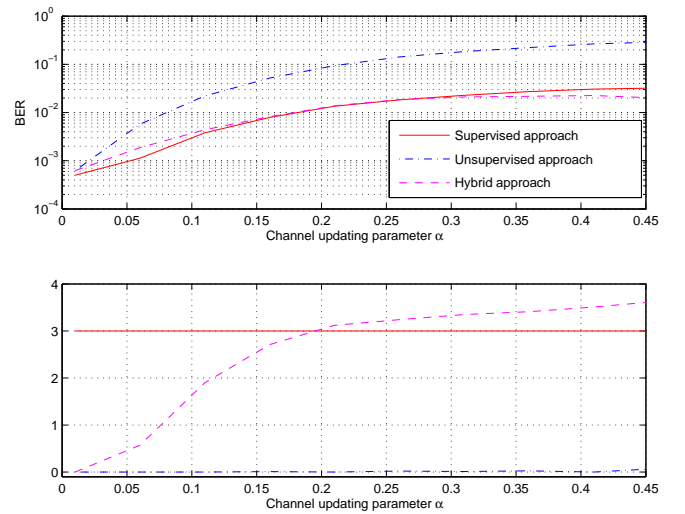


Fig. 5. Performance when the channel remains constant during a fixed number of symbols  $N_B = 2,000$ .

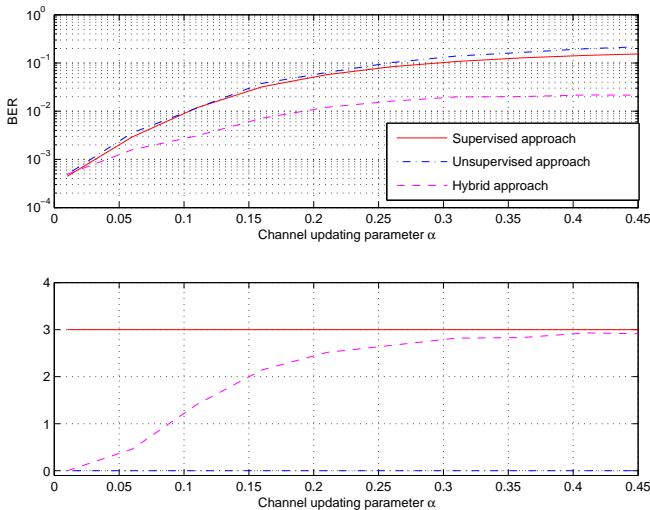


Fig. 6. Performance when the channel remains constant during a randomly generated number of symbols.

## VI. CONCLUSIONS

In order to reduce the overhead due to the transmission of pilots symbols in FDD-LP systems, we have proposed to combine supervised and unsupervised algorithms. The algorithm selection is done by using a simple decision rule that allows to determine the case when the channel has suffered a considerable variation. This information is sent to the transmitter using the feedback channel. The experiment results show that the hybrid approach is an attractive solution because it provides an adequate BER with a reduced number of pilot symbols. However, thinking in the transmission of an image with good quality, the hybrid approach is adequate when  $\alpha < 0.1$ .

## ACKNOWLEDGMENT

This work was supported by Xunta de Galicia, Ministerio de Educación y Ciencia, Ministerio de Ciencia e Innovación of Spain and FEDER funds of the European Union under grants number 09TIC008105PR, TEC2007-68020-C04-01, CSD2008-00010 and TIN2009-05736-E/TIN.

## REFERENCES

- [1] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. John Wiley & Sons, 2002.
- [2] M. Joham, *Optimization of Linear and Nonlinear Transmit Signal Processing*. PhD dissertation. Munich University of Technology, 2004.
- [3] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, November 1995.
- [4] R. Hunger, M. Joham, and W. Utschick, "Extension of linear and nonlinear transmit filters for decentralized receivers," in *European Wireless 2005*, April 2005, pp. 40–46, vol. 1.
- [5] M. Joham, K. Kusume, M. H. Gzara, W. Utschick, and J. A. Nossek, "Transmit Wiener Filter for the Downlink of TDD DS-CDMA Systems," in *Proc. ISSSTA*, vol. 1, September 2002, pp. 9–13.
- [6] R. L. Choi and R. D. Murch, "New Transmit Schemes and Simplified Receiver for MIMO Wireless Communication Systems," *IEEE Transactions on Wireless Communications*, vol. 2, no. 6, pp. 1217–1230, November 2003.
- [7] H. R. Karimi, M. Sandell, and J. Salz, "Comparison between Transmitter and Receiver Array Processing to Achieve Interference Nulling and Diversity," in *Proc. PIMRC*, vol. 3, September 1999, pp. 997–1001.
- [8] J. A. Nossek, M. Joham, and W. Utschick, "Transmit Processing in MIMO Wireless Systems," in *Proc. of the 6th IEEE Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Communication*, May/June 2004, pp. I–18 – I–23, Shanghai, China.
- [9] S. Haykin, *Neural Networks A Comprehensive Foundation*. Macmillan College Publishing Company, New York, 1994.
- [10] S.-I. Amari, "Gradient learning in structured parameter spaces: Adaptive blind separation of signal sources," in *Proc. WCNN'96*. San Diego, 1996, pp. 951–956.
- [11] C. Mejuto and L. Castedo, "A neural network approach to blind source separation," in *Proc. Neural Networks for Signal Processing VII*. Florida, USA, September 1997, pp. 486–595.
- [12] L. Castedo and O. Macchi, "Maximizing the information transfer for adaptive unsupervised source separation," in *Proc. SPAWC'97*. Paris, France, April 1997, pp. 65–68.
- [13] D. N. Godard, "Self-recovering equalization carrier tracking in twodimensional data communications systems," *IEEE Transactions on Communications*, vol. COM-28, pp. 1867–1875, 1980.