

Hybrid Channel Estimation Strategy for MIMO systems with Differential Feedback Equalizer

Héctor J. Pérez-Iglesias, Adriana Dapena,
Paula M. Castro, José A. García-Naya

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

Abstract. We propose to combine supervised and unsupervised algorithms in order to improve the performance of multiple-input multiple-output digital communication systems which make use of differential-feedback equalizers at the receiver. The basic idea is to avoid the periodical transmission of pilot symbols by using a simple criterion to determine the time instants when the performance obtained with an unsupervised algorithm is poor or, equivalently, those instants when pilot symbols must be transmitted. Simulation results show how the novel approach provides an adequate BER with a low overhead produced by the transmission of pilot symbols.

1 Introduction

Decision-Feedback Equalizer (DFE) was initially proposed to reduce the effect of multiple-delayed copies of a signal transmitted over *Single-Input Single-Output* (SISO) systems [1], i.e., to equalize the channel. It consists in 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 a feedback from past decisions to cancel the interference of the symbols already detected. Extensions of this idea to *Multiple-Input Multiple-Output* (MIMO) systems have been proposed by several authors (see, for instance, [2–4]). In fact, DFE has been included in several standards like *Digital Terrestrial Multimedia Broadcast* (DTMB) and 802.16 (WiMAX).

In this work, we propose a novel way to combine supervised and unsupervised algorithms in order to improve the performance of MIMO DFE systems. The basic idea is to use an unsupervised algorithm joined with a simple detection criterion to determine the instant where the channel has suffered a considerable variation. When this event occurs, the receiver sends an “alarm” to the transmitter by means of a feedback or reverse channel indicating that some pilot symbols must be transmitted. Such reverse channel is actually implemented in most of the standards [5]. In the rest of time, an unsupervised adaptive algorithm is used to track channel variations. Simulation results show that the proposed scheme leads to good performance in terms of *Bit Error Rate* (BER) and avoids periodical transmissions of pilot symbols.

2 System Model

We consider a MIMO system with N_t transmit antennas and N_r receive antennas. The data symbols $\mathbf{u} = [u_1, \dots, u_{N_t}]^T$ are transmitted to the different receive antennas such as, for flat fading channels, the received signals have the following form

$$y_j(t) = \sum_{i=1}^{N_t} h_{ji}(t)u_i(t) + n_j(t).$$

In a compact form, we can express the vector of received signals, $\mathbf{y}(t)$, as follows

$$\mathbf{y}(t) = \mathbf{H}(t)\mathbf{u}(t) + \mathbf{n}(t) \quad (1)$$

where $\mathbf{u}(t)$ is the source vector, $\mathbf{n}(t)$ is the noise vector, and $\mathbf{H}(t)$ contains the channel coefficients $h_{i,j}(t)$ from the i -th transmit antenna to the j -th receive antenna in this way

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

In general, if we let $f[n] = f(nT_s + \Delta)$ denote samples of $f(t)$ every T_s seconds with Δ 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

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

where $n = 0, 1, 2, \dots$ corresponds to samples spaced with T_s and q denotes the slot time. For brevity, we omit the slot index q in the sequel. The channel remains unchanged during a block of N_B symbols, i.e, over the data frame. We also assume that the transmit sources are independent and identically distributed and have unit power, i.e., $\mathbf{C}_{\mathbf{u}} = E[\mathbf{u}[n]\mathbf{u}^H[n]] = \mathbf{I}$, where \mathbf{I}_{N_t} denotes the identity matrix.

Note that the discrete time model in Equation (2) is equivalent to the continuous time model in Equation (1) only if the *Inter-Symbol Interference* (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 samples by means of interpolation. This channel model is known as time-varying flat block fading channel and this assumption is made in the sequel.

Channel equalization is often used at the receiver to combat the distortion introduced by the channel. As can be seen in Figure 1, the estimated signal $\hat{\mathbf{u}}[n]$ can be expressed as

$$\hat{\mathbf{u}}[n] = \mathbf{F}\mathbf{y}[n] + (\mathbf{I}_{N_t} - \mathbf{B})\tilde{\mathbf{u}}[n] \quad (3)$$

where \mathbf{F} , \mathbf{B} , and \mathbf{I}_{N_t} represent the feedforward filter, the feedback filter, and the identity matrix, respectively. The vector $\tilde{\mathbf{u}}[n]$ denotes the quantized symbols. Since the estimated signals can be recovered in a different order than the

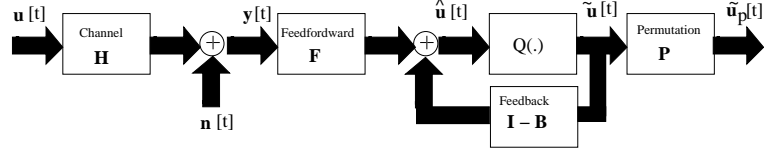


Fig. 1. MIMO system with DFE.

transmitted ones, we introduce the permutation matrix \mathbf{P} such as

$$\hat{\mathbf{u}}[n] = \mathbf{F}\mathbf{y}[n] + (\mathbf{I}_{N_t} - \mathbf{B})\mathbf{P}\tilde{\mathbf{u}}_p[n].$$

The DFE feedforward and feedback filters can be found by minimizing the *Mean Square Error* (MSE) between a permutation of the transmit signals and their respective estimates by restricting the feedback filter \mathbf{B} to being lower triangular [2, 3], i.e.

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

The procedure described in [2] to find the matrices \mathbf{P} , \mathbf{B} , and \mathbf{F} can be summarized as follows:

- Step 1: Compute $\Phi = (\mathbf{H}^H \mathbf{C}_n^{-1} \mathbf{H} + \mathbf{C}_u^{-1})^{-1}$
 Step 2: Initialize $\mathbf{P} = \mathbf{I}_{N_t}$ and $\mathbf{D} = \mathbf{O}_{N_t \times N_t}$
 Step 3: **for** $i = 1, \dots, N_t$
 • Find $q = \arg \min_{q'=1, \dots, N_t} \Phi(q', q')$
 • Set $\mathbf{P}_i = \mathbf{I}_{N_t}$ whose i -th and q -th rows are exchanged
 • Compute $\mathbf{P} = \mathbf{P}_i \mathbf{P}$ and $\Phi = \mathbf{P}_i \Phi \mathbf{P}_i^T$
 • Let $\mathbf{D}(i, i) = \Phi(i, i)$
 • Compute $\Phi(i : N_t, i) = \Phi(i : N_t, i) / \mathbf{D}(i, i)$
 • Compute $\Phi(i+1 : N_t, i+1 : N_t) = \Phi(i+1 : N_t, i+1 : N_t) - \Phi(i+1 : N_t, i) \Phi(i+1 : N_t, i)^H \mathbf{D}(i, i)$
 • Let \mathbf{L} be the lower triangular part of Φ
 Step 4: Compute $\mathbf{B} = \mathbf{L}^{-1}$ and $\mathbf{F} = \mathbf{D} \mathbf{L}^H \mathbf{P} \mathbf{H}^H \mathbf{C}_n^{-1}$

The *for* loop is used to choose the optimum order, so the data stream corresponding to the minimum MSE entry of the MSE matrix Φ is decoded first in order to minimize the effect derived from feeding back erroneous decisions.

In order to implement the DFE procedure described above (Steps 1 and 4), the receiver must acquire the channel matrix \mathbf{H} . Since the transmission model in Equation (2) corresponds to a linear combination of the transmit signals, we will consider a linear recovering system whose outputs are computed as follows

$$\mathbf{z}[n] = \mathbf{W}^H[n] \mathbf{y}[n] \quad (4)$$

where $\mathbf{W}[n]$ is a $N_r \times N_t$ matrix that can be found using different supervised and unsupervised algorithms.

The classical way of estimating \mathbf{H} uses pilot symbols periodically sent from the transmit to the receive antennas. An important family of unsupervised filtering algorithms arises from considering the minimization of the MSE between the outputs $\mathbf{z}[n]$ and the desired signals $\mathbf{u}[n]$ [6]. Mathematically, the cost function can be written as

$$J_{\text{MSE}} = \sum_{i=1}^{N_B} E [|z_i[n] - u_i[n]|^2] \quad (5)$$

$$= E [\text{tr} ((\mathbf{W}^H[n]\mathbf{y}[n] - \mathbf{u}[n])(\mathbf{W}^H[n]\mathbf{y}[n] - \mathbf{u}[n])^H)]. \quad (6)$$

In this case, the optimum separating matrix can be obtained by determining the points where the gradient of J vanishes, i.e.,

$$\nabla_{\mathbf{W}} J_{\text{MSE}} = 0 \Rightarrow \mathbf{W}_{\text{opt}} = \mathbf{C}_{\mathbf{y}}^{-1} \mathbf{C}_{\mathbf{y}\mathbf{u}} \quad (7)$$

where $\mathbf{C}_{\mathbf{y}} = E[\mathbf{y}[n]\mathbf{y}^H[n]]$ is the autocorrelation of the observations and $\mathbf{C}_{\mathbf{y}\mathbf{u}} = E[\mathbf{y}[n]\mathbf{u}^H[n]]$ is the cross-correlation between the observations and the desired signals.

The transmission of pilot symbols can be avoided by means of using *Blind Source Separation* (BSS) algorithms [7], which simultaneously estimate the channel matrix 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 [8]. 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 $\mathbf{y}[n]$. The learning rule of Infomax is given by

$$\mathbf{W}[n+1] = \mathbf{W}[n] + \mu \mathbf{W}[n] (\mathbf{z}[n]\mathbf{g}^H(\mathbf{z}[n]) - \mathbf{I}_{N_t}) \quad (8)$$

where $\mathbf{g}(\mathbf{z}[n]) = [-h'_1(z_1[n])/h''_1(z_1[n]), \dots, -h''_{N_t}(z_{N_t}[n])/h'_1(z_{N_t}[n])]^T$ depends on the activation function ($h'(\bullet)$ and $h''(\bullet)$ represent the first and the second derivative of $g(\bullet)$, respectively). The expression in Equation (8) admits an interesting interpretation by means of the use of the non-linear function $g(z) = z^*(1 - |z|^2)$. In this case, Castedo and Macchi in [9] have shown that the Bell and Sejnowski rule can be interpreted as a generalization of the *Constant Modulus Algorithm* (CMA).

3 Hybrid approach

We propose to combine the two methods above shown in order to obtain a performance close to the supervised approach, but by means of using a lower number of pilot symbols. Let $\mathbf{W}_u[n]$ and $\mathbf{W}_s[n]$ be the matrices of coefficients for the unsupervised and supervised modules, respectively. We start with an initial estimation of the channel matrix obtained using the supervised method

given by Equation (7), $\hat{\mathbf{H}} = \mathbf{W}_s^{-H}[n]$. 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 sample by sample the separating matrix using the rule (8) and the channel matrix needed by 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, an important variation on the channel. When this event occurs, the receiver sends an “alarm” to the transmitter. At this instant, a frame of pilot symbols must be sent by the transmitter. At the receiver, the supervised algorithm estimates the channel from the pilot symbols using Equation (7). This solution is used to initialize the unsupervised algorithm.

The important question is how to determine the instants where the unsupervised algorithm presents a poor performance. By combining both Equations (2) and (4), the output $\mathbf{z}[n]$ can be rewritten as a linear combination of the sources

$$\mathbf{z}[n] = \mathbf{\Gamma}[n]\mathbf{u}[n] + \mathbf{W}^H[n]\mathbf{n}[n] \quad (9)$$

where $\mathbf{\Gamma}[n] = \mathbf{W}^H[n]\mathbf{H}$ represents the overall mixing/separating system (or gain matrix). This means that each output contains a term corresponding to the desired source and another one due to the *MultiUser Interference* (MUI). It is interesting to note that the initialization of the unsupervised algorithm removes the permutation ambiguity inherent in this class of learning rules. Thus, each output will have the form

$$z_i[n] = \gamma_{ii}[n]u_i[n] + \sum_{j=1, i \neq j}^{N_t} \gamma_{ij}[n]u_j[n] + \mathbf{w}_i^H \mathbf{n}[n]. \quad (10)$$

By dividing this equation by $\gamma_{ii}[n]$ and considering that the noisy component is small compared to the other terms, we obtain that the power of each output is given by

$$\frac{E[|u_i[n]|^2]}{|\gamma_{ii}[n]|^2} = E[|u_i[n]|^2] + \sum_{j=1, i \neq j}^{N_t} \frac{|\gamma_{ij}[n]|^2}{|\gamma_{ii}[n]|^2} E[|u_j[n]|^2] = E[|u_i[n]|^2] + MUI_i$$

where the *MUI* term is implicitly defined. When *MUI* is high, the channel matrix estimate is poor. In that case, a pilot frame must be transmitted, i.e.,

$$MUI = \sum_{i=1}^{N_t} MUI_i = \sum_{i=1}^{N_t} \sum_{\substack{j=1 \\ i \neq j}}^{N_t} \frac{|\gamma_{ij}[n]|^2}{|\gamma_{ii}[n]|^2} > t \rightarrow \text{Send an “alarm”} \quad (11)$$

where t is a real positive number (threshold). 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 an initial estimation of the channel matrix obtained with the supervised approach. Obviously, a small value for t reduces the error but it increases the number of pilot symbols.

4 Experimental Results

In order to show the performance achieved with the proposed hybrid scheme, we present the results for several computer simulations performed considering that 4900 symbols are transmitted in frames of size $N_B = 100$ using a QPSK modulation. The system considers four transmit and receive antennas. The channel matrix is updated using the model $\mathbf{H} = (1 - \alpha)\mathbf{H} + \alpha\mathbf{H}_{\text{new}}$, where \mathbf{H}_{new} is a 4×4 matrix randomly generated according to a Gaussian distribution.

We have evaluated the performance of the following proposed schemes:

- *Perfect Channel State Information* (CSI) at the receiver, i.e., the DFE uses the real channel matrix.
- The supervised approach in Equation (7) computed using a frame of N_B pilot symbols transmitted each 10 frames.
- The generalized Infomax algorithm initialized to the matrix obtained with the supervised approach. The step-size parameter is $\mu = 0.001$ and $g(z) = z^*(1 - |z|^2)$.
- The hybrid approach using the generalized Infomax algorithm with different thresholds, step-size parameter given by $\mu = 0.001$, $g(z) = z^*(1 - |z|^2)$ and $t = 0.1$. A frame of N_B pilot symbols is used when the error is higher than the corresponding threshold.

The results have been obtained by averaging 1000 independent experiments. Pilot frames have not been considered to compute the Bit Error Rate.

In the first experiment we have considered that the channel remains constant during 10 frames. This assumption is an ideal situation for the supervised approach because it corresponds to the same instants where pilot frames are transmitted. Figure 2 (a) shows the BER and the number of transmit pilot symbols in terms of the channel updating parameter α for a SNR of 15 dB. Note the considerable improvement in BER obtained with the hybrid approach compared to the Infomax algorithm. Note also that the BER is close to that obtained with the supervised approach but now the number of pilots is considerable smaller.

Figure 3 (a) shows the BER and the number of pilot symbols for several values of SNR given $\alpha = 0.05$. Again, the hybrid approach overcomes the results obtained with the unsupervised algorithm and it achieves a BER close to the supervised approach with fewer pilot symbols. Note that the Infomax algorithm shows a floor effect at a SNR of 8 dB while this effect appears for the hybrid approach at about 12 dB where it is clear that the BER is considerable smaller.

In the second experiment, we have considered that the channel remains constant during a number of frames randomly generated in each computer simulation from the interval [10, 15]. Figure 2 (b) shows the performance at a SNR of 15 dB for different values of the updating channel parameter α . It can be seen that the supervised BER obtained with the hybrid approach matches the BER obtained with the supervised algorithm. Comparing with part (a), we observe that the number of pilot symbols is reduced. The same conclusion is obtained from Figure 3 (b), which corresponds to the performance for different SNRs considering $\alpha = 0.05$. Note that now the floor effect of the hybrid approach appears

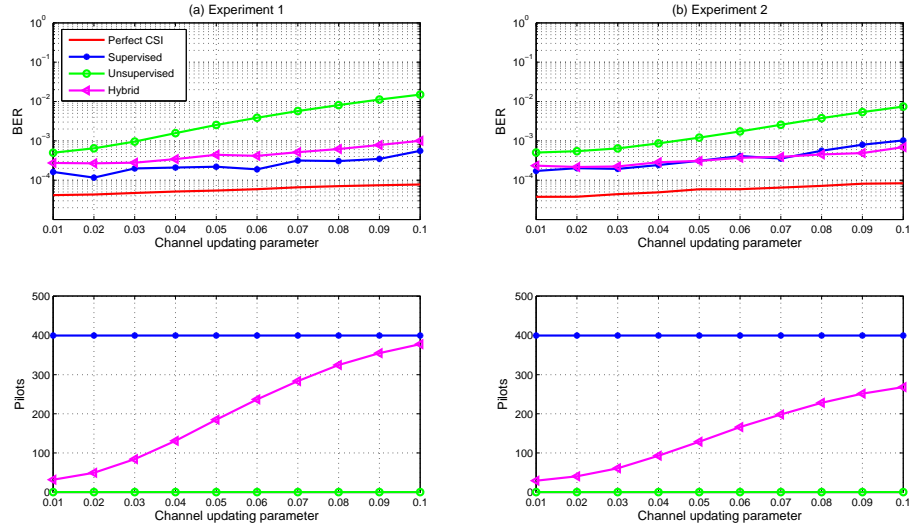


Fig. 2. BER and number of pilots in terms of the channel updating parameter for a SNR of 15 dB.

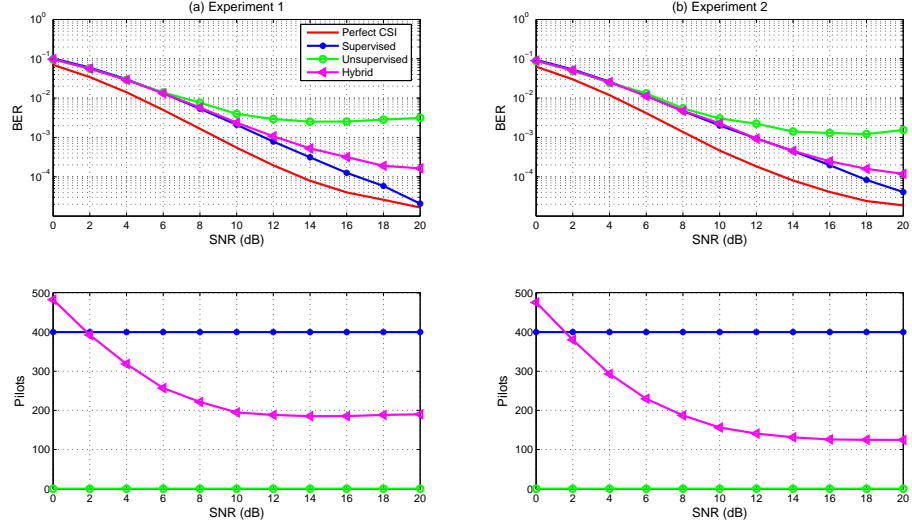


Fig. 3. BER and number of pilots in terms of SNR for $\alpha = 0.05$.

at a SNR of 15 dB corresponding to a BER of 2×10^{-4} . For this value only 130 pilot symbols are transmitted, which represents a considerable reduction in comparison to the 400 symbols needed by the supervised approach.

5 Conclusions

This paper deals with the utilization of supervised and unsupervised algorithms to estimate the channel matrix in systems with DFE at the receiver. Given a communication model where the channel is block flat fading, we have proposed a simple way to dynamically determine the instants when pilot symbols must be transmitted. Simulation results show that the novel approach provides an adequate BER with a low overhead produced by the transmission of these pilot symbols. The proposed strategy could be extended to other type of receive models.

Acknowledgment

This work was supported by Xunta de Galicia, 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.

References

1. Austin, M.E.: Decision feedback equalization for digital communication over dispersive channels. Technical report 437. Lincoln Laboratory (1967)
2. Kusume, K., Joham, M., Utschick, W.: MMSE block decision-feedback equalizer for spatial multiplexing with reduced complexity. *Proc. IEEE Global Telecommunications Conference* **4** (2004) 2540–2544
3. Fischer, R.F.H.: *Precoding and Signal Shaping for Digital Transmission*. John Wiley & Sons (2002)
4. Joham, M.: *Optimization of Linear and Nonlinear Transmit Signal Processing*. PhD dissertation. Munich University of Technology (2004)
5. Philips: Comparison Between MU-MISO Codebook-based Channel Reporting Techniques for LTE Downlink. 3GPP TSG RAN WG1, Tech. Rep. R1-062483 (2006)
6. Haykin, S.: *Neural Networks A Comprehensive Foundation*. Macmillan College Publishing Company, New York (1994)
7. Comon, P., Jutten, C.: *Handbook of Blind Source Separation, Independent Component Analysis and Applications*. Academic Press (2010)
8. Bell, A., Sejnowski, T.: An information-maximization approach to blind separation and blind deconvolution. *Neural Computation* **7**(6) (1995) 1129–1159
9. Castedo, L., Macchi, O.: Maximizing the information transfer for adaptive unsupervised source separation. In: *Proc. SPAWC'97, Paris, France* (1997) 65–68