Channel Estimation Techniques for Linear Precoded Systems: Supervised, Unsupervised, and Hybrid Approaches

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

Department of Electronics and Systems, University of A Coruña
Campus de Elviña 8, 15071 A Coruña, Spain

Abstract

Linear precoding is an attractive technique to combat interference in multiple-input multiple-output systems because it reduces costs and power consumption at the receiver end. Most of the frequency division duplex systems with linear precoding acquire the channel state information at the receiver side by using supervised algorithms. Such methods make use of pilot symbols periodically provided by the transmitter. In a later step, this channel state information is sent to the transmitter side through a low-cost feedback channel. Thus, the available channel information allows the transmitter to adapt signals to the channel conditions. Given that pilot symbols do not convey user data, the inclusion of such symbols penalizes the throughput, the spectral efficiency, and the transmission energy consumption of the system. In this work, we propose to mitigate the above-mentioned limitations by combining both supervised and unsupervised algorithms to acquire the channel state information needed by the transmitter. The key idea consists of introducing a simple criterion to determine whether the channel has suffered a significant variation which could require the transmission of pilot symbols. Otherwise, when small fluctuations happen, an unsupervised method is used to track these channel variations instead. This criterion will be evaluated by considering two types of strategies for the design of the linear precoders: Zero–Forcing (ZF) and Wiener criteria.

*Corresponding author. Phone: +34 98167000 ext. 1366, FAX: +34 981167160

Email addresses: `pcastro@udc.es` (Paula M. Castro), `jagarcia@udc.es` (José A. García-Naya), `adriana@udc.es` (Adriana Dapena), `dani@udc.es` (Daniel Iglesia)

Keywords:
Linear Precoding, MIMO, Blind Source Separation, Learning Rules.

1. Introduction

Linear precoding is a powerful method to reduce interferences in Multiple-Input Multiple-Output (MIMO) systems since it simplifies the receiver equipment. The design of linear precoding schemes has been widely studied for the ideal case in which the Channel State Information (CSI) is perfectly known at the transmitter (TX) side. Although the zero–forcing criterion for the filters design is very intuitive since the transmitter has not knowledge about channel noise [17, 19, 24, 26], its drawbacks about amplifying noise lead to better choices for the precoder, such that Wiener linear precoder [9, 17, 18, 20, 24].

In practice, most of communication systems, including Frequency Division Duplex (FDD) systems, the TX cannot obtain the CSI from the received signals —even under the assumption of perfect calibration— since down– and up-channels are not reciprocal. Thus, the CSI has to be estimated at the receiver (RX) side and transmitted back through a limited feedback channel. As a rule, current wireless communications standards include pilot symbols in the definition of the transmitted signals.

The best known and most widely used approach to acquire the CSI at the RX side is the so-called Least Squares (LS) criterion. The LS-based channel estimation consists of minimizing the Mean Square Error (MSE) between the transmitted pilot symbols and the observed ones. Such pilots are periodically sent by the TX which gets round whether the RX really needs or not such symbols to track channel variations. This situation produces a strong degradation in terms of throughput, spectral efficiency, and transmit energy.

The so-called *unsupervised techniques* are able to estimate the channel coefficients directly from the observations, without requiring pilot symbols. They only assume that the transmitted signals are statistically independent [25]. Unfortunately, such unsupervised approaches —also known as Blind Source Separation (BSS) techniques [10, 25]— exhibit three major drawbacks:

> A poor performance in the event of significant channel variations. This limitation can be mitigated by using a supervised approach to re-estimate the channel.

Two ambiguities: permutation and gain. This means that the transmitted signals could be recovered with a different ordering and with a distinct scale factor.

A considerable number of observations (received symbols) is required to ensure the convergence to the channel estimate. This limitation can be avoided by initializing the BSS algorithm to a valid channel estimate obtained, for example, by using a supervised method.

Most of current wireless communications standards make use of feedback channels (usually limited in terms of throughput) connecting the RX and the TX sides of the communications link to periodically send channel state information from the RX to the TX. 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 or not to send pilot symbols and, consequently, reduce the above mentioned penalties introduced by the use of pilot symbols. This idea constitutes the main motivation of our work. We will take advantage of the feedback channel to inform the TX when the channel has suffered a significant variation and pilot symbols are thus required. Othervise, we use unsupervised approaches instead of supervised ones to track the small fluctuations of the wireless channel.

Therefore, in this work we propose to use a hybrid channel estimation setup based on combining both unsupervised and supervised approaches. Such algorithms have recently been used in image classification [21] and beamforming [4], for example. However, the way of combining these two paradigms is an open issue for wireless systems, including the case of precoding systems. In fact, the first hybrid schemes for linear precoding have been proposed by the authors in [8, 11]. Both works evaluate the performance achieved with the proposed hybrid setups using a Wiener criterion for the filter design. Compared to those initial works, in this article we derive a less complex decision criterion to be utilized by the hybrid algorithm, without any penalties on the final performance. The Bit Error Rate (BER) performance will be evaluated considering ZF– and Wiener–precoders, which let us compare the two most widely used designs for transmit linear processing.

The proposed procedure is the following. At the beginning, the RX uses a supervised algorithm to quickly obtain a valid channel estimate. Then, the RX switches to an unsupervised adaptive method for tracking the small channel fluctuations without requiring the transmission of any pilot symbols.

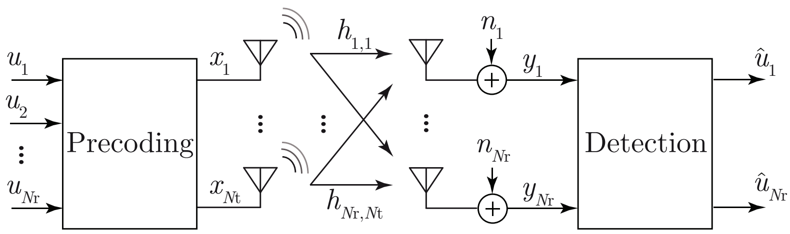

Figure 1: System with precoding over a flat MIMO channel.

When the channel suffers a significant variation, the RX commutes to the supervised working, instructs the TX to send pilot symbols, and re-estimates the channel using the supervised method. Finally, the RX switches again to run the unsupervised algorithm until a new significant channel change is observed. By making use of computer simulations, we will show how the proposed scheme leads to good performance in terms of Bit Error Ratio (BER) avoiding periodical transmission of pilot symbols. From the results, it is clear that the hybrid schemes for channel estimation could be implemented in the recent wireless communications standards (like Worldwide Interoperability for Microwave Access (WiMAX) or Long Term Evolution (LTE) standards), where precoding setups are included.

This work is organized as follows. Section 2 describes signal and channel models. Section 3 reviews supervised and unsupervised algorithms used for channel estimation, which are utilized in Section 4 for the proposed novel hybrid scheme. Illustrative omputer simulations results are presented in Section 5 and some concluding remarks are stated in Section 6.

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. We use $\mathrm{E}[\cdot]$, $\mathrm{tr}(\cdot)$, $(\cdot)^*$, $(\cdot)^\mathrm{T}$, $(\cdot)^\mathrm{H}$, $\det(\cdot)$ and $\|\cdot\|_2$, 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 $\boldsymbol{x}$ is denoted by x_i.

2. System Model

We consider a MIMO system with N_t TX antennas and N_r RX antennas, as plotted in Figure 1. For simplicity reasons, we consider $N_\mathrm{t} = N_\mathrm{r} = N$ Tx and RX antennas. The precoder generates the TX signal $\boldsymbol{x}$ from all data symbols $\boldsymbol{u} = [u_1, \ldots, u_N]^\mathrm{T}$ corresponding to the different RX antennas $1, \ldots, N$. We denote the equivalent lowpass channel impulse response between the j–th TX antenna and the i–th RX antenna as $h_{i,j}(\tau, t)$. Thus, the random time-varying channel is characterized by the $N \times N$ matrix $\boldsymbol{H}(\tau, t)$ defined as

$$\boldsymbol{H}(\tau, t) = \begin{pmatrix} h_{1,1}(\tau, t) & h_{1,2}(\tau, t) & \cdots & h_{1,N}(\tau, t) \\ h_{2,1}(\tau, t) & h_{2,2}(\tau, t) & \cdots & h_{2,N}(\tau, t) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N,1}(\tau, t) & h_{N,2}(\tau, t) & \cdots & h_{N,N}(\tau, t) \end{pmatrix}$$

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

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

where $*$ denotes the convolution between the channel and the transmitted signal and $\eta_j(t)$ is the additive noise. In matrix notation, this equation can be rewritten as

$$\boldsymbol{y}(t) = \boldsymbol{H}(\tau, t) * \boldsymbol{x}(t) + \boldsymbol{\eta}(t), \tag{1}$$

where $\boldsymbol{x}(t) = [x_1(t), \ldots, x_N(t)]^{\mathrm{T}} \in \mathbb{C}^N$, $\boldsymbol{y}(t) = [y_1(t), \ldots, y_N(t)]^{\mathrm{T}} \in \mathbb{C}^N$ and $\boldsymbol{\eta}(t) = [\eta_1(t), \ldots, \eta_N(t)]^{\mathrm{T}} \in \mathbb{C}^N$. For flat fading channels, the channel matrix $\boldsymbol{H}(\tau, t)$ is transformed into the matrix $\mathbf{H}(t)$ given by

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

and the received signal is now

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

which can be expressed in matrix form as

$$\boldsymbol{y}(t) = \boldsymbol{H}(t)\boldsymbol{x}(t) + \boldsymbol{\eta}(t). \tag{2}$$

Note that the convolution appeared in Equation(1) is transformed into a product since the channel is formed by only one tap.

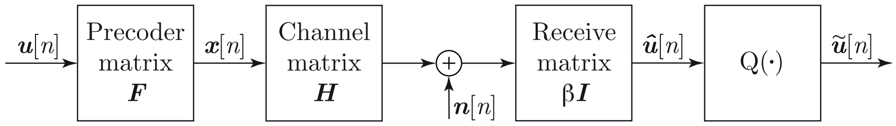

Figure 2: MIMO system with linear precoding.

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

$$\boldsymbol{y}[n] = \boldsymbol{H}[n]\boldsymbol{x}[n] + \boldsymbol{\eta}(n), \tag{3}$$

where $n = 0, 1, 2, \ldots$ corresponds to the sample index and the samples are spaced T_{s} (in seconds). Given that the channel remains constant during several frames of N_{B} symbols, we henceforth use $\boldsymbol{H}$ instead of $\boldsymbol{H}[n]$. Note that this discrete-time model is equivalent to the continuous-time model in Equation (2) only if 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. In the rest of this work, we assume this channel model, known as time-varying flat block fading channel.

2.1. Linear Precoding Design

The equalization task can be performed at the TX and thus the channel is pre-equalized or *precoded* before the transmission with the goal of simplifying the requirements at the RX. Such an operation is only possible when a centralized TX is employed (e.g. the base-station of the downlink of a cellular system). We assume hereinafter that the RX filter is an identity matrix (multiplied by a scalar β, with $\beta \in \mathbb{C}$), which allows the utilization of decentralized RX (see, for instance, [16]). Clearly, the restriction that all receivers apply the same scalar weight β is not necessary for decentralized receivers, but it ensures closed–form solutions for the filters design.

The goal is to find the optimum TX filter $\boldsymbol{F} \in \mathbb{C}^{N\times N}$ and the RX filter $\boldsymbol{G} = \beta\boldsymbol{I} \in \mathbb{C}^{N\times N}$. The resulting communications system is shown in Figure 2. It can be seen from the figure how the data symbols $\boldsymbol{u}[n]$ are passed through the transmit filter $\boldsymbol{F}$ to form the transmitted signal $\boldsymbol{x}[n] = \boldsymbol{F}\boldsymbol{u}[n] \in \mathbb{C}^{N}$. Note that the constraint for the transmitted energy must be fulfilled, i.e.

$$\mathrm{E}\left[\|\boldsymbol{x}[n]\|_2^2\right] = \mathrm{tr}\left(\boldsymbol{F}\boldsymbol{C}_{\boldsymbol{u}}\boldsymbol{F}^{\mathrm{H}}\right) \leq E_{\mathrm{tx}},$$

where E_{tx} is the fixed total transmitted energy. The cross–correlation of the uncoded symbols is represented by the matrix $\boldsymbol{C}_{\boldsymbol{u}} = E[\boldsymbol{u}[n]\boldsymbol{u}^H[n]]$. The received signal is thus given by

$$\boldsymbol{y}[n] = \boldsymbol{H}\boldsymbol{F}\boldsymbol{u}[n] + \boldsymbol{\eta}[n], \tag{4}$$

where $\boldsymbol{y}[n] \in \mathbb{C}^N$, $\boldsymbol{H} \in \mathbb{C}^{N\times N}$, and $\boldsymbol{\eta}[n] \in \mathbb{C}^N$ is the Additive White Gaussian Noise (AWGN).

After multiplying by the receive gain β, we get the estimated symbols

$$\hat{\boldsymbol{u}}[n] = \beta\boldsymbol{H}\boldsymbol{F}\boldsymbol{u}[n] + g\boldsymbol{\eta}[n], \tag{5}$$

where $\hat{\boldsymbol{u}}[n] \in \mathbb{C}^N$.

The most widely used optimizations for linear precoding are performed according to the *Zero–Forcing* (ZF) and Wiener criteria, in a similar way to the known respective criteria for the receive processing schemes [19, 24]. We focus on the standard approaches of MSE minimization with and without the zero-forcing constraint assuming a constraint of the total average transmit energy [26].

2.1.1. Linear Zero–Forcing Precoding

The most intuitive approach for precoding design is to apply the zero–forcing optimization since the transmitter has no influence on the noise at the receiver to remove interferences. Therefore, the transmit zero-forcing filter eliminates global interference at the output of the receive filter (given by $\boldsymbol{G} = \beta\boldsymbol{I}$) by means of forcing the chain formed by the precoder filter $\boldsymbol{F}$, the channel $\boldsymbol{H}$, and the receive filter $\boldsymbol{G} = \beta\boldsymbol{I}$, to be an identity matrix. This criterion minimizes the following MSE taking into account the total transmit energy constraint [17, 19, 24, 26]

$$\begin{aligned} \{\boldsymbol{F}_{\text{ZF}}, \beta_{\text{ZF}}\} = &\underset{\{\boldsymbol{F},\beta\}}{\operatorname{argmin}} \, \mathrm{E}\left[\|\boldsymbol{u}[n] - \hat{\boldsymbol{u}}[n]\|_2^2\right] \\ &\text{s.t.: } \beta\boldsymbol{H}\boldsymbol{F} = \mathbf{I} \quad \text{and} \quad \operatorname{tr}(\boldsymbol{F}\boldsymbol{C}_{\boldsymbol{u}}\boldsymbol{F}^{\mathrm{H}}) \le E_{\text{tx}}, \end{aligned} \tag{6}$$

where E_{tx} is the total transmitted energy.

As it can be seen in [17], the solution for the linear precoder designed using the ZF criterion is as follows

$$\begin{aligned} \boldsymbol{F}_{\text{ZF}} &= \beta_{\text{ZF}}^{-1}\boldsymbol{H}^H(\boldsymbol{H}\boldsymbol{H}^H)^{-1}, \\ \beta_{\text{ZF}} &= \sqrt{\frac{\operatorname{tr}((\boldsymbol{H}\boldsymbol{H}^H)^{-1}\boldsymbol{C}_{\boldsymbol{u}})}{E_{\text{tx}}}}. \end{aligned} \tag{7}$$

2.1.2. Linear Wiener Precoding

Although *Wiener Filtering* (WF) for precoding has been dealt with by only few authors [18] in comparison with other criteria for precoding, it is a very powerful transmit optimization that minimizes the MSE with a transmit energy constraint [9, 17, 20, 24], i.e.

$$\{\boldsymbol{F}_{\mathrm{WF}}, \beta_{\mathrm{WF}}\} = \underset{\{\boldsymbol{F},\beta\}}{\operatorname{argmin}} \mathrm{E}\left[\|\boldsymbol{u}[n] - \hat{\boldsymbol{u}}[n]\|_2^2\right] \quad \text{s.t.: } \operatorname{tr}(\boldsymbol{F}\boldsymbol{C}_{\boldsymbol{u}}\boldsymbol{F}^{\mathrm{H}}) \leq E_{\mathrm{tx}}. \tag{8}$$

As it has been demonstrated in [17], taking into account this transmit energy constraint $\operatorname{tr}(\boldsymbol{F}\boldsymbol{C}_{\boldsymbol{u}}\boldsymbol{F}^{\mathrm{H}}) = E_{\mathrm{tx}}$, the solution for the linear precoder designed using this MSE criterion is given by

$$\begin{aligned} \boldsymbol{F}_{\mathrm{WF}} &= \beta_{\mathrm{WF}}^{-1}\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \xi\mathbf{I}\right)^{-1}\boldsymbol{H}^{\mathrm{H}}, \\ \beta_{\mathrm{WF}} &= \sqrt{\frac{\operatorname{tr}\left(\left(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \xi\mathbf{I}\right)^{-2}\boldsymbol{H}^{\mathrm{H}}\boldsymbol{C}_{\boldsymbol{u}}\boldsymbol{H}\right)}{E_{\mathrm{tx}}}}, \end{aligned} \tag{9}$$

where $\xi = \operatorname{tr}\left(\frac{\boldsymbol{C}_{\boldsymbol{\eta}}}{E_{\mathrm{tx}}}\right)$.

It is interesting to note that the linear Wiener precoder converges to the linear zero–forcing precoder for $\xi = \frac{\operatorname{tr}(\boldsymbol{C}_{\boldsymbol{\eta}})}{E_{\mathrm{tx}}} \to 0$, i.e. for SNR $\to \infty$, where SNR is the ratio of transmit and noise powers at the receiver.

3. Adaptive Algorithms

The model described by Equations (3) and (4) can be summarized as follows

$$\boldsymbol{y}[n] = \boldsymbol{A}\boldsymbol{d}[n] + \boldsymbol{\eta}[n], \tag{10}$$

where the meaning of the mixing matrix $\boldsymbol{A}$ and of $\boldsymbol{d}[n]$ will depend on our target. In Equation (3), the mixing matrix is the channel matrix $\boldsymbol{H}$ with $\boldsymbol{d}[n]$ containing the coded signals, $\boldsymbol{x}[n]$. However, in Equation (4), it is considered the whole coding-channel system, given by $\boldsymbol{F}\boldsymbol{H}$, as the mixing matrix, and $\boldsymbol{d}[n]$ contains the uncoded signals, $\boldsymbol{u}[n]$, instead of the coded ones as before. For both cases, we assume that the mixing matrix is unknown, although it is full rank nevertheless. The sources are also assumed to be statistically independent. Without any loss of generality, we can suppose that the sources have a normalized power value equal to one since possible fluctuations of power values can be included into the mixing matrix $\boldsymbol{A}$.

The source data can be recovered by means of a linear system whose output is an instantaneous combination of the observations, expressed as

$$\boldsymbol{z}[n] = \boldsymbol{W}^{\mathrm{H}}[n]\boldsymbol{y}[n], \tag{11}$$

where $\boldsymbol{W}[n] \in \mathbb{C}^{N\times N}$ is a matrix containing the so–called *separation coefficients* obtained from adaptive algorithms as follows

$$\boldsymbol{W}[n+1] = \boldsymbol{W}[n] + \mu\boldsymbol{\Delta}(\boldsymbol{W}[n], \boldsymbol{y}[n]), \tag{12}$$

where μ is the step-size parameter that satisfies $0 \leq \mu \leq 1$ and $\boldsymbol{\Delta}(\cdot,\cdot)$ is a matrix computed from both the separating matrix $\boldsymbol{W}[n]$ and the observed signal vector $\boldsymbol{y}[n]$ at the time instant n. The exact expression for $\boldsymbol{\Delta}(\cdot,\cdot)$ depends on the separation criterion. Adaptive algorithms are especially attractive for real-time applications since each iteration only depends on the input at the instant n, thus avoiding the need of storing the values for the observations corresponding to previous instants, and it obtains an updated output sample each time a new observation is received.

The purpose of many adaptive algorithms is to minimize (or maximize) a cost function $J[n]$. In this case, the matrix $\boldsymbol{\Delta}(\boldsymbol{W}[n], \boldsymbol{y}[n])$ would be computed as the gradient of $J[n]$ with respect to $\boldsymbol{W}[n]$, denoted as $\nabla_{\boldsymbol{W}} J$. Therefore, the separation coefficients are obtained using the following rule

$$\boldsymbol{W}[n+1] = \boldsymbol{W}[n] - \mu \; \nabla_{\boldsymbol{W}} J[n]. \tag{13}$$

An alternative way to compute the minimum of this cost function consists in using a relative gradient. As shown in [1, 5], the relative gradient is obtained by means of multiplying the gradient $\nabla_{\boldsymbol{W}} J[n]$ by $\boldsymbol{W}[n]\boldsymbol{W}^{\mathrm{H}}[n]$, i.e. the recursion takes the form

$$\boldsymbol{W}[n+1] = \boldsymbol{W}[n] - \mu \; \boldsymbol{W}[n]\boldsymbol{W}^{\mathrm{H}}[n]\nabla_{\boldsymbol{W}} J[n]. \tag{14}$$

3.1. Supervised Approach

An important family of adaptive filtering algorithms arises from considering the minimization of the MSE between the outputs $\boldsymbol{y}[n]$ and the desired signals $\boldsymbol{d}[n]$ [14, 15]. Mathematically, the cost function is defined as

$$\begin{aligned} J_{\mathrm{MSE}} &= \sum_{i=1}^{N} \mathrm{E}\left[|z_i[n] - d_i[n]|^2\right] \\ &= \mathrm{E}\left[\mathrm{tr}\left((\boldsymbol{W}^{\mathrm{H}}[n]\boldsymbol{y}[n] - \boldsymbol{d}[n])(\boldsymbol{W}^{\mathrm{H}}[n]\boldsymbol{y}[n] - \boldsymbol{d}[n])^{\mathrm{H}}\right)\right]. \end{aligned} \tag{15}$$

The gradient of this cost function is obtained as

$$\nabla_{\boldsymbol{W}} J_{\mathrm{MSE}} = \mathrm{E}\left[\boldsymbol{y}[n](\boldsymbol{W}^{\mathrm{H}}[n]\boldsymbol{y}[n] - \boldsymbol{d}[n])^{\mathrm{H}}\right]. \tag{16}$$

The classical stability analysis for gradient-based algorithms consists in the study of the point where the gradient vanishes and in the definition of the Hessian matrix containing the second derivatives of J_{MSE} [3]. In particular, it can be demonstrated that the stationary points of the Least Mean Squares (LMS) rule are obtained as

$$\nabla_{\boldsymbol{W}} J_{\mathrm{MSE}} = 0 \Rightarrow \boldsymbol{W} = \boldsymbol{C}_{\boldsymbol{y}}^{-1}\boldsymbol{C}_{\boldsymbol{yd}}, \tag{17}$$

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

In practice, the desired signals are considered known only during a finite number of instants (pilot symbols) so estimation is used to recover several frames. For this reason, the performance of this type of algorithms is degraded in presence of calibration errors.

3.2. Unsupervised Approach

Unsupervised approaches allow the mixing matrix $\boldsymbol{A}$ to be estimated directly from the observations, without using pilot symbols. In this section, we show two strategies for obtaining the estimate of the joint channel-precoding matrix ($\boldsymbol{HF}$).

The first approach directly arises from Equation (16) by assuming that the sources are uncorrelated with unitary power, i.e. $\boldsymbol{C}_{\boldsymbol{u}} = \boldsymbol{I}$. Then, Equation (16) can be expressed as follows

$$\begin{aligned}
\nabla_{\boldsymbol{W}} J_{\mathrm{MSE}} &= \mathrm{E}\left[\boldsymbol{y}[n](\boldsymbol{W}^{\mathrm{H}}[n]\boldsymbol{y}[n] - \boldsymbol{d}[n])^{\mathrm{H}}\right] \\
&= \mathrm{E}\left[\boldsymbol{y}[n](\boldsymbol{z}[n] - \boldsymbol{d}[n])^{\mathrm{H}}\right] \\
&= \mathrm{E}\left[\boldsymbol{y}[n]\boldsymbol{z}^{\mathrm{H}}[n]\right] - \mathrm{E}\left[\boldsymbol{y}[n]\boldsymbol{d}^{\mathrm{H}}\,[n]\right] \\
&= \mathrm{E}\left[\boldsymbol{y}[n]\boldsymbol{z}^{\mathrm{H}}[n]\right] - \boldsymbol{A}\,\mathrm{E}\left[\boldsymbol{d}[n]\boldsymbol{d}^{\mathrm{H}}[n]\right] \\
&= \mathrm{E}\left[\boldsymbol{y}[n]\boldsymbol{z}^{\mathrm{H}}[n]\right] - \boldsymbol{A},
\end{aligned} \tag{18}$$

since $\mathrm{E}[\boldsymbol{d}[n]\boldsymbol{d}^{\mathrm{H}}[n]] = \mathrm{E}[\boldsymbol{u}[n]\boldsymbol{u}^{\mathrm{H}}[n]] = \boldsymbol{C}_{\boldsymbol{u}} = \boldsymbol{I}$ and where the expectation only uses one sample. Thus, the following adaptive algorithm is derived

$$\boldsymbol{W}[n+1] = \boldsymbol{W}[n] - \mu\left(\boldsymbol{y}[n]\boldsymbol{z}^{\mathrm{H}}[n] - \boldsymbol{A}\right), \tag{19}$$

with μ being the step-size parameter. Equation (19) can be interpreted as a generalization of the **P-vector** algorithm described by Griffiths [13]. In such a case, the separation matrix requires information about the mixing matrix $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{F}$, which can be acquired using the supervised approach previously explained. This supervised setup gives us the channel matrix estimate and, consequently, the precoder matrix $\boldsymbol{F}$, as shown in Equation (9).

Both pilot symbols and prior knowledge about mixing matrices could be avoided by using Blind Source Separation (BSS) algorithms. BSS methods simultaneously obtain the mixing matrix as well as the realizations of the source vector, $\boldsymbol{d}[n]$, all from the corresponding realizations of the observed vector, $\boldsymbol{y}[n]$. This prior unknowledge may limit the achievable performance, but makes blind approaches more robust against calibration errors (i.e. deviations of the model assumptions from the resulting ones) than conventional array processing techniques [6]. A property commonly exploited in BSS methods is that sources are statistically independent.

One of the best known BSS algorithms has been approached by Bell and Sejnowski [2]. Given an activation function $h(\cdot)$, the idea proposed by these authors is to obtain the weighted coefficients of a Neural Network, $\boldsymbol{W}[n]$, in order to maximize the mutual information between the outputs of the activation function, given by $\boldsymbol{h}(\boldsymbol{z}[n]) = \boldsymbol{h}(\boldsymbol{W}^{\mathrm{H}}[n]\boldsymbol{y}[n])$, and the inputs $\boldsymbol{y}[n]$. The resulting method can be expressed as

$$J_{\mathrm{MI}} = \ln(\det(\boldsymbol{W}^{\mathrm{H}}[n])) + \sum_{i=1}^{N} \mathrm{E}[\ln(h_i'(z_i[n]))], \tag{20}$$

with MI being the abbreviation for *Mutual Information.* h_i is the i–th element of the vector $\boldsymbol{h}(\boldsymbol{z}[n])$, and $'$ denotes the first derivative. The maximum of this cost function can be obtained using a gradient algorithm [2], or a relative gradient algorithm [1, 22]. Both approaches use the gradient of Equation (20) given by

$$\begin{aligned}
\nabla_{\boldsymbol{W}} J_{MI} &= \nabla_{\boldsymbol{W}} \left(\ln(\det(\boldsymbol{W}^{\mathrm{H}}[n]))\right) + \nabla_{\boldsymbol{W}} \left(\sum_{i=1}^{N_{\mathrm{B}}} E[\ln(h_i'(z_i[n]))] \right) \\
&= \frac{\operatorname{adj}(\boldsymbol{W}^{\mathrm{H}}[n])}{\det(\boldsymbol{W}^{\mathrm{H}}[n])} - E[\boldsymbol{y}[n]\boldsymbol{g}^{\mathrm{H}}(\boldsymbol{z}[n])] \\
&= \boldsymbol{W}^{-\mathrm{H}}[n] - E[\boldsymbol{y}[n]\boldsymbol{g}^{\mathrm{H}}(\boldsymbol{z}[n])], \qquad (21)
\end{aligned}$$

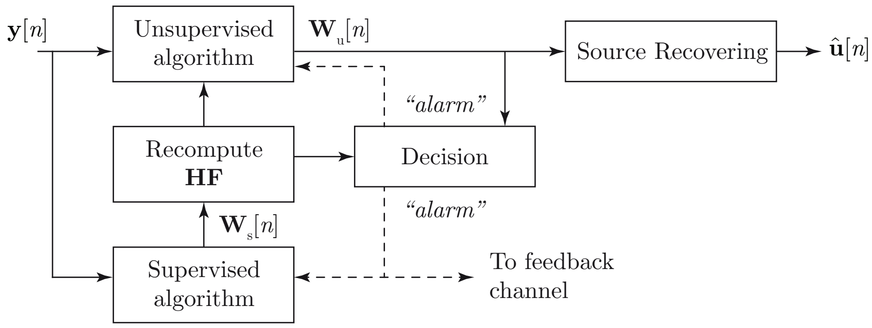

Figure 3: Block diagram for the hybrid approach.

where $adj(\cdot)$ is the adjunct of a matrix and

$$\boldsymbol{g}(\boldsymbol{z}[n]) = [-h_1''(z_1[n])/h_1'(z_1[n]), \cdots, -h_N''(z_N[n])/h_N'(z_N[n])]^{\mathrm{T}},$$

depends on the activation function. Finally, for the case in which the above expectation is estimated using only one sample, we obtain the following relative gradient algorithm (termed as **Infomax**):

$$\begin{aligned} \boldsymbol{W}[n+1] &= \boldsymbol{W}[n] + \mu \boldsymbol{W}[n]\boldsymbol{W}^{\mathrm{H}}[n]\left(\boldsymbol{y}[n]\ \boldsymbol{g}^{\mathrm{H}}(\boldsymbol{z}[n]) - \boldsymbol{W}^{-\mathrm{H}}[n]\right) \\ &= \boldsymbol{W}[n] + \mu \boldsymbol{W}[n]\left(\boldsymbol{z}[n]\boldsymbol{g}^{\mathrm{H}}(\boldsymbol{z}[n]) - \boldsymbol{I}\right), \end{aligned} \tag{22}$$

where we have used that $\boldsymbol{z}[n] = \boldsymbol{W}^{\mathrm{H}}[n]\boldsymbol{y}[n]$. The expression in Equation (22) 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 [7] have shown that the Bell and Sejnowski rule is equivalent to the Constant Modulus Algorithm (CMA) proposed by Godard [12].

4. Hybrid Approach

The main advantage of adaptive unsupervised algorithms is their ability to track low channel variations. On the contrary, supervised solutions provide a fast channel estimate for both low and high variations at the cost of using pilot symbols. In this section, we propose to combine these two strategies in order to obtain a performance close to that offered by supervised approaches, but using a lower number of pilot symbols. Fig 3 shows a simplified block diagram for this hybrid approach.

We denote by $\boldsymbol{W}_u[n]$ and $\boldsymbol{W}_s[n]$ the matrices containing the separation coefficients for the unsupervised and supervised modules, respectively. We start with an initial estimate of the channel matrix obtained using the Widrow-Hoff solution given by Equation (17) with $\boldsymbol{d}[n] = \boldsymbol{x}[n]$. This channel estimate is used at the TX to design the optimum precoding matrix $\boldsymbol{F}$, and at the RX, to initialize the unsupervised algorithm to $\boldsymbol{W}_u[n] = (\hat{\boldsymbol{H}}\boldsymbol{F})^{-\mathrm{H}}$.

When the module termed as *"decision module"* determines that the channel has not suffered a significant variation, the matrix $\boldsymbol{W}_u[n]$ is adapted and the data symbols $\boldsymbol{u}[n]$ are recovered by using $\boldsymbol{z}[n] = \boldsymbol{W}_u^{\mathrm{H}}[n]\boldsymbol{y}[n]$, being $\boldsymbol{z}[n] = \hat{\boldsymbol{u}}[n]$. Otherwise, if a significant variation has occurred, the RX sends an "alarm" to the TX through the feedback channel. In that moment, pilot symbols have to be sent by the TX to the RX, so at the RX, the supervised algorithm estimates the channel from those pilots. For that purpose, we consider the Widrow-Hoff solution of Equation (17). This solution provides us the channel matrix estimate which is sent to the TX for updating the precoding matrix. The RX also computes that precoding matrix $\boldsymbol{F}$, together with the whole reference matrix $\hat{\boldsymbol{H}}\boldsymbol{F}$, and initializes the unsupervised algorithms according to $\boldsymbol{W}_u[n] = (\hat{\boldsymbol{H}}\boldsymbol{F})^{-H}$.

4.1. Decision Rule

An important issue is how to determine when the channel has suffered a significant variation. By combining Equations (10) and (11) and considering $\boldsymbol{d}[n] = \boldsymbol{u}[n]$, the output $\boldsymbol{z}[n]$ can be rewritten as a linear combination of the sources

$$\boldsymbol{z}[n] = \boldsymbol{\Gamma}[n]\boldsymbol{u}[n], \tag{23}$$

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

$$\boldsymbol{\Gamma}[n] = \boldsymbol{D}\boldsymbol{P}, \tag{24}$$

where $\boldsymbol{D}$ is a diagonal invertible matrix and $\boldsymbol{P}$ is a permutation matrix. The difference of $\boldsymbol{\Gamma}[n]$ with respect to the aforementioned diagonal matrix can be measured using the following "error" criterion proposed by Macchi and

Moreau [23] to evaluate the performance of BSS algorithms

$$\mathrm{Error}[n] = \sum_{i=1}^{N}\left(\sum_{j=1}^{N}\frac{|\gamma_{i,j}[n]|^2}{|\max_l(\gamma_{i,l}[n])|^2}-1\right)+\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\frac{|\gamma_{i,j}[n]|^2}{|\max_l(\gamma_{l,j}[n])|^2}-1\right), \tag{25}$$

where $\gamma_{i,j}[n]$ denotes the element of the i–th row and j-th column of $\boldsymbol{\Gamma}[n]$. Note that the first summation is zero when each row of $\boldsymbol{\Gamma}[n]$ contains only one non-zero entry, i.e. in the event of each output extracting only one source although the same source could be extracted by several outputs. The second summation is zero when each column contains only one non-zero entry. In such a case, each source is extracted by only one output. Obviously, $\mathrm{Error}_i[n] = 0$ if $\boldsymbol{\Gamma}[n]$ is diagonal.

An interesting consequence of using a linear precoder is that the permutation indeterminacy associated to unsupervised algorithms is avoided because of the initialization $\boldsymbol{W}_u[n] = (\hat{\boldsymbol{H}}\boldsymbol{F})^{-\mathrm{H}}$. This means that the sources are recovered in the same ordering as they were transmitted, i.e, $\max_l(\gamma_{i,l}[n]) = \gamma_{i,i}[n]$ and $\max_l(\gamma_{l,j}[n]) = \gamma_{j,j}[n]$. This property has been already used in [8, 11] to simplify Equation (25).

A way of computing the distortion with regard to a diagonal matrix consists on measuring

$$\mathrm{Error}_i[n] = \sum_{j=1, j\neq i}^{N}\frac{|\gamma_{i,j}[n]|^2}{|\gamma_{i,i}[n]|^2}, \tag{26}$$

where we are considering only the rows of $\boldsymbol{\Gamma}[n]$. Thus, the $\mathrm{Error}_i[n]$ is zero when only the element $\gamma_{i,i}[n]$ is not equal to zero. Compared to [8, 11], the above error has not to be computed for each row instead. If some value of $\mathrm{Error}_i[n]$ computed for a certain row is greater than the fixed threshold, then pilot symbols must be required. Therefore, this decision criterion is less complex than that explained in [8, 11], which implies a reduction in terms of computational complexity and a more intuitive way of deciding whether pilots are needed or not. Table 1 summarizes the aforementioned method to detect the instants in which pilot symbols are required.

5. Experimental Results

In order to show the performance achieved with the proposed hybrid scheme, we present the results for several computer simulations performed by considering that 10 000 QPSK symbols have been transmitted through a MIMO

1. Compute the gain matrix $\boldsymbol{\Gamma}[n] = \boldsymbol{W}_u^{\mathrm{H}}[n]\hat{\boldsymbol{HF}}$,
   where $\boldsymbol{W}_u[n]$ is obtained after processing all the frame symbols.
2. Estimate $\hat{\boldsymbol{HF}}$,
   computed by Equations (7) (ZF criterion) or (9) (Wiener criterion), using the supervised approach as a reference.
3. Compute the error for each row, $\mathrm{Error}_i(n)$, using Equation (26).
4. Decide when the channel has significantly changed.
   If some $\mathrm{Error}_i[n] > t \rightarrow$ Request pilot symbols.

Table 1: Algorithm steps for tracking large channel variations.

system in blocks of 100 symbols each one. The system consists of four TX and four RX antennas. All the simulation results have been obtained by averaging 100 independent channel realizations. From one realization to another, the channel matrix is changed using the following model:

$$\boldsymbol{H} = (1-\alpha)\boldsymbol{H} + \alpha\boldsymbol{H}_{\mathrm{new}}, \tag{27}$$

where $\boldsymbol{H}_{\mathrm{new}}$ is a 4×4 complex matrix randomly generated according to a Gaussian distribution, and α is termed as *channel updating parameter*, which lies on the interval $[0, 1]$. Thus, if α is high (close to one), the channel totally changes from one realization to another, and if it is low (close to zero), the channel almost remains unchanged from one realization to another.

By computer simulations, the performance of the following schemes has been evaluated:

As the supervised approach, the Widrow-Hoff algorithm described in Equation (17) is computed using 100 pilot symbols transmitted each 2 000 symbols.

As unsupervised approaches, the generalized P-vector of Equation (19) and the Infomax algorithm of Equation (22), both initialized to the whole matrix $\boldsymbol{HF}$ obtained from the above Widrow-Hoff solution. The step-size parameter has been set to $\mu = 0.001$.

By combining supervised and unsupervised approaches, the hybrid proposal which makes use of the generalized P-vector and of the Infomax algoritms for different values of the threshold t. The step-size parameter has also been set to $\mu = 0.001$. A frame of 100 pilot symbols is sent when the measured error is greater than the given threshold, i.e. when the hybrid approach switches to the supervised working.

5.1. Experiment 1: Fixed channel coherence time

In the first set of experiments, the channel updating parameter α in Equation (27) is fixed to 0.1. Moreover, it is considered that the channel remains constant during the transmission of 2 000 symbols. This is an ideal scenario for the supervised approach since the pilots are transmitted each 2 000 symbols, i.e. just in the time instants when the channel changes. Consequently, the performance achieved for the supervised approach constitutes the lower bound for the performances experienced by all the schemes.

Figure .4 shows the BER performance as a function of the SNR for a ZF precoding setup. For both, the generalized P–vector and the Infomax algorithms, note the considerable improvement in terms of the BER obtained with the hybrid approaches compared to the unsupervised ones. We can also conclude that Infomax is always the best choice independently from the threshold previously fixed. However, the number of pilot symbols sent from the TX to the RX is lightly higher, specially for medium SNR, as depicted in Figure .5. In this figure, the overhead of the feedback channel expressed in terms of the number of pilot symbols needed to track channel variations is plotted. In this case, the lower bound is given by the unsupervised approaches, since only 100 pilots are needed to initialize its respective algorithms. As considered in the computer simulations, the figure represents the results for two different values of the threshold t, which has an influence on the switching of hybrid proposal working. As can be observed from the figure, a good choice for the threshold t, for example $t = 0.2$, let us reduce the number of pilots in a significant percentage (about 50% for 10 dB), with respect to a bad choice, $t = 0.05$, but as a consequence a loss in BER performance appeared. Therefore, by choosing appropriate parameters for the hybrid approach, its adequate trade–off between BER performance and feedback channel overhead leads the hybrid proposal to be quite interesting in scenarios with transmit processing and unavailable channel state information at the transmitter side.

Figures .6 and .7 also show the BER performance and the number of pilots sent through the feedback channel versus SNR but for a MIMO WF-precoding system instead. Again, the performance of the hybrid approach is much better than that obtained with the unsupervised approach. Taking into account these results, we can affirm that the hybrid setup implementing the Infomax algorithm for $t = 0.2$ is an attractive technique since it achieves a BER close to the one offered by the supervised approach, but requiring fewer pilot symbols, and consequently, increasing the spectral efficiency of the system. Compared to the ZF approach, we always obtain better BER performances for the Wiener designs, as expected, due to the drawback of amplifying noise inherent to ZF designs. However, looking at the number of pilots, it can be observed that the Wiener hybrid proposal needs more pilots for very low SNR (SNR $< 4\,$dB). This effect is not very important for Infomax, but it is quite obvious for P-vector, and therefore, P–vector is clearly overcome by Infomax in Wiener implementations.

5.2. Experiment 2: Variable channel coherence time

In the second set of experiments, we consider that the channel remains constant during the transmission of a random number of symbols, ranging from 2 000 to 3 000.

The results of BER and number of pilots versus SNR for $\alpha = 0.1$ with a ZF linear precoder are shown in Figures .8 and .9, respectively. For a Wiener linear precoder, the results are respectively plotted in Figures .10 and .11. It can be seen that the supervised approach presents a considerable performance loss —in terms of BER— since the pilot symbols are transmitted periodically each 2 000 symbols, although the channel may change between the transmission of two consecutive sendings of pilots. However, as it can be seen from the figures, the hybrid approach is robust against calibration errors. Again, the use of the Infomax algorithm improves the BER results, and similar behaviour as before is observed for low SNR with regard to the number of pilots needed by the Wiener hybrid proposal.

Finally, Figures .12 and .13 illustrate the BER achieved by all the schemes under study as a function of the channel updating parameter α restricting the transmit filter to be designed according to the ZF criterion. The respective results for the Wiener case are plotted in Figures .14 and .15, respectively. For both setups, the SNR is set to 15 dB. Note that the BER curves corresponding to the hybrid approaches implementing Infomax converge to the P-vector curves for $\alpha \to 1$ but sending more pilot symbols to reach that performance. As can be observed from all the figures, our hybrid proposal always works better than exclusively supervised or unsupervised appraches for any α. Anyway, for a given channel fading value α, and an adequate choice of t (for example, $t = 0.2$ and $\alpha = 0.1$ or smaller than that value) leads to get a reduction of about 50% in the number of pilots compared to the supervised number needed to track channel variations. Moreover, the BER reduction is of about 2×10^{-2} and 6×10^{-3} at the operation point corresponding to $\alpha = 0.1$, for ZF and Wiener lienar precoding, respectively.

5.3. Computational Complexity

Table 2 shows the computational complexity of all the recovering algorithms compared in this section by computer simulations, as well as the complexity associated to the detection criterion used by the proposed hybrid scheme. It can be seen that the use of this criterion implies a negligible increasing of the computational complexity since it is computed once for each frame. In this table, N_B represents the total number of symbols in a frame and N the number of antennas at each side of the link.

6. Conclusions

This paper deals with the utilization of supervised and/or unsupervised algorithms in FDD systems together with ZF and Wiener linear precoding. We have introduced the well–known Wiener-Hoff solution as an adequate supervised method to estimate channel coefficients with reduced MSE between observations and coded signals. We have also derived the unsupervised algorithm termed as generalized P-vector, which allows us to estimate the whole precoding-channel matrix without sending pilot symbols. This algorithm uses the whole matrix estimate previously obtained making use of the supervised technique. Moreover, we have introduced the Infomax algorithm, proposed in the context of BSS, as an adaptive method that makes possible the estimation of the whole precoding-channel matrix directly from the observations, with the only assumption of that sources and channel are completely unknown.

Considering a communications model where the channel is block flat fading, we have proposed a simple scheme to detect channel variations in those FDD systems with linear precoding. The hybrid approach combines both supervised and unsupervised algorithms making use of the new decision criterion. When channel variations are significant, the system implements a supervised approach to get channel estimates to be sent to the TX through the feedback channel. In other case, an unsupervised adaptive algorithm is used to recover the transmitted signals. The aforementioned decision rule determines when the system has to switch from the unsupervised working to

Widrow-Hoff Solution

Compute $\mathbf{C_y}$ (or $\mathbf{C_{yd}}$)	$N^2 \times N_\mathrm{B}$ complex multiplications $N^2 \times (N_\mathrm{B}-1)$ complex summations
Matrix inversion $\mathbf{C}_{\boldsymbol{y}}^{-1}$	$O(N^3)$ for the Gauss-Jordan method
Compute $\mathbf{C}_{\boldsymbol{y}}^{-1}\mathbf{C_{yd}}$	N^3 complex multiplications $N^2 \times (N-1)$ complex summations
Total	$O(N^2 \times N_\mathrm{P})$

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 $\mathbf{P}_1 = \mathbf{z}\mathbf{g}^H(\mathbf{z}) - \mathbf{I}$	N^2 complex multiplications N complex summations
Compute $\mathbf{P}_2 = \mathbf{W}[n]\mathbf{P}_1$	N^3 complex multiplications $N^2 \times (N-1)$ complex summations
Update $\mathbf{W}[n+1] = \mathbf{W}[n] + \mu\mathbf{P}_2$	N^2 real-complex multiplications N^2 complex summations $N(N-1)$ real summations
Total (for N_B symbols)	$O(N^3 \times N_\mathrm{B})$

Unsupervised approach P-vector

For each iteration:	
Compute $\mathbf{P}_1 = \mathbf{y}\mathbf{z}^H - \mathbf{A}$	N^2 complex multiplications N complex summations
Update $\mathbf{W}[n+1] = \mathbf{W}[n] - \mu\mathbf{P}_1$	N^2 real-complex multiplications N^2 complex summations $N(N-1)$ real summations
Total (for N_B symbols)	$O(N^3 \times N_\mathrm{B})$

Decision criterion

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

Table 2: Computational cost of the supervised and unsupervised algorithms and of the decision criterion.

the supervised one. Simulation results have shown that the combined proposal is an attractive solution for communications systems since it provides an adequate BER with a low overhead due to the transmission of pilot symbols. Moreover, when the ZF linear precoding designs are compared to the respective MMSE designs, the first criterion is clearly outperformed by the second one, as expected.

Acknowledgments

This work was supported by 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, and CSD2008-00010.

References

[1] Amari, S. (1996). Gradient learning in structured parameter spaces: Adaptive blind separation of signal sources. In *Proc. WCNN'96* (pp. 951–956). San Diego.

[2] Bell, A., & Sejnowski, T. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*, 1129–1159.

[3] Benveniste, A., Métivier, M., & Priourent, P. (1990). *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag, New York.

[4] Bourgeois, J., & Minker, W. (2009). Comparison of LCMV Beamforming and Second–Order Statistics BSS. *Lecture Notes in Computer Science: Time-Domain Beamforming and Blind Source Separation*, *3*, 125–146.

[5] Cardoso, J.-F., & Laheld, B. (1996). Equivariant adaptive source separation. *IEEE Transactions on Signal Processing*, *44*, 3017–3030.

[6] Cardoso, J. F., & Souloumiac, A. (1993). Blind Beamforming for Non–Gaussian Signals. *IEEE Proceedings-F*, *140*, pp.362–370.

[7] Castedo, L., & Macchi, O. (April 1997). Maximizing the information transfer for adaptive unsupervised source separation. In *Proc. SPAWC'97* (pp. 65–68). Paris, France.

[8] Castro, P., García-Naya, J., Iglesia, D., & Dapena, A. (2010). A novel hybrid approach to improve performance on frequency division duplex

systems with linear precoding. In *Lecture Notes on Artificial Intelligence* (pp. 248–255). volume vol.II.

[9] Choi, R. L., & Murch, R. D. (2003). New transmit schemes and simplified receiver for MIMO wireless communication systems. *IEEE Transactions on Wireless Communications*, *2*, 1217–1230.

[10] Comon, P., & Jutten, C. (2010). *Handbook of Blind Source Separation, Independent Component Analysis and Applications*. Academic Press.

[11] Dapena, A., Castro, P., & Labrador, J. (2010). Combination of supervised and unsupervised algorithms for communication systems with linear precoding. In *Proc. of World Congress on Computational Intelligence (WCCI 2010)*. Barcelona, Spain.

[12] Godard, D. (November 1980). Self-recovering equalization and carrier tracking in two-dimensional data communication systems. *IEEE Transactions on Communications*, *28*, 1867–1875.

[13] Griffiths, L. J. (1969). A simple adaptive algorithm for real-time processing in antenna arrays. *Proceedings IEEE*, *57*, 1696–1704.

[14] Haykin, S. (1994). *Neural Networks A Comprehensive Foundation*. Macmillan College Publishing Company, New York.

[15] Haykin, S. (1996). *Adaptive Filter Theory*. 3rd. ed., Prentice Hall.

[16] Hunger, R., Joham, M., & Utschick, W. (2005). Extension of linear and nonlinear transmit filters for decentralized receivers. In *European Wireless 2005* (pp. 40–46). volume 1.

[17] Joham, M. (2004). *Optimization of Linear and Nonlinear Transmit Signal Processing. PhD dissertation*. Munich University of Technology.

[18] Joham, M., Kusume, K., Gzara, M. H., Utschick, W., & Nossek, J. A. (2002). Transmit Wiener filter for the downlink of TDD DS-CDMA systems. In *Proc. ISSSTA* (pp. 9–13). volume 1.

[19] Joham, M., Utschick, W., & Nossek, J. A. (2005). Linear Transmit Processing in MIMO Communications Systems. *IEEE Transactions on Signal Processing*, *53*, 2700–2712.

[20] Karimi, H. R., Sandell, M., & Salz, J. (1999). Comparison between transmitter and receiver array processing to achieve interference nulling and diversity. In *Proc. PIMRC* (pp. 997–1001). volume 3.

[21] Markowska-Kaczmar, U., & Switek, T. (2009). Combined unsupervised–supervised classification method. *Lecture Notes in Computer Science: Knowledge-Based and Intelligent Information and Engineering Systems*, *5712*, 861–868.

[22] Mejuto, C., & Castedo, L. (1997). A neural network approach to blind source separation. In *Proc. Neural Networks for Signal Processing VII* (pp. 486–595). Florida, USA.

[23] Moreau, E., & Macchi, O. (1994). Complex self-adaptive algorithms for source separation based on higher order contrasts. *Signal Processing VII: Theories and Applications*, (pp. 1157–1160).

[24] Nossek, J. A., Joham, M., & Utschick, W. (2004). 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* (pp. 18–23). volume vol.I. Shanghai, China.

[25] Stone, J. V. (2004). *Independent Component Analysis: A Tutorial Introduction*. MIT Press.

[26] Vojčić, B. R., & Jang, W. M. (1998). Transmitter Precoding in Synchronous Multiuser Communications. *IEEE Transactions on Communications*, *46*, 1346–1355.

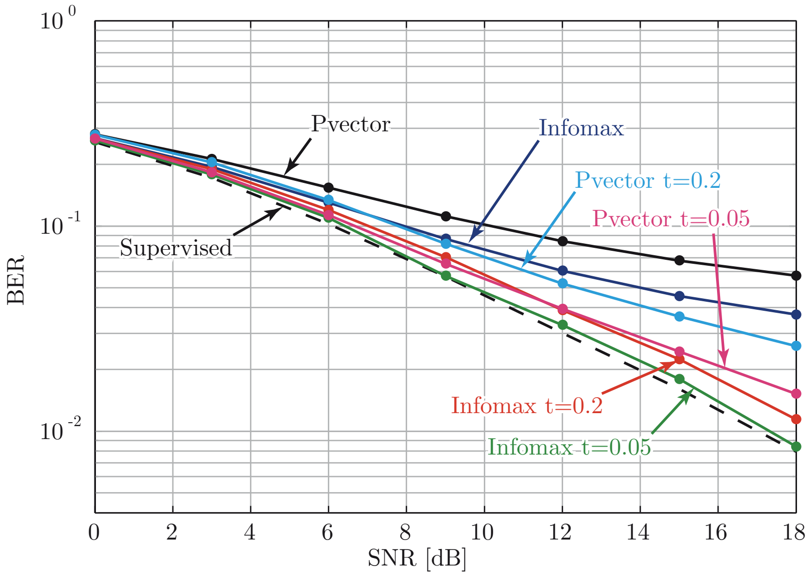

Figure .4: Experiment 1. Fixed distance for channel updating: BER versus SNR for ZF–precoding.

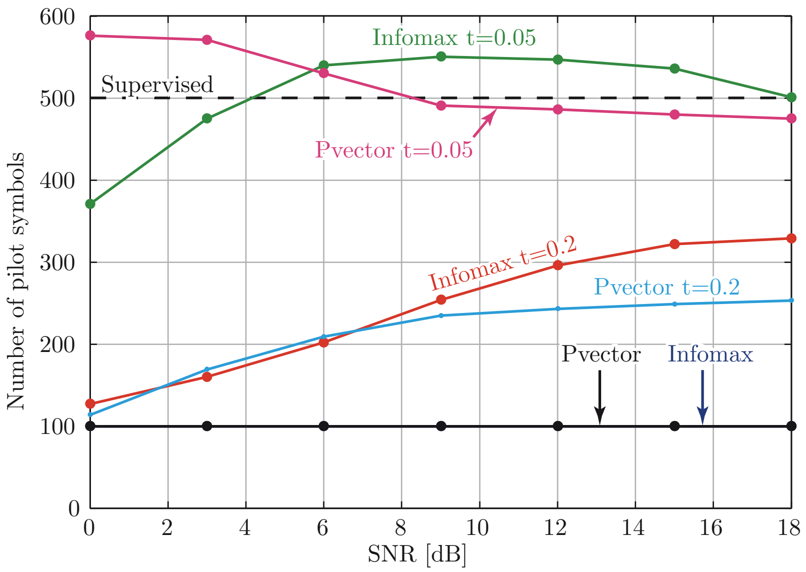

Figure .5: Experiment 1. Fixed distance for channel updating: Number of pilots versus SNR for ZF–precoding.

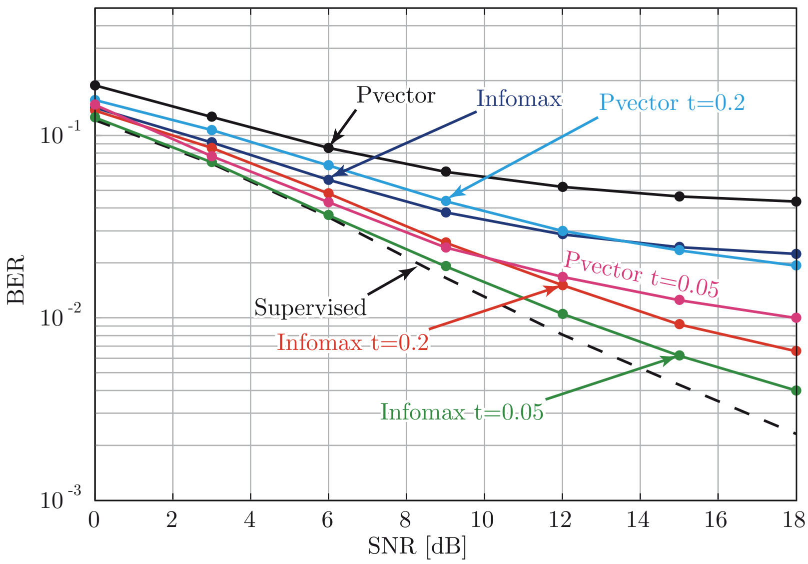

Figure .6: Experiment 1. Fixed distance for channel updating: BER versus SNR for WF–precoding.

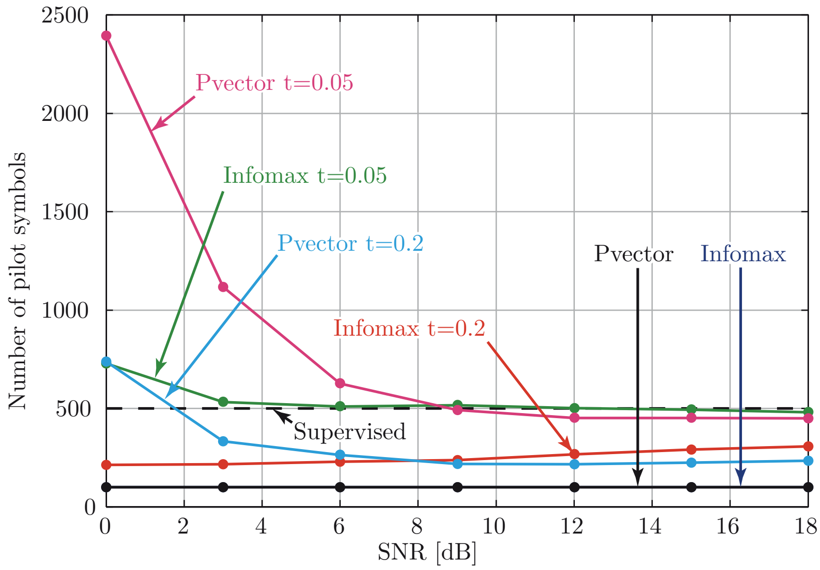

Figure .7: Experiment 1. Fixed distance for channel updating: Number of pilots versus SNR for WF–precoding.

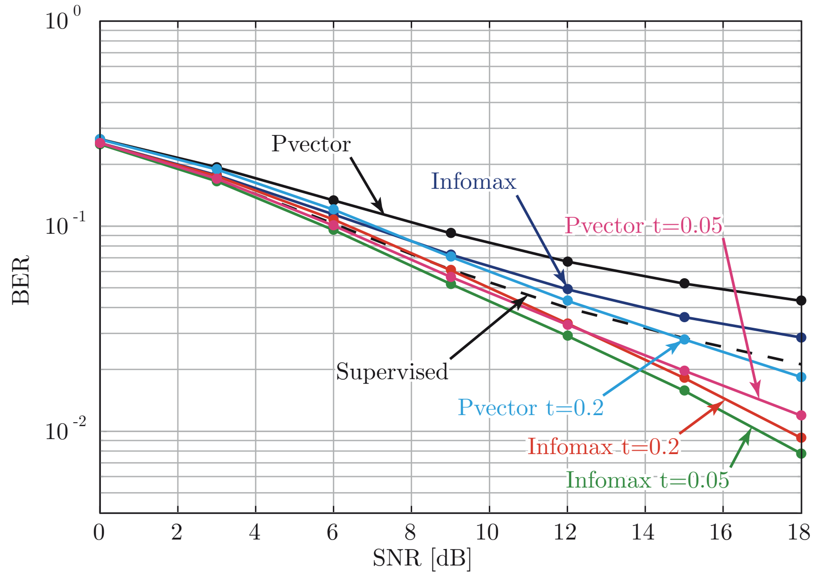

Figure .8: Experiment 2. Variable distance for channel updating: BER versus SNR for ZF–precoding.

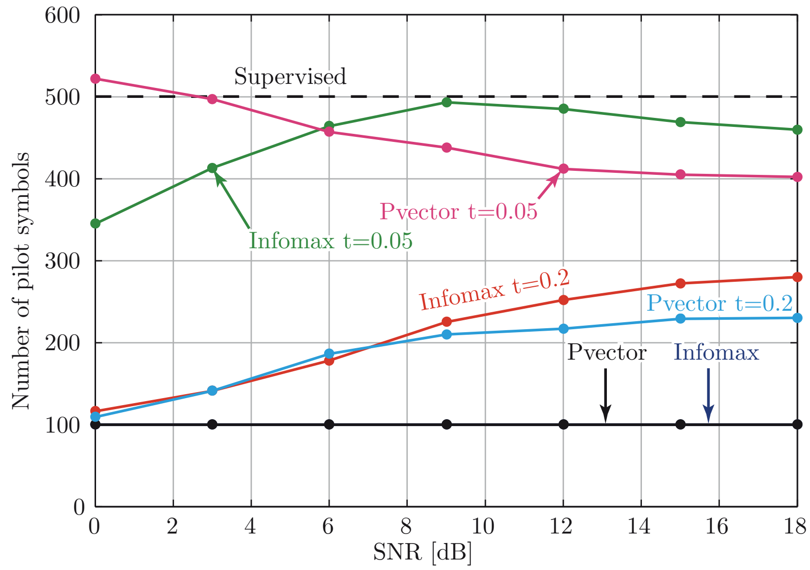

Figure .9: Experiment 2. Variable distance for channel updating: Number of pilots versus SNR for ZF–precoding.

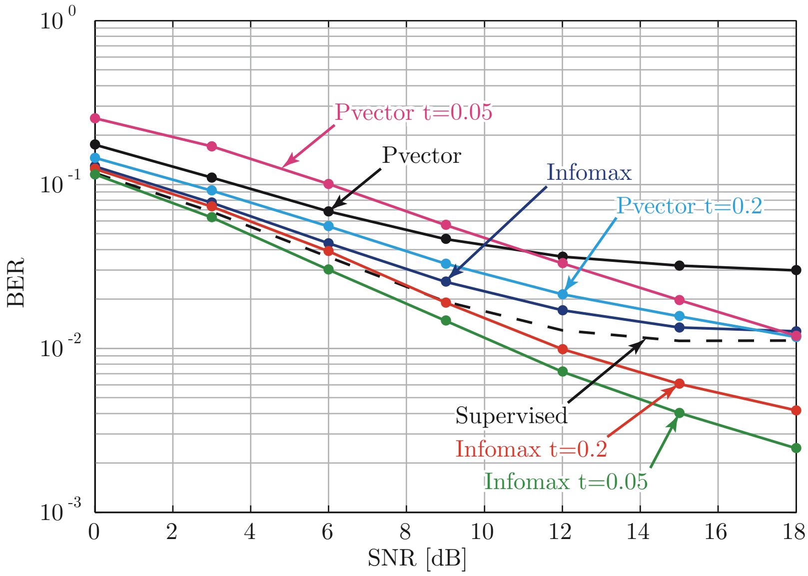

Figure .10: Experiment 2. Variable distance for channel updating: BER versus SNR for WF–precoding.

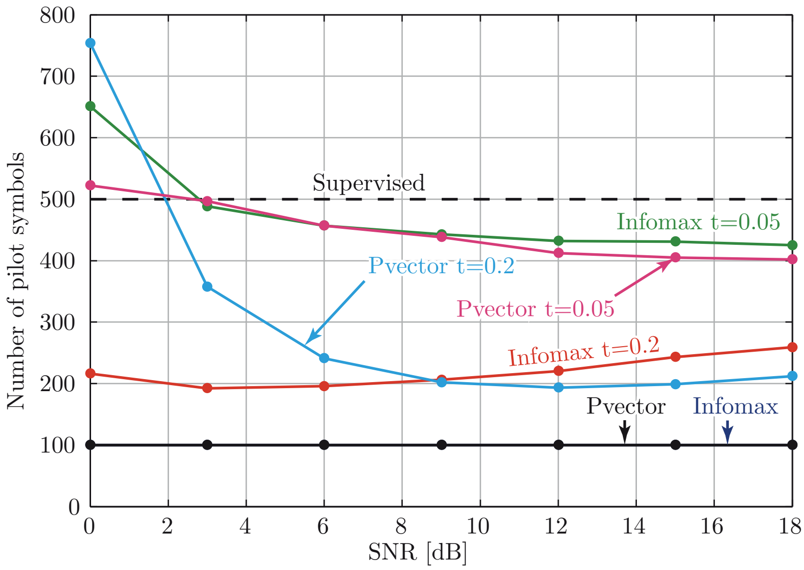

Figure .11: Experiment 2. Variable distance for channel updating: Number of pilots versus SNR for WF–precoding.

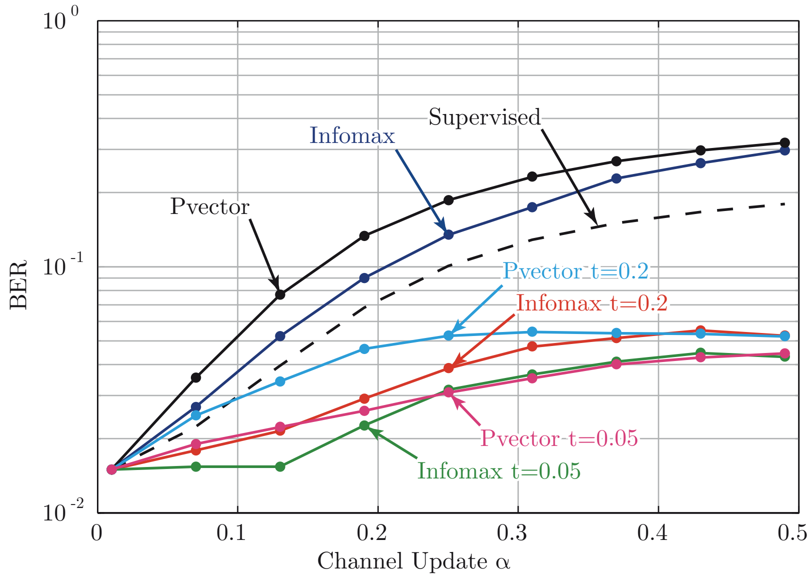

Figure .12: Experiment 2. Variable distance for channel updating: BER in terms of the channel updating parameter for ZF–precoding.

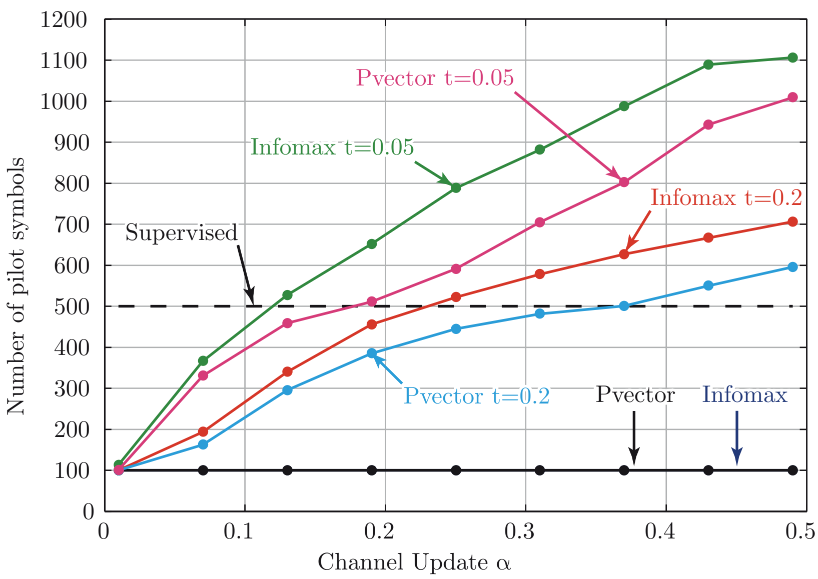

Figure .13: Experiment 2. Variable distance for channel updating: Number of pilots in terms of the channel updating parameter for ZF–precoding.

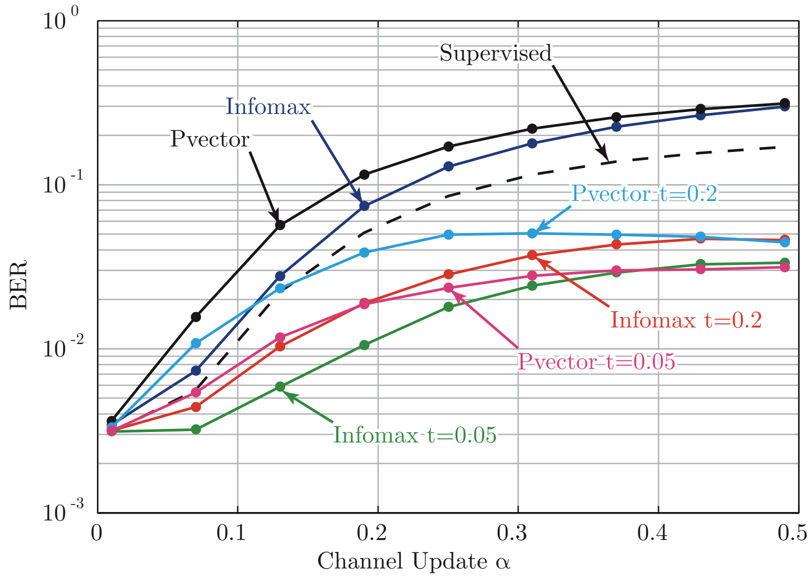

Figure .14: Experiment 2. Variable distance for channel updating: BER in terms of the channel updating parameter for WF–precoding.

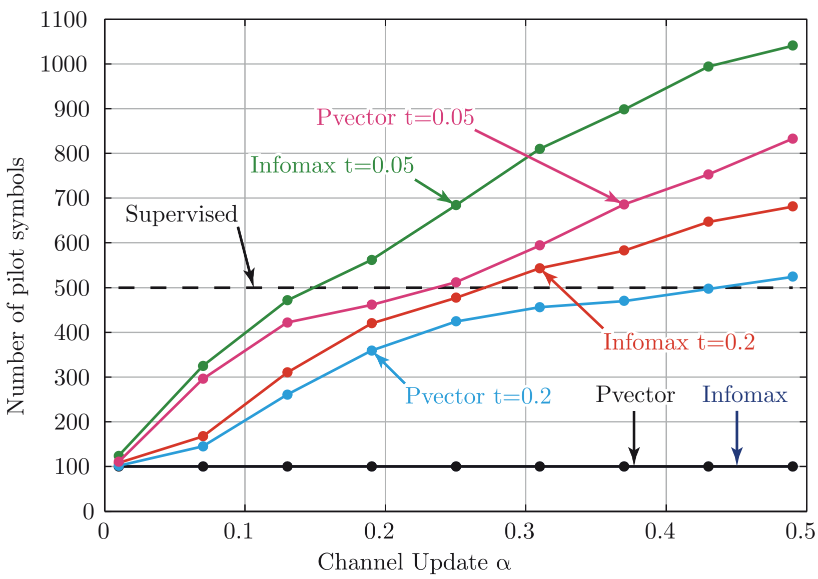

Figure .15: Experiment 2. Variable distance for channel updating: Number of pilots in terms of the channel updating parameter for WF–precoding.