Training-Based Channel Estimation in Multiple Antennae and Multicarrier Systems

Dimitrios Katselis *

National and Kapodistrian University of Athens
Department of Informatics and Telecommunications

Abstract. This work examines the problem of training-based channel estimation in multiple antennae and multicarrier systems. In the case of multiple antennae systems, optimal training sequences w.r.t. minimizing the estimation mean square error are derived for optimal linear estimators that use partial knowledge about the state of the unknown channel, additive noise and multiuser interference. This partial knowledge refers to the first and second order statistics of the aforementioned random processes. Two estimators accompanied by their optimal training sequences are proposed: the Bayesian Minimum Disturbance Estimator (BMDE) and the Two-Sided Bayesian Minimum Disturbance Estimator (2S BMDE). Their performances and feedback requirements are investigated both analytically and via simulations, while certain properties of the Toeplitz matrices are explored in order to reduce the feedback requirements for both schemes. In the second part of this work the Orthogonal Frequency Division Multiplexing (OFDM) system using a Quadrature Amplitude Modulation (QAM) and a cyclic prefix (CP) is compared with the OFDM system employing an Offset QAM (OQAM) transmission strategy, w.r.t. to the channel estimation performances of the optimal preambles for both systems. Optimal training sequences are derived for both systems, while their fundamental differences and similarities are highlighted via this exposition.

1 Introduction

A crucial task in the operation of a transmission system is the availability of exact knowledge about the effect of the physical channel on the transmitted signal. Exact knowledge of this effect offers the possibility of coherent detection of the transmitted symbol, which is superior than semicoherent and noncoherent methods [10]. In the case of multiple antennae systems, this knowledge is more difficult to be acquired, while it is crucial to be obtained in order to exploit the huge capacity that is promised by theoretical studies of these systems [11, 1]. In the case of multicarrier systems, accurate channel knowledge is desirable for similar reasons.

Due to complexity and pure signal processing reasons, the estimators that we usually study in the context of practical transmission systems are linear.

* Dissertation Advisor: Sergios Theodoridis, Professor

The complexity justification is almost obvious, while from the aspect of signal processing the justification is based on the fact that nonlinear transformations of a certain input spectrum lead to undesirable generation of new frequencies in the output spectrum. The optimal algorithms derived in this work belong to the general class of *parametric* estimation methods. They are based on a model for the transmitted and received signal, while the unknown channel is a parameter of this model.

This work is divided in two parts. The first part examines the problem of optimal channel estimation w.r.t. the mean square error (MSE) in multiple antennae systems. It is used to call these systems Multiple Input Multiple Output (MIMO), although the term MIMO may refer to any system (e.g. a filter) that possess multiple inputs and multiple outputs. The great interest on these systems is based on the fact that theoretical results of the current and the previous decade showed that these systems can increase greatly the information rate without sacrificing spectral efficiency or increasing the power. Accurate knowledge of the channel affects the achievable capacity. Two estimators with their optimal training sequences are proposed in this context: The *Bayesian Minimum Disturbance Estimator* (BMDE) and the *Two Sided Bayesian Minimum Disturbance Estimator* (2S BMDE). For the latter estimator, it is shown that it has less feedback requirements than the former with a comparable performance. The second part deals with the same problem but in multicarrier (MC) systems. The Cyclic Prefixed Orthogonal Frequency Division Multiplexing (CP-OFDM) system transmitting a Quadrature Amplitude Modulation (QAM) and the OFDM system implemented based on the Offset QAM modulation (OFDM/OQAM) are compared from the aspect of preamble-based channel estimation performance. To this end, three different preambles are defined: the *sparse*, the *full* and the *sparse-data*. Detailed definitions of these are given later on. It is showed that the sparse sparse preamble containing L_h equispaced and equal or equipowered pilot tones for the CP-OFDM and OFDM/OQAM systems, respectively, is the optimal one when all preambles transmit the same training power. Then, the sparse preambles for these systems are compared and it turns out that the optimal sparse preamble for the OFDM/OQAM performs $3-9$ dB better that the corresponding optimal preamble for the CP-OFDM system. Additional unknown results accompanied the above analysis.

2 MIMO Signal Model

Assume that we have a MIMO system with M_T transmit and M_R receive antennae for any use[1]. There L interfering sources in the communication between the desired user and the desired receiver. All the MIMO channels involved are assumed to be quasi-stationary for the duration of one block, while the can change randomly from a block to another (block fading). These channels are also considered to be narrowband (flat fading). The nth received vector (n is a temporal

[1] This assumption can be eliminated. It is maintained here for simplicity.

index) is given by:

$$\boldsymbol{y}(n) = \boldsymbol{H}_0\boldsymbol{x}_0(n) + \sum_{i=1}^{L} \boldsymbol{H}_i\boldsymbol{x}_i(n) + \boldsymbol{w}(n) \tag{1}$$

$\boldsymbol{y}(n), \boldsymbol{w}(n)$ are $M_R \times 1$ vectors, $\boldsymbol{x}_i(n), i = 0, 1, \ldots, L$ are $M_T \times 1$ vectors and $\boldsymbol{H}_i, i = 0, 1, \ldots, L$ are $M_R \times M_T$ channel matrices. The quantities with index $i = 0$ correspond to the desired user and the rest of the indices to the interferers. $\boldsymbol{H}_i$ contains the channel gains between the receiver and the ith user. $\boldsymbol{x}_i(n)$ is the vector signal transmitted at the nth time instant by the array of the ith user. We assume that the transmitted signals have zero mean and they are spatially and/or temporally correlated. They are considered uncorrelated among them and w.r.t. the additive noise vector $\boldsymbol{w}(n)$. Additionally, $\boldsymbol{w}(n)$ is considered to have zero mean.

3 The Unvectorized Linear Minimum Mean Square Error (LMMSE) Estimator

Suppose that M time slots per frame are devoted for training. Grouping together the received vectors in a $M_R \times M$ matrix, the aggregated signal model becomes:

$$\begin{aligned}
\boldsymbol{Y} &= \left[\boldsymbol{y}(n)\ \boldsymbol{y}(n-1) \cdots \boldsymbol{y}(n-M+1)\right] \\
&= \boldsymbol{H}_0 \left[\boldsymbol{x}_0(n)\ \boldsymbol{x}_0(n-1) \cdots \boldsymbol{x}_0(n-M+1)\right] \\
&+ \underbrace{\left[\boldsymbol{H}_1\ \boldsymbol{H}_2 \cdots \boldsymbol{H}_L\right]}_{\boldsymbol{H}_{\text{int}}} \underbrace{\begin{bmatrix} \boldsymbol{x}_1(n) & \boldsymbol{x}_1(n-1) & \cdots & \boldsymbol{x}_1(n-M+1) \\ \boldsymbol{x}_2(n) & \boldsymbol{x}_2(n-1) & \cdots & \boldsymbol{x}_2(n-M+1) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_L(n) & \boldsymbol{x}_L(n-1) & \cdots & \boldsymbol{x}_L(n-M+1) \end{bmatrix}}_{\boldsymbol{X}_{\text{int}}} \\
&+ \underbrace{\left[\boldsymbol{w}(n)\ \boldsymbol{w}(n-1) \cdots \boldsymbol{w}(n-M+1)\right]}_{\boldsymbol{W}}
\end{aligned}$$

or

$$\boldsymbol{Y} = \boldsymbol{H}_0\boldsymbol{X}_0 + \underbrace{\boldsymbol{H}_{\text{int}}\boldsymbol{X}_{\text{int}} + \boldsymbol{W}}_{\boldsymbol{\mathcal{E}}} = \boldsymbol{H}_0\boldsymbol{X}_0 + \boldsymbol{\mathcal{E}} \tag{2}$$

where the matrix $\boldsymbol{\mathcal{E}} = \boldsymbol{H}_{\text{int}}\boldsymbol{X}_{\text{int}} + \boldsymbol{W}$ contains the total interference (multiuser interference plus thermal noise) at the receiver for the duration of time slots. The matrix $\boldsymbol{X}_0$ is assumed to be known to the receiver and it will be called the *training matrix*.

Our goal is to estimate $\boldsymbol{H}_0$. We need as many measurements as our unknowns. Thus, we assume that $\geq M_T$. Suppose that $\boldsymbol{C}$ is the estimating filter of size $M \times M_T$, which is applied to the matrix $\boldsymbol{Y}$ from the *right*. The minimum mean square error (MMSE) criterion can be formulated here with the use of the Frobenius norm as follows:

$$\min_{\boldsymbol{C}} E\left[\|\boldsymbol{Y}\boldsymbol{C} - \boldsymbol{H}_0\|_F^2\right] \tag{3}$$

and the solution, $\boldsymbol{C}$, is given by the Wiener filter [8, 2]:

$$\boldsymbol{C} = \boldsymbol{R}_{\boldsymbol{Y}_0}^{-1} \boldsymbol{R}_{\boldsymbol{Y}\boldsymbol{H}_0} \tag{4}$$

where $\boldsymbol{R}_{\boldsymbol{Y}} = E\left[\boldsymbol{Y}^H \boldsymbol{Y}\right]$ and $\boldsymbol{R}_{\boldsymbol{Y}\boldsymbol{H}_0} = E\left[\boldsymbol{Y}^H \boldsymbol{H}_0\right]$. Using the eq. (2), we can derive $\boldsymbol{C}$ as:

$$\boldsymbol{C} = \boldsymbol{R}_{\boldsymbol{\mathcal{E}}}^{-1} \boldsymbol{X}_0^H \left(\boldsymbol{X}_0 \boldsymbol{R}_{\boldsymbol{\mathcal{E}}}^{-1} \boldsymbol{X}_0^H + \boldsymbol{R}_{\boldsymbol{H}_0}^{-1}\right)^{-1} \tag{5}$$

where $\boldsymbol{R}_{\boldsymbol{\mathcal{E}}} = E\left[\boldsymbol{\mathcal{E}}^H \boldsymbol{\mathcal{E}}\right]$ and $\boldsymbol{R}_{\boldsymbol{H}_0} = E\left[\boldsymbol{H}_0^H \boldsymbol{H}_0\right]$. Also,

$$\min_{\boldsymbol{C}} E\left[\|\boldsymbol{Y}\boldsymbol{C} - \boldsymbol{H}_0\|_F^2\right] = \operatorname{tr}\left[\left(\boldsymbol{X}_0 \boldsymbol{R}_{\boldsymbol{\mathcal{E}}}^{-1} \boldsymbol{X}_0^H + \boldsymbol{R}_{\boldsymbol{H}_0}^{-1}\right)^{-1}\right] \tag{6}$$

4 The BMDE

The $M_T \times M$ training matrix $\boldsymbol{X}_0$, as it is incorporated in the signal model (2), can be designed in such a way that the MMSE, which occurs by the processing of the received data by $\boldsymbol{C}$ to be reduced further. To eliminate the possibility of trivial solutions, we impose a training energy constraint on $\boldsymbol{X}_0$. An optimization problem can be formulated in this way, which is given by:

$$\min_{\boldsymbol{X}_0} \operatorname{tr}\left[\left(\boldsymbol{X}_0 \boldsymbol{R}_{\boldsymbol{\mathcal{E}}}^{-1} \boldsymbol{X}_0^H + \boldsymbol{R}_{\boldsymbol{H}_0}^{-1}\right)^{-1}\right] \tag{7}$$

$$\text{s.t.} \qquad \operatorname{tr}\left[\boldsymbol{X}_0 \boldsymbol{X}_0^H\right] \leq E_T \tag{8}$$

where E_T is the available training energy. We observe that the objective function (7) equals the sum of the inverse eigenvalues of $\boldsymbol{X}_0 \boldsymbol{R}_{\boldsymbol{\mathcal{E}}}^{-1} \boldsymbol{X}_0^H + \boldsymbol{R}_{\boldsymbol{H}_0}^{-1}$. Suppose that $\boldsymbol{R}_{\boldsymbol{H}_0} = \boldsymbol{Q}\boldsymbol{K}\boldsymbol{Q}^H$, where $\boldsymbol{Q}$ is a unitary matrix and $\boldsymbol{K} = \operatorname{diag}(\kappa_1, \kappa_2, \cdots, \kappa_{M_T})$ is the EVD of channel autocorrelation matrix. In addition, $\boldsymbol{R}_{\boldsymbol{\mathcal{E}}} = \boldsymbol{G}\boldsymbol{\Lambda}\boldsymbol{G}^H$, where $\boldsymbol{G}$ is unitary and $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_M)$, is the interference autocorrelation. We assume that $\kappa_1 \geq \kappa_2 \geq \cdots \geq \kappa_{M_T}$ and that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_M$. The eigenvalues of $\boldsymbol{X}_0 \boldsymbol{R}_{\boldsymbol{\mathcal{E}}}^{-1} \boldsymbol{X}_0^H$ are $\mu_1, \mu_2, \ldots, \mu_{M_T}$ and we assume that $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_{M_T}$. The selection of the optimal training matrix for problem (7)-(8) is summarized bellow:

Proposition 1 *The optimal training matrix for the problem (7)-(8) is:*

$$\boldsymbol{X}_0^* = \boldsymbol{Q} \begin{bmatrix} \sqrt{\mu_1^* \lambda_1} & 0 & \cdots & 0 & 0 \cdots 0 \\ 0 & \sqrt{\mu_1^* \lambda_1} & \cdots & 0 & 0 \cdots 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \ddots \vdots \\ 0 & 0 & \cdots & \sqrt{\mu_{M_T}^* \lambda_{M_T}} & 0 \cdots 0 \end{bmatrix} \boldsymbol{G}^H \tag{9}$$

where

$$\mu_i^* = \begin{cases} \frac{E_T + \sum_{j=1}^{m^*} \lambda_j / \kappa_j}{\sqrt{\lambda_i} \sum_{j=1}^{m^*} \sqrt{\lambda_j}} - \frac{1}{\kappa_i}, & i = 1, 2, \ldots, m^* \\ 0, & i = m^* + 1, \ldots, M_T \end{cases} \tag{10}$$

are the optimal values for the mu_i's and m^* is chosen so that all the μ_i's to be positive.

$$m^* = \max\left\{ m \in \{1, 2, \ldots, M_T\} : \frac{\sqrt{\lambda_m}}{\kappa_m} \sum_{i=1}^{m} \sqrt{\lambda_i} - \sum_{i=1}^{m} \frac{\lambda_i}{\kappa_i} < E_T \right\} \tag{11}$$

5 The Unvectorized LMMSE in Two Steps

The unvectorized LMMSE in the last section can be seen to occur as a two-step procedure. In the first step, the interference is minimized under a *zero forcing* (ZF) constraint, which aims at the preservation of the desired channel after the minimization has taken place. We call this first filter $\boldsymbol{C}_1$ and it is a $M \times M_T$ matrix, which applies to $\boldsymbol{Y}$ from the *right*:

$$\boldsymbol{Y}_1 \stackrel{\triangle}{=} \boldsymbol{Y}\boldsymbol{C}_1 = \boldsymbol{H}_0\boldsymbol{X}_0\boldsymbol{C}_1 + \boldsymbol{\mathcal{E}}\boldsymbol{C}_1 = \boldsymbol{H}_0\boldsymbol{X}_0\boldsymbol{C}_1 + \boldsymbol{\mathcal{E}}_1 \tag{12}$$

where $\boldsymbol{\mathcal{E}} \stackrel{\triangle}{=} \boldsymbol{\mathcal{E}}\boldsymbol{C}_1$ is the residual interference at the filter's output. The problem of the optimal selection of $\boldsymbol{C}_1$ is formulated as follows:

$$\min_{\boldsymbol{C}_1} E(\|\boldsymbol{\mathcal{E}}_1\|_F^2)$$
$$\text{s.t. } \boldsymbol{X}_0\boldsymbol{C}_1 = \boldsymbol{I}_{M_T}$$

Taking the constraint into account, we have:

$$\boldsymbol{Y}_1 = \boldsymbol{H}_0 + \boldsymbol{\mathcal{E}}_1 \tag{13}$$

The last equation shows that $\boldsymbol{Y}_1$ is a first estimate of the unknown channel and it holds: $E\left[\boldsymbol{Y}_1\right] = \boldsymbol{H}_0$. I.e. in this first step, the unknown channel is considered to be deterministic.

The solution of the last problem is known to be given by [2]:

$$\boldsymbol{C}_1 = \boldsymbol{R}_{\boldsymbol{\mathcal{E}}}^{-1}\boldsymbol{X}_0^H(\boldsymbol{X}_0\boldsymbol{R}_{\boldsymbol{\mathcal{E}}}^{-1}\boldsymbol{X}_0^H)^{-1} \tag{14}$$

if and only if $\boldsymbol{X}_0$ is full row rank.

We can further refine this first estimate using a second filter $\boldsymbol{C}_2$, which is chosen through the unconstrained problem $E[\|\boldsymbol{Y}_1\boldsymbol{C}_2 - \boldsymbol{H}_0\|^2]$ as a $M_T \times M_T$ matrix. This filter incorporates reduces the MSE by incorporating in the final channel estimate our knowledge about the statistics of the channel. The solution of the last problem is the Wiener filter [2, 8]:

$$\boldsymbol{C}_2 = \boldsymbol{R}_{\boldsymbol{Y}_1}^{-1}\boldsymbol{R}_{\boldsymbol{Y}_1\boldsymbol{H}_0} \tag{15}$$

where $\boldsymbol{R}_{\boldsymbol{Y}_1} \stackrel{\triangle}{=} E(\boldsymbol{Y}_1^H\boldsymbol{Y}_1)$ and $\boldsymbol{R}_{\boldsymbol{Y}_1\boldsymbol{H}_0} \stackrel{\triangle}{=} E(\boldsymbol{Y}_1^H\boldsymbol{H}_0)$. Using the above equations, it is:

$$\boldsymbol{C}_2 = [\boldsymbol{R}_{\boldsymbol{H}_0} + (\boldsymbol{X}_0\boldsymbol{R}_{\boldsymbol{\mathcal{E}}}\boldsymbol{X}_0^H)^{-1}]^{-1}\boldsymbol{R}_{\boldsymbol{H}_0} \tag{16}$$

We can easily show that the composition of the two steps leads to the unvectorized LMMSE estimator:

Proposition 2 *The composition of the above two steps is given by the expression:*

$$\boldsymbol{C}_1\boldsymbol{C}_2 = \boldsymbol{R}_{\boldsymbol{\mathcal{E}}}^{-1}\boldsymbol{X}_0^H(\boldsymbol{X}_0\boldsymbol{R}_{\boldsymbol{\mathcal{E}}}^{-1}\boldsymbol{X}_0^H + \boldsymbol{R}_{\boldsymbol{H}_0}^{-1})^{-1} = \boldsymbol{C}$$

5.1 The 2S BMDE

In the last section, the filters $\boldsymbol{C}_1, \boldsymbol{C}_2$ are applied on the measurement matrix from the *right*. Observing eq. (13), one can easily verify that the Wiener filter $\boldsymbol{C}_2$ does not need to be applied on $\boldsymbol{Y}_1$ from the right. We can define a corresponding filter $\bar{\boldsymbol{C}}_2$ which will be applied on $\boldsymbol{Y}_1$ from the *left* side. This modification leads to an alternative unvectorized LMMSE estimator and allows a great reduction of the feedback requirements of the BMDE. It also reduces the transmitter's and receiver's complexity, while it usually maintains a comparable perfromance.

The first step filter remains as it is. The filter of the second step will be given by:

$$\bar{\boldsymbol{Y}}_2 = \bar{\boldsymbol{C}}_2\boldsymbol{Y}_1 = \bar{\boldsymbol{C}}_2\boldsymbol{H}_0 + \bar{\boldsymbol{C}}_2\boldsymbol{\mathcal{E}}_1$$

The selection of $\bar{\boldsymbol{C}}_2$ is performed via the following optimization problem:

$$\min_{\bar{\boldsymbol{C}}_2} E\left(\|\bar{\boldsymbol{C}}_2\boldsymbol{Y}_1 - \boldsymbol{H}_0\|^2\right)$$

The solution of this problem is again the Wiener filter, which now has the form:

$$\bar{\boldsymbol{C}}_2 = \bar{\boldsymbol{R}}_{\boldsymbol{Y}_1\boldsymbol{H}_0}\bar{\boldsymbol{R}}_{\boldsymbol{Y}_1}^{-1}, \tag{17}$$

where $\bar{\boldsymbol{R}}_{\boldsymbol{Y}_1} \triangleq E\left(\boldsymbol{Y}_1\boldsymbol{Y}_1^H\right)$ and $\bar{\boldsymbol{R}}_{\boldsymbol{Y}_1\boldsymbol{H}_0} \triangleq E\left(\boldsymbol{Y}_1\boldsymbol{H}_0^H\right)$. After some algebra,

$$\bar{\boldsymbol{C}}_2 = \bar{\boldsymbol{R}}_{\boldsymbol{H}_0}\left(\bar{\boldsymbol{R}}_{\boldsymbol{H}_0} + \bar{\boldsymbol{R}}_{\boldsymbol{\mathcal{E}}_1}\right)^{-1} \tag{18}$$

We can also write:

$$\bar{\boldsymbol{C}}_2 = \bar{\boldsymbol{R}}_{\boldsymbol{H}_0}\left(\bar{\boldsymbol{R}}_{\boldsymbol{\mathcal{E}}_1}^{-1}\bar{\boldsymbol{R}}_{\boldsymbol{H}_0} + \boldsymbol{I}_{M_R}\right)^{-1}\bar{\boldsymbol{R}}_{\boldsymbol{\mathcal{E}}_1}^{-1} = \left(\bar{\boldsymbol{R}}_{\boldsymbol{H}_0}^{-1} + \bar{\boldsymbol{R}}_{\boldsymbol{\mathcal{E}}_1}^{-1}\right)^{-1}\bar{\boldsymbol{R}}_{\boldsymbol{\mathcal{E}}_1}^{-1}$$

The MSE will be given by:

$$\begin{aligned}\min E\left(\|\bar{\boldsymbol{C}}_2\boldsymbol{Y}_1 - \boldsymbol{H}_0\|^2\right) &= \mathrm{tr}\left[\bar{\boldsymbol{R}}_{\boldsymbol{H}_0} - \bar{\boldsymbol{R}}_{\boldsymbol{H}_0}(\bar{\boldsymbol{R}}_{\boldsymbol{H}_0} + \bar{\boldsymbol{R}}_{\boldsymbol{\mathcal{E}}_1})^{-1}\bar{\boldsymbol{R}}_{\boldsymbol{H}_0}\right] \\ &= \mathrm{tr}\left[(\bar{\boldsymbol{R}}_{\boldsymbol{H}_0}^{-1} + \bar{\boldsymbol{R}}_{\boldsymbol{\mathcal{E}}_1}^{-1})^{-1}\right] \end{aligned} \tag{19}$$

where we have used the matrix inversion lemma.

The optimal training matrix can be selected as:

$$\min_{\boldsymbol{X}_0} \mathrm{tr}\left[(\bar{\boldsymbol{R}}_{\boldsymbol{H}_0}^{-1} + \bar{\boldsymbol{R}}_{\boldsymbol{\mathcal{E}}_1}^{-1})^{-1}\right] \tag{20}$$

$$\text{s.t.}\ \ \mathrm{tr}\left[\boldsymbol{X}_0\boldsymbol{X}_0^H\right] \leq E_T \tag{21}$$

In the general case, the dependence on trace operators makes the last problem very difficult to be solved analytically. To achieve an analytical solution we will relax the above problem to:

$$\min_{\boldsymbol{X}_0} \operatorname{tr}\left(\bar{\boldsymbol{R}}_{\boldsymbol{\mathcal{E}}_1}\right) \tag{22}$$

$$\text{s.t. } \operatorname{tr}\left(\boldsymbol{X}_0 \boldsymbol{X}_0^H\right) \leq E_T \tag{23}$$

The solution of this problem is summarized below:

Theorem 1 *A class of training matrices which optimizes the criterion (22), (23) is given by the expression:*

$$\bar{\boldsymbol{X}}_0^* = \boldsymbol{U} \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 & 0 \cdots 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 \cdots 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \ddots \vdots \\ 0 & 0 & \cdots & \sigma_{M_T} & 0 \cdots 0 \end{pmatrix} \boldsymbol{G}^H \tag{24}$$

where

$$\sigma_i = \sqrt{\frac{\sqrt{\lambda_i}}{\sum_{j=1}^{M_T} \sqrt{\lambda_j}} E_T}, \quad i = 1, 2, \ldots, M_T \tag{25}$$

and $\boldsymbol{U}$ any *unitary* $M_T \times M_T$ *matrix.*

6 CP-OFDM and OFDM/OQAM: A Fair Comparison Framework

Suppose that we want to built a transmission system based on blocks of data. If we desire to estimate the channel with the use of training sequences, then at the beginning of each block we introduce some symbols or a priori known data vectors devoted for training. This part of the block is called *preamble*.

For the CP-OFDM system, the preamble will consist of a complex training vector, which is a common assumption in the litterature [9]. Each complex CP-OFDM symbols is equivalent to two real vector symbols in the OFDM/OQAM system. We consider an equivalent preamble for the OFDM/OQAM system, which consists of a nonzero first training vector followed by a zero second vector.

Clarifying the above, we have:

Definition 1 *Assume that a preamble consists of a certain number of training vectors and one among them is nonzero. This nonzero vector will be called* preamble vector*.*

Assume that T_1, T_2 are the sampling periods at the output of the SFBs for the two OFDM systems[2] and that the minimum possible number of output samples

[2] Either CP-OFDM or OFDM/OQAM

at the SFB output for the reconstruction of the preamble vector at the receiver is R_1, R_2 for each system respectively. Then the following quantity will be useful in the definition of a fair comparison framework of different preambles:

Definition 2 *The Training Power Ratio (TPR) for the preambles of the two systems is defined as:*

$$\text{TPR}^{\boldsymbol{p}_1,\boldsymbol{p}_2} = \frac{\frac{1}{R_1 T_1}\mathcal{E}^{\boldsymbol{p}_1}}{\frac{1}{R_2 T_2}\mathcal{E}^{\boldsymbol{p}_2}}$$

where $\mathcal{E}^{\boldsymbol{p}_1}, \mathcal{E}^{\boldsymbol{p}_2}$ are the energies of the corresponding preambles at the SFB outputs for the minimum required sample numbers R_1, R_2 respectively.

We observe that if we define different preambles within the same system, then $T_1 = T_2$ and the sampling periods can be neglected from the last definition. We can now define a fair comparison framework.

Suppose that $\boldsymbol{p}_1, \boldsymbol{p}_2$ are two different preambles in two systems. To guarantee a fair comparison among these preambles, it necessary that $\text{TPR}^{\boldsymbol{p}_1,\boldsymbol{p}_2} = 1$. In other words, we require that both systems put the amount of power in the training data at the transmit antenna.

7 Summary of Definitions and Useful Results

We use the assumption that the ratio of the number of subcarriers, M, in the systems over the channel length L_h is an integer number and a power of two[3] If M is selected to be a power of 2, the last can always be achieved through zero padding of the channel.

Definition 3 Sparse *preamble vector is a $M \times 1$ training vector, which consists of L_h isolated pilot tones and nulls to the rest of the positions.*

Definition 4 *A preamble vector will be called* full *if it contains pilot symbols on all subcarriers.*

Definition 5 *A preamble vector with L_h isolated pilot symbols and data symbols to the rest of the positions will be called* sparse-data *preamble vector.*

The rest of the results will be useful.

Theorem 2 *For the CP-OFDM system, the sparse preamble that minimizes the MSE of the channel frequency response (CFR) estimates subject to an energy constraint both on the useful part of the signal as well as on the CP is the one containing* equispaced *and* equal *pilot symbols.*

Even if the last result holds for the full preamble vector as well, we can show the following result:

[3] FFT implementation of the DFT.

Theorem 3 *There full preambles which are MSE-optimal subject to an energy constraint both on the useful part of the signal and the CP which contain simply* equipowered *symbols.*

For the OFDM/OQAM system, the corresponding results are:

Theorem 4 *For the OFDM/OQAM, the sparse preamble vector which minimizes the MSE of the CFR estimates subject to an enery constraint is built with* equispaced *and* equipowered *pilot symbols.*

Theorem 5 *Full OFDM/OQAM preamble vectors with equal symbols are locally MSE-optimal under an energy constraint. Their global optimality can be shown when the enery constraint is translated to the SFB input.*

Result: It can be proved that the sparse preambles are MSE-optimal when compared to full and sparse-data preambles carrying the same training energy. Therefore, the comparison between CP-OFDM and OFDM/OQAM can be performed based on their optimal sparse preambles under the same transmit training power.

8 OFDM/OQAM Sparse Preamble vs. CP-OFDM Sparse Preamble

We may neglect the mathematical analysis for brevity and simply give the final expression that relates the MSE's achieved by the optimal sparse preambles of the two systems, when both systems transmit the same training power:

$$\mathrm{MSE}^{\mathrm{s}}_{\mathrm{OFDM/OQAM}} = \frac{M + L_h - 1}{L_g} \mathrm{MSE}^{\mathrm{s}}_{\mathrm{CP-OFDM}}$$

where L_g is the length of the prototype pulse used by the OFDM/OQAM system and it is usually given by $L_g = KM$, where $K \in \{1, 2, 3, 4, 5\}$ is called *overlapping factor*. For example, suppose that $L_h = 32$. Then, for $L_g = M$, the CP-OFDM sparse preamble is better than the corresponding OFDM/OQAM sparse preamble, but for $L_g = KM$ with $2 \leq K \leq 5$, the OFDM/OQAM sparse preamble is $3 - 9$ dB better.

9 Simulations

9.1 BMDE and 2S BMDE

The performance of the analyzed schemes is demonstrated in this section via a simulation result for a wide range of SINRs and for a certain system setup due to space limitations. We assume three interferers (i.e., $L = 3$) with transmit array size (M_T) the same as that of the desired user. All channel matrices are generated as [12]

$$\boldsymbol{H}_i = \boldsymbol{R}_{r_i}^{1/2} \boldsymbol{H}_{w_i} \boldsymbol{R}_{t_i}^{1/2},$$

where $\boldsymbol{H}_{w_i}$, $i = 0, 1, \ldots, L$ is an uncorrelated channel, represented by an $M_R \times M_T$ matrix of i.i.d. zero mean, unit variance, circularly symmetric complex Gaussian entries, and $\boldsymbol{R}_{t_i}^{1/2}$ and $\boldsymbol{R}_{r_i}^{1/2}$ are Hermitian square roots of the ith transmit and receive fade correlation matrices, respectively. The fading correlation matrices follow the exponential model [**?**], that is, they are built as Hermitian with entries:

$$\boldsymbol{R}_{i,j} = r^{j-i}, \quad j \geq i$$

where r is the (complex) normalized correlation coefficient with magnitude $\rho = |r| < 1$. The correlation strengths for the receive and transmit sides of the desired channel will be denoted by ρ_r and ρ_t, respectively. Each interfering source is generated as a linearly filtered i.i.d. QPSK sequence.

Remark that the 2S B-MDE is the only one, among the schemes under study, which employs the *receive* fading correlation matrix. This feature suggests that there might be cases where the B-MDE is outperformed by the two-sided scheme, namely when $\rho_r \gg \rho_t$. In Fig. 1 we demonstrate such a scenario. These results show that in cases that the receive correlation is significantly stronger than the correlation at the transmit side, the 2S B-MDE can provide the best performance of all the schemes, while being also the most economic one in the amount of information to be fed back. This holds for both the optimal schemes and their sub-optimal variants.

Remark: In this figure, the GM/LS estimator is a common name for the Gauss-Markov and Least Squares Estimators which is shown in this thesis that for their optimal training they coincide. The suboptimal variants of the schemes use DFT matrices of proper dimensions as the modal matrices of the interference and channel correlation matrices.

9.2 CP-OFDM vs. OFDM/OQAM

The sparse preambles for the CP-OFDM and OFDM/OQAM systems are compared in Fig. 2. The superiority of the OFDM/OQAM sparse preamble, when the entire transmit pulse duration is considered, is evident. The analytical results can be seen to be approximately verified. Thus, for Fig. 2a, the theoretical difference is $10 \log_{10} [KM/(M + L_h - 1)] \approx 4.5$ dB, while for Fig. 2b, it is approximately 5.9 dB. These values agree with the difference of the experimental curves.

References

1. G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications,* vol. 6, pp. 311–335, 1998.
2. T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, Prentice-Hall, 2000.
3. T. Kaiser *et al.* (eds.), *Smart Antennas – State of the Art,* Hindawi, 2005.
4. D. Katselis, E. Kofidis, and S. Theodoridis, "Training-based estimation of correlated MIMO fading channels in the presence of colored interference," *Signal Processing*, vol. 87, pp. 2177–2187, Sept. 2007.

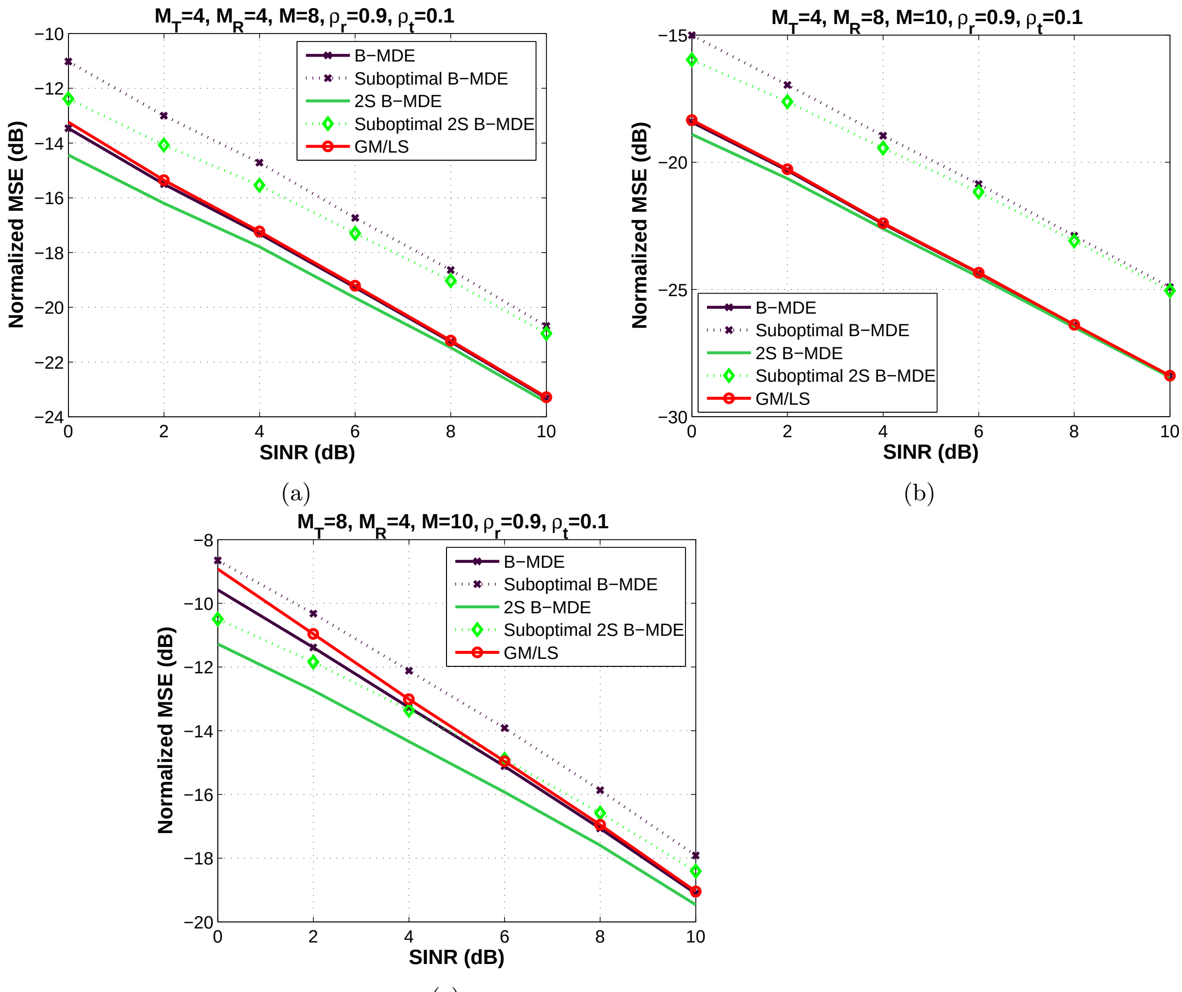
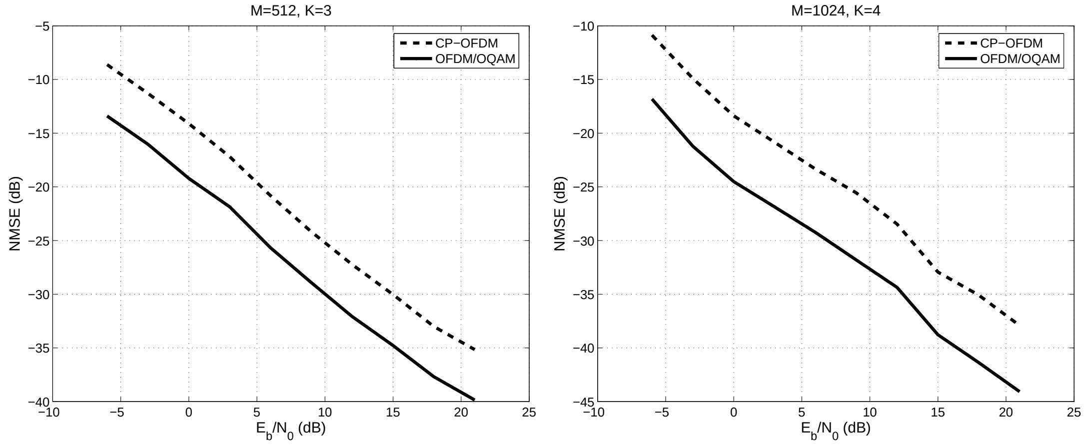


Fig. 1. MSE performance of the estimators for channels with strong receive ($\rho_r = 0.9$) and weak transmit ($\rho_t = 0.1$) correlation.



(a) (b)

Fig. 2. NMSE performance of the CP-OFDM and OFDM/OQAM sparse preambles: (a) $M = 512$, $K = 3$; (b) $M = 1024$, $K = 4$.

5. ——, “A low-cost approach to training-based MIMO channel estimation in interference-limited environments,” *Proc. SPAWC-2007,* Helsinki, Finland, June 2007.
6. ——, ”On training optimization for estimation of correlated MIMO channels in the presence of multiuser interference”, IEEE trans on Signal Processing, Volume: 56, 2008, Page(s): 4892-4904.
7. ——, ”A low-cost approach to training-based MIMO channel estimation in interference-limited environments”, Proc. SPAWC-2007, Helsinki, Finland, June 2007.
8. S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory,* Prentice-Hall, 1993.
9. M. Morelli and U. Mengali, “A comparison of pilot-aided channel estimation methods for OFDM systems,” *IEEE Trans. Signal Processing,* vol. 49, no. 12, pp. 3065–3073, Dec. 2001.
10. J. G. Proakis, *Digital Communications*, McGraw-Hill, 3rd edition, 1995.
11. I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Trans. on Telecommunications (ETT)*, vol. 10, no. 6, pp. 585-595, Nov.-Dec. 1999. (See also a previous version of the paper in AT& T Bell Labs Internal Tech. Memo, June 1995.).
12. K. Yu, M. Bengtsson, and B. Ottersten, “MIMO channel models,” pp. 271–292 in [3].