CRAMER-RAO BOUNDS FOR MIMO CHANNEL ESTIMATION

Lamia Berriche, Karim Abed-Meraim, and Jean-Claude Belfiore

Ecole Nationale Superieure des Telecommunications
46 rue Barrault, 75634 Paris Cedex 13, France

ABSTRACT

We investigate the performances of pilot-aided channel estimation and data detection for Multi-Input Multi-Output (MIMO) systems. We analyze and compare the Cramer-Rao Bound (CRB) on channel taps estimation error for different design models of the pilot sequences. Then, Minimum Mean Square Error (MMSE) estimation method with deflation is performed for channel identification for the embedded pilot scheme.

1. INTRODUCTION

The potential capacity of MIMO channels has recently been recognized [1]. High data rates with low error probabilities could be achieved especially when the wireless channel response is known at the receiver. Pilot symbols insertion had been introduced to make channel estimation for coherent detection. Accuracy of the channel estimator depends on the pilot symbol design and placement [2]. Many studies have been performed to find out the optimal design according to different criteria such as maximizing the channel capacity [3], minimizing the MSE or minimizing the CRB [2]. The objective of this paper is to compare the CRB for different pilot insertion schemes, and to introduce a channel estimation method for the embedding pilot case.

2. PROBLEM STATEMENT

2.1. Data Model

Assume that we have K-input and M-output channels. The channels are assumed to be frequency-selective block-fading, i.e., the random channels taps remain constant for some data packets and change to independent values for the next. The baseband received signal can be written as:

$$\mathbf{y}(l) = [y_1(l), ..., y_M(l)]^T = \sum_{k=1}^{K} \sum_{i=0}^{L_k} \mathbf{h}_k(i) s_k(l-i) + \mathbf{n}(l)$$

where $\mathbf{h}_k(i) = [h_{k,1}(i) \dots h_{k,M}(i)]^T \in \mathcal{C}^M$ for $i = 0 \dots L_k$ is the channel impulse response vector from the k^{th} transmit antenna to the M receive antennae and L_k is the channel order for all M subchannels. For simplicity, we assume the same channel length for all users, i.e, $L_k = L$ for all k. s_k denotes the transmitted symbol from the k^{th} antenna and $\mathbf{n}(l)$ represents additive noise. We consider the stretch of output vector $\mathbf{y} = [\mathbf{y}(N)^T \dots \mathbf{y}(1)^T]^T$, N being the sample size

$$\mathbf{y} = \mathcal{T}(H)\mathbf{s} + \mathbf{n} \tag{1}$$

with $\mathbf{s} = [\mathbf{s}_1^T \dots \mathbf{s}_K^T]^T$ where $\mathbf{s}_k$ is the N' dimensional vector ($N' = N + L$) defined by $\mathbf{s}_k = [s_k(N), \cdots, s_k(1-L)]^T$ and $\mathcal{T}(H) = [\mathcal{T}(H_1) \dots \mathcal{T}(H_K)]$ where $\mathcal{T}(H_k)$ denotes the block Sylvester matrix associated to the k^{th} channel which is given by $\mathbf{H}_k = [\mathbf{h}_k(0), \ \cdots, \ \mathbf{h}_k(L)]$. Due to the commutativity of convolution, the stretched received vector could as well be expressed by the following expression:

$$\mathbf{y} = \mathcal{S}\mathbf{h} + \mathbf{n} \tag{2}$$

where

$$\begin{aligned} \mathbf{h} &= [\mathbf{h}_1^T, \dots, \mathbf{h}_K^T]^T \text{ with } \mathbf{h}_k = vec(\mathbf{H}_k) \\ \mathcal{S} &= [\mathbf{S}_1 \ \dots \ \mathbf{S}_K] \otimes \mathbf{I}_M \end{aligned}$$

$$\mathbf{S}_k = \begin{pmatrix} s_k(N) & \cdots & s_k(N-L) \\ s_k(N-1) & \cdots & s_k(N-L-1) \\ \vdots & & \vdots \\ s_k(1) & \dots & s_k(1-L) \end{pmatrix} \tag{3}$$

$\otimes$ denotes the Kronecker product and $vec(.)$ is the column vectorization operator. In the sequel, we make the following assumptions:

1. Data symbols are i.i.d. with probability density function[1] $p_{s_d}(.)$ with zero mean and variance σ_d^2
2. Taps of the channel $\mathbf{h}$ are random i.i.d. with probability density function $p_h(.)$ and variance σ_h^2.
3. The noise vector $\mathbf{n}$ is complex circular white gaussian with zero mean, and variance σ_n^2 per complex dimension.
4. The noise vector $\mathbf{n}$, the channel taps $\mathbf{h}$ and data symbols are jointly independent.

2.2. Pilot Design

We propose to analyze and compare the effect of pilot symbol design on the CRB. We will calculate the performances of the traditional scheme in which pilot symbols are time-multiplexed with data symbols, and the so called embedded scheme in which pilots are transmitted simultaneously with linearly precoded or non precoded data.

2.2.1. *Time-multiplexed pilot and data*

This is the conventionally used pilot insertion method where each antenna transmits a sequence of P pilot symbols $s_{pk}(i)$ and a sequence of $N' - P$ data symbols $s_{dk}(i)$ (see Fig. 1). One desired

[1] For simplicity, we assume the same power and probability distribution for all users.

property of such pilot sequences is the orthogonality which is restrictive in practice, especially for short pilot sequences [2]. Using the scheme of staggered small pilot clusters simplifies the pilot design in MIMO systems [2]. However, the channel estimation problem becomes more difficult. The optimal placement for time-multiplexed pilot and data is proposed in [2]. If we assume that

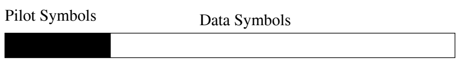

Fig. 1. Time multiplexed pilot and data symbols

$s_{pk}(i) = 0$ (resp. $s_{dk}(i) = 0$) when a data (resp. pilot) symbol is transmitted, the i^{th} transmitted symbol from the k^{th} antenna can be expressed as:

$$s_k(i) \quad = \quad s_{dk}(i) + s_{pk}(i) \tag{4}$$

Hence, the FIR channels response in (1) and (2) could be expressed as follows:

$$\mathbf{y} \quad = \quad \mathcal{T}(H)\mathbf{s}_d + \mathcal{T}(H)\mathbf{s}_p + \mathbf{n} \tag{5}$$
$$\quad = \quad (\mathcal{S}_d + \mathcal{S}_p)\mathbf{h} + \mathbf{n} \tag{6}$$

where $\mathbf{s}_d$ and $\mathbf{s}_p$ (resp. $\mathcal{S}_d$ and $\mathcal{S}_p$) are the vectors (resp. the block Hankel matrices, i.e., $\mathcal{S}_d = \mathbf{S}_d \otimes \mathbf{I}_M$ and $\mathcal{S}_p = \mathbf{S}_p \otimes \mathbf{I}_M$) corresponding to the data symbols and to the pilot symbols, respectively.

2.2.2. Embedded Pilot

In this case, data symbols $s_{dk}(i)$ are sent during N' symbol periods simultaneously with pilot symbols $s_{pk}(i)$ (see Fig. 2). This scheme has been already proposed for channel estimation for Space-Time orthogonal block codes [4]. In our work, we consider only the uniform power allocation which means that the same power σ_p^2 is allocated to all added pilot symbols. The principal design parameter in this scheme is the ratio σ_p^2/σ_d^2. The transmitted symbols can still be expressed by (4). Also, the FIR channels response is expressed as in (5) and (6).

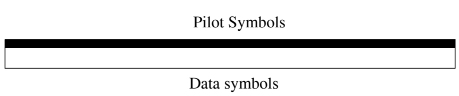

Fig. 2. Embedded pilot symbols with data symbols

2.2.3. Precoded data

At each transmitter k, a set of $N_d \leq N'$ data symbols are gathered forming the vector $\mathbf{s}_{dk}$ which is linearly precoded. Actually, depending on the precoder, this will increase the system diversity. At each time i, a set of symbols $s_k(i)$ is transmitted. We assume that embedded within each $s_k(i)$ are contributions from the set of N_d data symbols $\mathbf{s}_{dk}$ together with contribution from a training symbol $s_{pk}(i)$ (see Fig.3). Consequently, the following model is considered for the transmitted symbols from the k^{th} transmitter:

$$\mathbf{s}_k = \mathbf{U}\mathbf{s}_{dk} + \mathbf{s}_{pk}$$

Fig. 3. Precoded data with embedded pilot

where $\mathbf{U}$ is the $N' \times N_d$ matrix of the linear precoder. Therefore, the i^{th} symbol transmitted from the k^{th} antenna is:

$$s_k(i) = \mathbf{u}_{N-i+1}\mathbf{s}_{dk} + s_{pk}(i) \ \text{ with } \ i = N, \ldots, 1-L \tag{7}$$

where $\mathbf{u}_i$ is the i^{th} row of the precoding matrix.

3. CRAMER-RAO LOWER BOUND

The Cramer-Rao Bound is a performance criterion that gives a lower bound to the mean square error of estimation in the set of unbiased estimates. It is used here as a performance measure for the design and placement of pilot symbols.

3.1. Fisher Information Matrix

Under regularity conditions, the Fisher Information Matrix (FIM) associated with a complex stochastic parameter vector $\boldsymbol{\theta}$ is defined as [5]:

$$\begin{aligned} J(\boldsymbol{\theta}) &= E\left\{\left[\frac{\partial \ln p(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]\left[\frac{\partial \ln p(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]^H\right\} \\ &= E\left\{E\left\{\left[\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]\left[\frac{\partial \ln p(\mathbf{y}\mid\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]^H \;\middle|\; \boldsymbol{\theta}\right\}\right\} \\ &+ E\left\{\left[\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]\left[\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]^H\right\} \end{aligned}$$

$(.)^H$ denotes the Hermitian transpose conjugate and $(.)^*$ denotes the complex conjugate. The CRB is defined as the inverse of the FIM $J(\boldsymbol{\theta})$, i.e., for any unbiased estimate $\hat{\boldsymbol{\theta}}$ of the parameter vector $\boldsymbol{\theta}$, we have:

$$E\left\{\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^H\right]\right\} \quad \geq \quad J(\boldsymbol{\theta})^{-1}$$

Throughout our investigation, we denote by $\boldsymbol{\theta}$ the complex vector composed of the unknown parameters in (1), i.e., $\boldsymbol{\theta} = [\mathbf{s}_d^T \ \ \mathbf{h}^T]^T$. Under regularity condition and assumptions 1-4, the complex FIM is given by [2]:

$$\mathbf{J}(\theta) = \begin{bmatrix} \frac{1}{\sigma_n^2}E[\mathcal{T}(H)^H\mathcal{T}(H)] + \rho_{s_d}^2\mathbf{I} & 0 \\ 0 & \frac{1}{\sigma_n^2}E[\mathcal{S}^H\mathcal{S}] + \rho_h^2\mathbf{I} \end{bmatrix} \tag{8}$$

where $\rho_h^2 = E\left[\mid \frac{\partial \ln p_h(h)}{\partial h^*}\mid^2\right]$ and $\rho_{\mathbf{s}_d}^2 = E\left[\mid \frac{\partial \ln p_{\mathbf{s}_d}(\mathbf{s}_d)}{\partial \mathbf{s}_d{}^*}\mid^2\right]$. In the Rayleigh fading case, we have $\rho_h^2 = 1/\sigma_h^2$. Note that, in our simulations, we chose $\sigma_h^2 = 1/(L+1)$ so that the total power of the channel vector of each user is normalized to one.

3.2. Channel CRB workout

Let $\hat{\mathbf{h}}$ be the vector of channel taps estimates. The CRB on the channel taps is given by:

$$E\left\{[(\mathbf{h}-\hat{\mathbf{h}})(\mathbf{h}-\hat{\mathbf{h}})^H]\right\} \quad \geq \quad \left[\frac{1}{\sigma_n^2}E[\mathcal{S}^H\mathcal{S}] + \rho_h^2\mathbf{I}\right]^{-1} \tag{9}$$

It was assumed that data symbols are i.i.d. with zero mean. In addition, pilot symbol sequences are assumed to be known at both transmitters and receivers. Consequently:

$$E[\mathcal{S}^H\mathcal{S}] = \begin{bmatrix} E[\mathbf{S}_1^H\mathbf{S}_1] & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & E[\mathbf{S}_K^H\mathbf{S}_K] \end{bmatrix} \otimes \mathbf{I}_M$$

with:

$$E[\mathbf{S}_k^H\mathbf{S}_k] = E[\mathbf{S}_{dk}^H\mathbf{S}_{dk}] + \mathbf{S}_{pk}^H\mathbf{S}_{pk} \tag{10}$$

3.2.1. Time-multiplexed Data and Pilot

The optimal CRB for the channel taps was proposed by *Theorem 4* in [2]

$$\text{MSE}[h_{k,j}(i)] \geq \frac{\sigma_n^2}{\sigma_n^2\rho_h^2 + (N'-P)\sigma_d^2 + P\sigma_p^2} \tag{11}$$

We also calculate the CRB for the channel estimate in the case where the conventional time-multiplexed scheme is used. We recall that the conventional scheme is the scheme where P pilot symbols are sent simultaneously from each transmitter followed by $N'-P$ data symbols. In that case, using simple calculation, the expression of (10) is given by:

$$E[\mathbf{S}_k^H\mathbf{S}_k] = (N-P)\sigma_d^2\mathbf{I}_{L+1} + \sigma_d^2\text{diag}([L\ \ (L-1)\ \ \dots\ 0]) + \mathbf{S}_{pk}^H\mathbf{S}_{pk} \tag{12}$$

where $\mathbf{S}_{pk}$ is the Hankel matrix corresponding to the N' dimensional vector $[\mathbf{0}_{1\times(N'-P)}\ \ s_{pk}(P-L)\ \ \dots\ \ s_{pk}(1-L)]^T$. The analytic computation of the inverse of the expression in (12) is not simple because it depends on the pilot correlation matrix $\mathbf{S}_{pk}^H\mathbf{S}_{pk}$.

3.2.2. Embedded Pilot

We calculate (10) for the embedded pilot scheme when a sequence of length N' is transmitted from each transmitter. For large sample sizes, i.e, $N \gg 1$, and in the case where pilots are drawn from i.i.d. sequences, we can use the following approximation:

$$\begin{aligned} E[\mathbf{S}_k^H\mathbf{S}_k] &= N\sigma_d^2\mathbf{I}_{L+1} + \mathbf{S}_{pk}^H\mathbf{S}_{pk} \\ &\approx N(\sigma_d^2+\sigma_p^2)\mathbf{I}_{L+1} \end{aligned}$$

Consequently, the MSE of the i^{th} channel impulse response $h_{k,j}(i)$ from the k^{th} transmitter to the j^{th} receiver is lower bounded by:

$$\text{MSE}[h_{k,j}(i)] \geq \frac{\sigma_n^2}{\sigma_n^2\rho_h^2 + N(\sigma_d^2+\sigma_p^2)} \tag{13}$$

3.2.3. Precoded Data

Obviously, when we calculate (10) the channel CRB depends on the precoding matrix $\mathbf{U}$ by means of $E[\mathbf{S}_{dk}^H\mathbf{S}_{dk}]$.

$$E[\mathbf{S}_{dk}^H\mathbf{S}_{dk}]_{ij} = \sigma_d^2\sum_{l=0}^{N-1}\mathbf{u}_{N'-1+i-l}\mathbf{u}_{N'-1+j-l}^H \tag{14}$$

When we make the assumption of large sample sizes or pseudo random pilot sequences we can use the same approximation as in subsection 3.2.2, i.e,

$$\mathbf{S}_{pk}^H\mathbf{S}_{pk} \approx \sigma_p^2 N\ \ \mathbf{I}_{L+1}$$

To get more insight into the CRB in the precoded data case, we consider the Trailing-Zero OFDM (TZ-OFDM) case [6], where the precoding matrix is given by $\mathbf{U} = [\mathbf{F}_N^T\ \ \mathbf{0}_{L\times N}^T]^T$, $\mathbf{F}_N$ being the $N\times N$ Fourier transform matrix (we chose here $N_d = N$). In that case, the expression of (10) becomes:

$$\begin{aligned} E[\mathbf{S}_k^H\mathbf{S}_k] &= \sigma_d^2\text{diag}([N, N-1,\ \dots N-L]) + \mathbf{S}_{pk}^H\mathbf{S}_{pk} \\ &\approx \sigma_d^2\text{diag}([N, N-1,\ \dots N-L]) + \sigma_p^2 N\ \ \mathbf{I}_{L+1} \end{aligned}$$

Then

$$\text{MSE}[h_{k,j}(i)] \geq \frac{\sigma_n^2}{\sigma_n^2\rho_h^2 + (N-i)\sigma_d^2 + N\sigma_p^2} \tag{15}$$

3.3. Power Allocation Scheme

Let us denote by $\overline{\sigma}_d^2$ and (resp. $\overline{\sigma}_p^2$) the per-data-symbol (resp. per-pilot-symbol) power transmitted in the embedded scheme and σ_d^2 (resp. σ_p^2) per-data-symbol (resp. per-pilot-symbol) power transmitted in the time-multiplexed scheme. The power allocation scheme is such that the energy allocated to N' data symbols superposed with pilot symbols is the same as that allocated to disjointly sent $N'-P$ data symbols and P pilot symbols.

$$N'(\overline{\sigma}_d^2 + \overline{\sigma}_p^2) = (N'-P)\sigma_d^2 + P\sigma_p^2$$

4. CHANNEL ESTIMATION ALGORITHMS FOR THE EMBEDDED PILOT CASE

We propose an iterative MMSE algorithm to estimate the MIMO channel and perform data detection. The first step of the algorithm consists in an MMSE channel estimation according to :

$$\hat{\mathbf{h}}_k = (\mathcal{S}_{pk}^H\mathcal{S}_{pk})^{-1}\mathcal{S}_{pk}{}^H\mathbf{y}$$

where $\mathcal{S}_{pk} = \mathbf{S}_{pk}\otimes\mathbf{I}_M$. The estimation error yielded by this estimator is :

$$\tilde{\mathbf{h}}_k = (\mathcal{S}_{pk}^H\mathcal{S}_{pk})^{-1}\mathcal{S}_{pk}^H\left(\mathcal{S}_d\mathbf{h} + \sum_{l=1,l\neq k}^{K}\mathcal{S}_{pl}\mathbf{h}_l\right) + (\mathcal{S}_{pk}^H\mathcal{S}_{pk})^{-1}\mathcal{S}_{pk}\mathbf{n}$$

with $\mathcal{S}_d = \mathbf{S}_d\otimes\mathbf{I}_M$. This expression shows that the estimation error depends on the correlation between data and pilot symbols and the correlation between the pilot symbols of the different users. If data and pilot sequences are orthogonal this interference term will be null (this orthogonality condition is approximately verified for long quasi-random sequences). Otherwise, at high SNR, when the noise could be neglected, an estimation error will always be present due to this interference term [7]. For this reason, we use the MMSE estimation in conjunction with a deflation for interference mitigation. It consists in an iterative process where each iteration is composed of the following two steps.

1. A Least Squares pilot-aided estimation of each channel k:

$$\begin{aligned} \hat{\mathbf{h}}_k^{(i+1)} &= (\mathcal{S}_{pk}^H\mathcal{S}_{pk})^{-1}\mathcal{S}_{pk}^H\overline{\mathbf{y}} \\ \overline{\mathbf{y}} &= \mathbf{y} - \sum_{l=1,l\neq k}^{K}\mathcal{S}_{p_l}\hat{\mathbf{h}}_l^{(i)} - \hat{\mathcal{S}}_d^{(i)}\hat{\mathbf{h}}^{(i)} \end{aligned}$$

   $\hat{\mathbf{h}}^{(i)}$ and $\hat{\mathcal{S}}_{dl}^{(i)}$ denote the channel estimate and the estimated data block Hankel matrix at the i^{th} iteration. We initialize our algorithm by $\hat{\mathbf{h}}^{(0)} = \mathbf{0}$ and $\hat{\mathcal{S}}_d^{(0)} = \mathbf{0}$.

2. The second step of the algorithm consists in a data equalization followed by data detection. For that, MMSE (or zero forcing) equalizers are used (calculated from the previously estimated channel coefficients) and applied to the enhanced observation vectors $\tilde{\mathbf{y}}$ given by:

$$\tilde{\mathbf{y}} = \mathbf{y} - \mathcal{S}_p \hat{\mathbf{h}}^{(i+1)} - \sum_{l=1, l \neq k}^{K} \hat{\mathcal{S}}_{dl}^{(i)} \hat{\mathbf{h}}_l^{i+1}$$

A hard (eventually a soft) decision is applied after data equalization.

The iterative method is stopped when there is no significant difference between two successive channel estimates.

5. EXPERIMENTAL RESULTS

5.1. Cramer-Rao Bound plots

We plot global channel CRB given by the trace value of the matrix in (9) for the three previous schemes of pilot symbol design. For the conventional pilot scheme, we average the CRB for different pilot sequences generated randomly at each run (over 100 runs). The simulation context corresponds to: $L = 3$, $N = 80$, $P = 20$ and $\sigma_p^2/\sigma_d^2 = 2$ for the time multiplexed scheme. As observed in Fig. 4, the optimal time multiplexed scheme and the embedded one have almost the same CRB. They outperform slightly the conventional and precoded scheme.

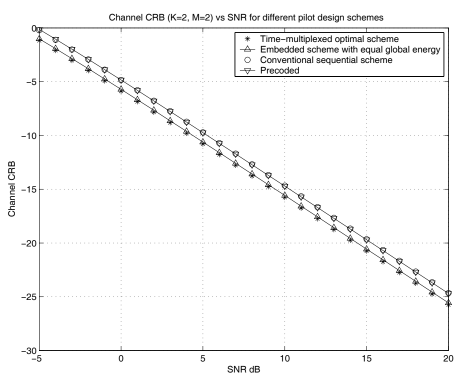

Fig. 4. Comparison of the Channel CRBs for different pilot design schemes

5.2. Channel estimation for the embedded scheme

The performance metric that is used throughout the simulations is the channel estimate MSE. The simulation results are obtained by averaging the MSE over 500 random channel realizations. For each channel realization we consider random BPSK data and pilot sequences chosen such that $\overline{\sigma}_p^2/\overline{\sigma}_d^2 = 0.5$. In Fig. 5, we compare the performances of the method with (after iteration) and without (before iteration) deflation. Note that by increasing the number of receivers we decrease the gap between the CRB and the performance of the proposed iterative MMSE algorithm.

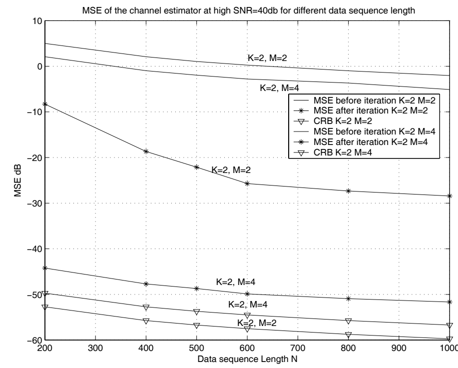

Fig. 5. MSE on channel taps.

6. CONCLUDING REMARKS

This paper presents a comparative study of different pilot design strategies for channel identification. In addition, it introduces an iterative MMSE channel estimation method for the embedded pilot scheme. The following observations can be made out of this study: (i) For large pilot sizes, the obtained CRBs are very close and almost independent from the design strategy.
(ii) The CRB corresponding to the embedded pilot case is almost the same as that corresponding to the time-multiplexed case for a same averaged power. However, in the former case, there is no waste of the channel throughput as we transmit N information symbols during the N time periods.
(iii) The performances of the proposed iterative MMSE algorithm become close to the CRB when we increase the number of receive antennae as compared to the number of transmit antennae (users).

7. REFERENCES

[1] E.Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecomm. ETT*, vol. 10, no. 6, pp. 585–596, Nov 1999.

[2] M. Dong and L.Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE. Tr. on Signal Processing*, vol. 50, no. 12, pp. 3055–3069, Dec 2002.

[3] X. M. Georgios and B. Giannakis, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE. Tr. on Signal Processing*, vol. 51, no. 5, pp. 1351–1366, May 2003.

[4] C. Budianu and L. Tong, "Channel estimation for space-time orthogonal block codes," *IEEE. Tr. on Signal Processing*, vol. 50, no. 10, pp. 2515–2527, Oct 2002.

[5] H. L. Van Trees, *Detection, Estimation and Modulation Theory*, vol. 1, New York: John Wiley, 1968.

[6] G. B. Giannakis Z. Wang B. Muquet, M. de Courville and P.Duhamel, "Reduced complexity equalizers for zero-padded ofdm transmissions," *Proc. Int. Conf. on Acoustic Speech and Signal Processing*, vol. 5, pp. 2973–2976, June 2000.

[7] E. Jorswieck V. Pohl V. Jungnickel, T. Haustein and C. von Helmot, "Performance of a mimo system with overlay pilots," *GLOBCOM'01, IEEE*, vol. 1, pp. 594–598, 2001.