Channel Estimation for Space-Time Orthogonal Block Codes

Cristian Budianu and Lang Tong†
School of Electrical and Computer Engineering
384 Frank H. T. Rhodes Hall
Cornell University
Ithaca, NY 14853
Tel: (607) 255-3900. Fax: (607)255-9079
Email: ltong@ece.cornell.edu

Technical report

Cornell University, School of Electrical and Computer Engineering

Abstract

Channel estimation is one of the key components of space-time systems design. The transmission of pilot symbols, referred to as training, is often used to aid channel acquisition. In this paper, a class of generalized training schemes that allow the superposition of training and data symbols is considered. First, the Cramér-Rao lower bound (CRLB) is derived as a function of the power allocation matrices that characterize different training schemes. Then, equivalent training schemes are obtained and the behavior of CRLB is analyzed under different power constraints. It is shown that, for certain training schemes, superimposing data with training symbols increases CRLB, and concentrating training power reduces CRLB. On the other hand, once the channel is acquired, uniformly superimposed power allocation maximizes the mutual information, and hence the capacity.

Index Terms— Space-Time Codes, Orthogonal Designs, Channel Estimation, Cramér-Rao Bound, Pilot Symbols, Power Allocation.

†Corresponding author
This work was supported in part by the Army Research Office under Grant ARO-DAAB19-00-1-0507 and the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564.

I. Introduction

A major challenge in wireless space-time communications is coping with channel uncertainties. While Shannon theory does not mandate channel estimation [22], the idea of acquiring the channel state before decoding, either blindly or through the use of pilot symbols, is entrenched in practice and has also been proposed for space-time systems [1], [15], [16], [17], [18]. The use of pilot symbols, however, may impose an unacceptable overhead that limits the effective data throughput. Here system designers must consider two contradictory goals. On the one hand, it is desirable to minimize the number of pilot symbols in a data packet so that more information carrying symbols can be transmitted. On the other hand, more pilot symbols result in better channel estimation hence reducing the symbol error rate and the need for packet retransmissions.

Conventionally, each transmitted symbol is either a pilot or a data symbol. Furthermore, pilot symbols are clustered so that training-based techniques which use received samples corresponding only to the pilot symbols can be applied. For such schemes, observations affected by the unknown data are discarded. Although training-based techniques simplify receiver design, they may carry a substantial penalty in performance for two reasons. First, the received samples corresponding to the unknown data contain valuable information about the channel. It was first established by de Carvalho and Slock [19] that the channel estimation errors can be reduced significantly by using semiblind techniques which utilize all observations for channel estimation. The second reason comes from the placement of pilot symbols in clusters suitable only for training-based techniques. It has been established recently that placing pilot symbols optimally provides gain in channel capacity [13], [9], [8] and reduction in symbol and channel estimation error [10], [20], [12].

A more general form of training that allows the superposition of pilot and data symbols has attracted attention recently [5], [8]. Such schemes, proposed earlier in [6] and [4], allow us to allocate power to data and training differently, perhaps in an adaptive fashion. It is hoped that, despite the additional complexity introduced by the mixing of pilot and data symbols, some performance gain over the conventional techniques can be realized. Furthermore, it is also hoped that the constant presence of pilot symbols in the data stream will somehow improve the tracking capability of the receiver for time varying channels.

In this paper, we consider the channel estimation problem for multiple-input multiple-output (MIMO) systems that use the orthogonal block codes proposed by Tarokh, Jafarkhani and Calderbank [2]. In addition to the placement of pilot symbols in time, we must now take the spatial domain into consideration. Within the framework of semiblind channel estimation that utilizes all observations for channel estimation, and using the Cramér-Rao Lower Bound (CRLB) as the performance measure, we examine general training strategies that allow the superposition of pilot and data symbols. In particular, we consider the effect of number of training symbols, specific training signal used, and power allocation of training symbols on CRLB. To this end, we characterize general training schemes by the power allocation matrices that specify, for each transmitted symbol in the space-time coordinate, the amount of power used for training and data respectively.

The challenge of finding the optimal (even a good) training strategy is twofold. First, one needs an expression of CRLB as a function of the power allocation matrices. Although conceptually simple, such an expression is in general complicated and not easy to optimize. Fortunately, by exploiting special properties of the orthogonal codes, we are able to simplify the CRLB expression to the point that equivalence among certain power allocation schemes can be established. The second challenge is to minimize, under a certain power constraint, the CRLB with respect to the power allocation matrices. This is again intractable in general. For the orthogonal codes presented in [2], however, we are able to show a convexity property of the CRLB. This leads to an optimal power allocation strategy under the per-symbol power constraint among those training schemes that have one pilot symbol transmitted in each block. It turns out that superimposing training with data is not optimal for channel estimation, although, with other considerations such as channel tracking and capacity enhancement, such a technique may be an appropriate compromise between accuracy in channel estimation and high rate in data transmission. While the optimal power allocation for the most general case is still unknown, our investigation reveals power allocation patterns that favor channel estimation in the acquisition stage and the optimal allocation once the channel has been acquired.

Finally, one must question that whether the CRLB is the appropriate measure. The use of CRLB as the performance measure is motivated by the consideration that training placement is a transmitter technique, and its design should not be affected by the specific technique used at the receiver. Furthermore, the asymptotic efficiency of maximum likelihood (ML) technique lends support for the use of CRLB. In this paper, we have also implemented the ML estimator and found that, for the case of using finite data samples, the performance of the ML estimator is still close to the CRLB.

This paper is organized as follows. In Section II, we present the framework and the assumptions used. The CRLB is computed in section III and it is followed by the analysis of its behavior in section IV. Numerical results that complement the theorems are presented in Section V. We conclude in Section VI. The proofs of the theorems are presented in the Appendix.

Notations used in this paper: matrices and vectors are in boldface with matrices usually in capital letters, the vectors are column vectors, $\otimes$ is the Kronecker product, $\text{diag}(\mathbf{A})$ is the vector obtained from the diagonal entries of matrix $\mathbf{A}$, $\text{diag}(\mathbf{a})$ is the diagonal matrix having $\mathbf{a}$ on the diagonal, $\text{tr}(\mathbf{A})$ is the trace of the matrix $\mathbf{A}$, $\det(\mathbf{A})$ is the determinant of $\mathbf{A}$. $\text{vec}(\mathbf{A})$ is a vector obtained by stacking the columns of $\mathbf{A}$. $(\cdot)^T$ denotes the transpose, $\mathcal{R}\mathrm{e}\{\cdot\}$ and $\mathcal{I}\mathrm{m}\{\cdot\}$ are the real and imaginary part respectively. $\mathbb{E}$ denotes the expectation. $\mathbf{I}$ is the identity matrix, $\mathbf{1}$ and $\mathbf{0}$ are the vectors that have all the elements 1 and 0 respectively. $\mathrm{cov}(\mathbf{y})$ is the covariance matrix of the random vector $\mathbf{y}$. $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ denotes a Gaussian probability distribution function (pdf) with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. $\mathbf{A} > \mathbf{B}$ with $\mathbf{A}$ and Bbf sqare matrices means that their difference $\mathbf{A} - \mathbf{B}$ is positive definite, similar $\mathbf{A} \geq B$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite. $[\mathbf{A}]_{ij}$ means the element of $\mathbf{A}$ with coordinates (i, j), the vector $\mathbf{e}_k$ is the $k - th$ vector of the standard basis.

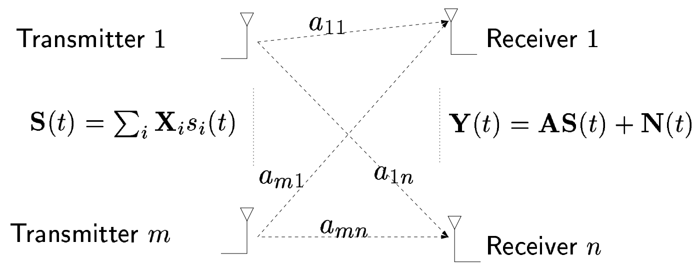

Fig. 1: An m-transmitter n-receiver space-time system.

figure

II. Model/Problem Description

A. Space-Time Block Codes

Consider a multiple antenna system with m transmitters and n receivers as shown in Fig. 1. In this paper we consider only rate one codes and real symbols, which means that a block of N symbols is transmitted within N symbol periods. For block t, denote by $\mathbf{S}(t) \in \mathbb{R}^{m \times N}$ the input of the m antennas; the kth column of $\mathbf{S}(t)$ corresponds to the transmitted vector in the kth symbol interval. The space-time code proposed by Tarokh *et al.* [2] has the following form

$$\mathbf{S}(t) = \sum_{i=1}^{N} \mathbf{X}_i s_i(t), \tag{1}$$

where $\{s_1(t), \ldots, s_N(t)\}$ is the block of N transmitted symbols and $\{\mathbf{X}_i \in \mathbb{Z}^{m \times N}\}_{i=1,\ldots,N}$ are the space-time block code (STBC) integer matrices, that satisfy

$$\mathbf{X}_j \mathbf{X}_i^T = \begin{cases} \mathbf{I} & i = j \\ -\mathbf{X}_i \mathbf{X}_j^T & i \neq j \end{cases} . \tag{2}$$

The theory of orthogonal designs also shows that (see [2]), for rate one codes and real symbols, the family $\{\mathbf{X}_i,\ i = 1, \ldots, N\}$ exists if and only if N is 2, 4 or 8. It is shown in [7] that using N single user detectors in parallel, the choice of the matrices $\{\mathbf{X}_i,\ i = 1, \ldots, N\}$ as above provide the best SNR.

Under the *quasi-static flat fading* model with coherent time of B blocks, the received signal matrix for the

t-th block is given by

$$\begin{aligned}\mathbf{Y}(t) &= \mathbf{A}\mathbf{S}(t) + \mathbf{N}(t) \\ &= \mathbf{A}\sum_{i=1}^{N}\mathbf{X}_i s_i(t) + \mathbf{N}(t), \quad t = 1, \cdots, B, \end{aligned} \tag{3}$$

where $\mathbf{A} \in \mathbb{C}^{n\times m}$ is the channel matrix and $\mathbf{N}(t)$ is the additive complex Gaussian noise.

In the sequel we need the received signal and the parameters represented as column vectors. Denote

$$\mathbf{y}(t) \triangleq \text{vec}\left(\mathbf{Y}(t)\right), \mathbf{n}(t) \triangleq \text{vec}\left(\mathbf{N}(t)\right), \tag{4}$$

$$\mathbf{a} \triangleq \text{vec}\left(\mathbf{A}^T\right), \mathbf{w}_i \triangleq \left(\mathbf{I}_n \otimes \mathbf{X}_i^T\right)\mathbf{a}. \tag{5}$$

The received signal in one block can then be written as

$$\mathbf{y}(t) = \sum_{i=1}^{N}\mathbf{w}_i s_i(t) + \mathbf{n}(t). \tag{6}$$

For real symbols and white noise, the structure of the space-time code does not depend on the number of receiving antennas, and Eq. (6) can be rewritten by separating the real and imaginary parts of the channel and noise:

$$\begin{bmatrix} \mathcal{R}\mathrm{e}\{\mathbf{y}(t)\} \\ \mathcal{I}\mathrm{m}\{\mathbf{y}(t)\} \end{bmatrix} = \left(\mathbf{I}_{2n} \otimes \sum_{i=1}^{N}\mathbf{X}_i^T s_i(t)\right)\begin{bmatrix} \mathcal{R}\mathrm{e}\{\mathbf{a}\} \\ \mathcal{I}\mathrm{m}\{\mathbf{a}\} \end{bmatrix} + \begin{bmatrix} \mathcal{R}\mathrm{e}\{\mathbf{n}\} \\ \mathcal{I}\mathrm{m}\{\mathbf{n}\} \end{bmatrix}.$$

Thus a system with n receiving antennas and complex channel coefficients is equivalent to a system with $2n$ receiving antennas with Gaussian noise $\mathcal{N}(\mathbf{0}, \frac{\sigma^2}{2}\mathbf{I})$. For simplicity, in the rest of the paper we consider the system described by (6) with all the channel coefficients real and real noise $\mathbf{n}(t) \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$.

B. Generalized Model for Training Symbols

In this section we introduce a generalized model for training strategies. To allow superimposed placements, the ith transmitted symbol $s_i(t)$ of block t is expressed as a linear combination of a pilot symbol and a data symbol

$$s_i(t) = \sqrt{\phi_{it}}v_i(t) + \sqrt{\gamma_{it}}u_i(t), \; i = 1, \ldots, N, \; t = 1, \ldots, B, \tag{7}$$

where $v_i(t)$ is the known pilot taking values from $\{\pm 1\}$ and $u_i(t)$ a data symbol drawn independently from a distribution with zero mean and unit variance. The coefficients ϕ_{it} and γ_{it} specify the power of the training and data symbols respectively. Therefore, the placement of pilot symbols within the coherent time of B blocks can be completely specified by the two $N \times B$ power allocation matrices

$$\mathbf{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_B], \; \boldsymbol{\phi}_t = [\phi_{1t}, \ldots, \phi_{Nt}]^T, \tag{8}$$

$$\mathbf{\Gamma} = [\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_B], \; \boldsymbol{\gamma}_t = [\gamma_{1t}, \ldots, \gamma_{Nt}]^T. \tag{9}$$

It is necessary to impose constraints on power allocation schemes. In this paper, we consider two types of constraints:

1. *Average Power Constraint (APC):* We assume that the average power per B blocks, each transmitting N symbols, is BN, *i.e.*,

$$\sum_{t=1}^{B}\sum_{i=1}^{N}\mathbb{E}\left\{(s_i(t))^2\right\} = \sum_{t=1}^{B}\sum_{i=1}^{N}(\phi_{it}+\gamma_{it}) = BN. \tag{10}$$

2. *Per-symbol Power Constraint (PPC):* As a special case of APC, PPC is a stronger constraint imposed on each symbol:

$$\mathbb{E}\left\{(s_i(t))^2\right\} = \phi_{it}+\gamma_{it} = 1. \tag{11}$$

Power allocation schemes that specify training schemes can be illustrated graphically. Shown in Fig. 2 is a general power allocation scheme under PPC applied to the transmission of $B = 12$ consecutive blocks with $N = 4$ symbols per block. Each column corresponds to one block of symbols transmitted together within $N = 4$ symbol intervals. Within each square, the shaded part represents the percentage of power allocated to the training part. To illustrate the power allocation under APC, a similar 3D bar-diagram may be necessary.

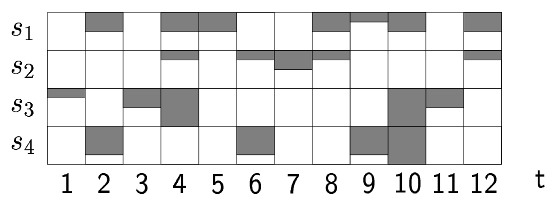

Fig. 2: A general power allocation scheme under the power constraint (11)

figure

Of particular interest are two special classes of power allocation schemes:

1. The *horizontal placement*, as illustrated in Fig. 3(a), is a scheme that places pilot symbols only in one symbol subsequence, say, without loss of generality, $s_1(t)$. The power allocation matrices $\{\mathbf{\Phi}, \mathbf{\Gamma}\}$ satisfy

$$\phi_{it} = 0, \gamma_{it} = 1, \ \ \forall\, i = 2, \ldots, N, \ \forall\, t.$$

The *periodic horizontal placement* is a horizontal placement that repeats itself every N blocks - fig. 3(b). The *uniform horizontal placement*, shown in Fig. 3(c), refers to the case when all pilot symbols in the horizontal placement have the same magnitude, *i.e.*, $\gamma_{1t} = \gamma$.

2. The *vertical* and *uniform vertical* placements, as illustrated in Fig. 3(c-d), are defined similarly. The *periodic vertical placement* is when the pilot symbols are placed periodically with period N, as shown in Fig. 3(e). It is important to note that the uniform schemes are under PPC.

Note that the conventional training-based technique corresponds to the periodic vertical placement with $\gamma = 0$.

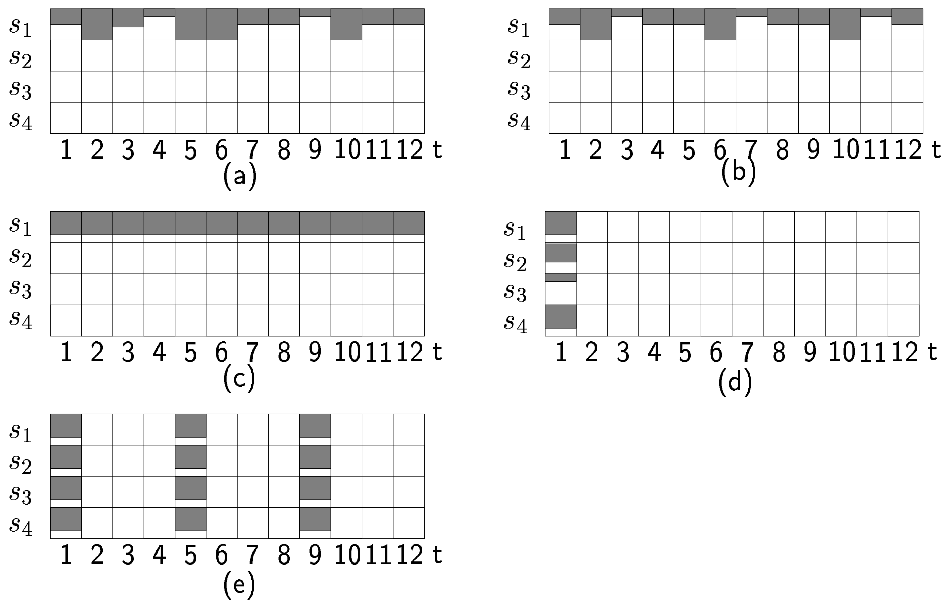

Fig. 3: Power allocation schemes defined previously- (a)-(nonuniform) horizontal scheme, (b)-uniform horizontal scheme, (c)-uniform vertical scheme

figure

C. Assumptions

The following assumptions will be imposed throughout this paper.

A1 The noise $\mathbf{n}(t)$ is i.i.d. Gaussian with zero mean and covariance $\sigma^2\mathbf{I}$.

A2 The pilot symbols are binary, $v_i(t) \in \{\pm 1\}$.

A3 The data symbols $u_i(t)$ are i.i.d. (in both t and i) Gaussian with zero mean and unit variance, *i.e.,* $u_i(t) \overset{iid}{\sim} \mathcal{N}(0, 1)$.

A4 Random variables $\{u_i(t), \mathbf{n}(t), \forall i, t\}$ are mutually independent.

A5 The code matrix $\mathbf{X}_1$ satisfies $\mathbf{X}_1 = \mathbf{I}_N$ or $\mathbf{X}_1 = [\mathbf{I}_m; 0_{N-m\times m}]$.

Collecting all observations in a vector $\mathbf{y} \overset{\Delta}{=} [\mathbf{y}^T(1), \cdots, \mathbf{y}^T(B)]^T$, under A1-A4, we have $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ where

$$\boldsymbol{\mu} = [\boldsymbol{\mu}_1^T, \cdots, \boldsymbol{\mu}_B^T]^T, \quad \boldsymbol{\mu}_t \overset{\Delta}{=} \mathbb{E}\{\mathbf{y}(t)\} = \sum_{i=1}^{N} \mathbf{w}_i \sqrt{\phi_{it}} v_i(t) \tag{12}$$

$$\mathbf{C} = \text{diag}\left[\mathbf{C}_{11}, \ldots, \mathbf{C}_{BB}\right], \quad \mathbf{C}_{tt} \overset{\Delta}{=} \text{cov}\{\mathbf{y}(t)\} = \sum_{i=1}^{N} \mathbf{w}_i \mathbf{w}_i^T \gamma_{it} + \sigma^2 \mathbf{I}_{nN}. \tag{13}$$

The assumption that pilot symbols are binary is not critical. The Gaussian assumption on data symbols, however, is essential for obtaining the CRLB and other results in this paper. In addition to making the problem analytically tractable, this assumption is partially justified because the capacity attaining signaling, under the assumption that the channel is known, is Gaussian [23]. For unknown and time invariant channels, the Gaussian signal is optimal for minimizing outage probability if the channel estimator used in the decoder is consistent [9]. Assumption A5 is made without loss of generality since the CRLB does not change if columns of all code matrices are permuted the same way or if one column of all code matrices changes the sign.

D. An Information Theoretical Perspective

Before we tackle the problem of channel estimation using superimposed training, it is relevant to examine the ultimate gain of such a scheme. Here we assume that the transmitter does not know when and if the receiver has acquired the channel. Therefore, the pilot symbols are transmitted indefinitely. Let us also assume that the receiver uses an estimator with strong consistency. Following the same argument in [9] where it is shown that the achievable transmission rate is not affected by the use of a strongly consistent channel estimator at the receiver, we can then assume that the channel is known at the receiver. We now ask: what is the placement strategy that maximizes the mutual information between the transmitter and the receiver?

Theorem 1: The mutual information between the input and the output $I(\mathbf{y}; s_1(1), \cdots, s_N(1), \cdots, s_1(B), \cdots, s_N(B))$ does not change under any permutation of the coefficients $\{\gamma_{kt}\}$. If we constrain the total amount of power that can be used to transmit pilot signals to P_{tr}, *i.e.,*

$$\sum_{t=1}^{B} \sum_{k=1}^{N} \gamma_{kt} \leq BN - P_{tr},$$

then the uniform scheme with

$$\gamma_{kt} = 1 - \frac{P_{tr}}{NB} \;\; \forall k = 1 \ldots N, t = 1 \ldots N \tag{14}$$

maximizes the mutual information between the input and the output.

Proof: see the appendix □

Of course, the above theorem tells only what happens if the channel has already been acquired. Nonetheless, it shows that if the transmitter always needs to include pilot symbols in its transmission, the superimposed strategy with a uniform placement of pilot symbols is the best. Next we look at the other part of the problem that addresses the issue of channel estimation, using a different information measure—the Fisher information.

III. Fisher Information and Cramér-Rao Lower Bound

The Fisher Information Matrix (FIM) is defined as

$$\mathbb{F}(\mathbf{a}, \mathbf{\Phi}, \mathbf{\Gamma}, \mathbf{v}, \sigma) \;\; = \;\; -\mathbb{E}\left\{ \nabla^2_{\mathbf{a}} \ln p(\mathbf{y}; \mathbf{a}, \mathbf{\Phi}, \mathbf{\Gamma}, \mathbf{v}, \sigma) \right\},$$

and the CRLB matrix $\mathbb{R}(\mathbf{a}, \mathbf{\Phi}, \mathbf{\Gamma}, \mathbf{v}, \sigma)$ is given by the inverse of FIM. In particular, for any unbiased estimator $\hat{\mathbf{a}}$, we have

$$\mathbb{E}(\hat{\mathbf{a}} - \mathbf{a})(\hat{\mathbf{a}} - \mathbf{a})^T \geq \mathbb{F}^{-1}(\mathbf{a}, \mathbf{\Phi}, \mathbf{\Gamma}, \mathbf{v}, \sigma) \stackrel{\Delta}{=} \mathbb{R}(\mathbf{a}, \mathbf{\Phi}, \mathbf{\Gamma}, \mathbf{v}, \sigma).$$

In the definition above, we have specified all relevant parameters. In order to simplify the notation in the rest of the paper, we will write only those parameters of interest. In most of cases, only the power allocation matrices $\mathbf{\Phi}$ and $\mathbf{\Gamma}$ or some columns of these matrices are listed. If the power allocation matrices have a special form and depend only on one parameter, we will use only that parameter in the argument of FIM or CRLB.

The i.i.d. assumptions on noise and data make the FIM additive. Specifically,

$$\mathbb{F}(\mathbf{\Phi}, \mathbf{\Gamma}) = \sum_{t=1}^{B} \mathbb{F}_t(\boldsymbol{\phi}_t, \boldsymbol{\gamma}_t), \tag{15}$$

where $\mathbb{F}_t(\boldsymbol{\phi}_t, \boldsymbol{\gamma}_t)$ is the FIM defined for block t. Under the Gaussian assumption (A1,A3), the FIM has a well known special form given by [24]

$$[\mathbb{F}_t(\boldsymbol{\phi}_t, \boldsymbol{\gamma}_t)]_{ij} = \left[\frac{\partial \boldsymbol{\mu}_t}{\partial a_i}\right]^T \mathbf{C}_{tt}^{-1} \left[\frac{\partial \boldsymbol{\mu}_t}{\partial a_j}\right] + \frac{1}{2}\mathrm{tr}\left[\mathbf{C}_{tt}^{-1}\frac{\partial \mathbf{C}_{tt}}{\partial a_i}\mathbf{C}_{tt}^{-1}\frac{\partial \mathbf{C}_{tt}}{\partial a_j}\right],$$

where a_i is the ith component of vector $\mathbf{a}$.

A. FIM and CRLB Expressions

We now present the expressions of FIM and CRLB on which the derivation of further properties and optimizations is based. The first expression is for the general case followed by a more compact expression for the horizontal placement. Necessary notations are listed in Table I.

$\rho \stackrel{\Delta}{=} \sigma^{-2}$	$\mathbf{v}_t \stackrel{\Delta}{=} [v_1(t), \ldots, v_N(t)]^T$
$\mathbf{W} \stackrel{\Delta}{=} [\mathbf{w}_1, \ldots, \mathbf{w}_N]$	$\mathbf{F}_i \stackrel{\Delta}{=} \frac{\partial \mathbf{W}}{\partial a_i} = [\mathbf{f}_{i1}, \ldots, \mathbf{f}_{iN}]$
$\mathbf{G}_t \stackrel{\Delta}{=} \mathrm{diag}[\gamma_{1t}, \ldots, \gamma_{Nt}]$	$\mathbf{P}_t \stackrel{\Delta}{=} \mathrm{diag}\left[\sqrt{\phi_{1t}}, \ldots, \sqrt{\phi_{Nt}}\right]$
$\mathbf{D}_t \stackrel{\Delta}{=} \mathrm{diag}[\Delta_{1t}, \ldots, \Delta_{Nt}]$	$\Delta_{it} \stackrel{\Delta}{=} -\frac{\Gamma_{it}}{\sigma^2(\|\mathbf{a}\|^2\gamma_{it}+\sigma^2)}$
$\mathbf{H}_i \stackrel{\Delta}{=} \mathbf{W}^T\mathbf{F}_i - a_i\mathbf{I}$	

TABLE I: Notations in the CRLB expressions

table

Theorem 2: The FIM for estimating the channel from the received vector $\mathbf{y}(t)$ is given by

$$[\mathbb{F}_t(\mathbf{\Gamma}, \mathbf{\Phi})]_{ij} = [\mathbf{T}_1(t)]_{ij} + [\mathbf{T}_2(t)]_{ij}, \tag{16}$$

where $\mathbf{T}_1(t)$ is the part corresponding to the mean

$$[\mathbf{T}_1(t)]_{ij} \stackrel{\Delta}{=} \mathbf{v}_t^T \mathbf{P}_t(-\mathbf{H}_i + a_i\mathbf{I})\mathbf{D}_t(\mathbf{H}_j + a_j\mathbf{I})\mathbf{P}_t\mathbf{v}_t + \rho\,\mathrm{tr}(\mathbf{P}_t^2)\delta_{ij}, \tag{17}$$

and $\mathbf{T}_2(t)$ the covariance

$$[\mathbf{T}_2(t)]_{ij} = a_i a_j \left(2q^2\,\mathrm{tr}(\mathbf{D}_t^2\mathbf{G}_t^2) + 3\rho q\,\mathrm{tr}(\mathbf{D}_t\mathbf{G}_t^2) + \rho^2\,\mathrm{tr}(\mathbf{G}_t^2)\right)$$

$$
\begin{aligned}
&+\delta_{ij}(\rho q^2\ \mathrm{tr}(\mathbf{D}_t\mathbf{G}_t^2) + \rho^2 q\ \mathrm{tr}(\mathbf{G}_t^2)) \\
&+q^2\ \mathrm{tr}(\mathbf{D}_t\mathbf{H}_i\mathbf{G}_t\mathbf{D}_t\mathbf{H}_j\mathbf{G}_t) - q^2\ \mathrm{tr}(\mathbf{D}_t\mathbf{H}_i\mathbf{G}_t^2\mathbf{D}_t\mathbf{H}_j) \\
&+2\rho q\ \mathrm{tr}(\mathbf{D}_t\mathbf{H}_i\mathbf{G}_t\mathbf{H}_j\mathbf{G}_t) - \rho q\ \mathrm{tr}(\mathbf{D}_t\mathbf{H}_i\mathbf{G}_t^2\mathbf{H}_j) \\
&+\rho^2\ \mathrm{tr}(\mathbf{G}_t\mathbf{H}_i\mathbf{G}_t\mathbf{H}_j). \qquad (18)
\end{aligned}
$$

Proof: The proof of this theorem involves a direct evaluation of FIM under the condition of code orthogonality (2) which implies that $\mathbf{w}_i$ are orthogonal vectors. The proof is given in the Appendix. □

The formula given above is for the most general placement scheme, and is difficult to analyze. However, in the special case of the horizontal placement, the FIM has a compact form.

Theorem 3: Consider only one block (t) of a horizontal power allocation scheme with $N \in \{2, 4\}$, with power allocation vectors $\boldsymbol{\gamma}_t = [\gamma, 1, \dots, 1]^T$ and $\boldsymbol{\phi}_t = [\phi, 0, \dots, 0]^T$. The FIM is given by

$$
\begin{aligned}
\mathbb{F}_t(\boldsymbol{\gamma}_t, \boldsymbol{\phi}_t) &= g_0(t)\mathbf{I} + g_1(t)\mathbf{J}_1^T\mathbf{w}_k\mathbf{w}_k^T\mathbf{J}_1 + g_2(t)\sum_{k=2}^{N}\mathbf{J}_1^T\mathbf{w}_k\mathbf{w}_k^T\mathbf{J}_1, &(19)\\
g_0(t) &\triangleq \rho\phi + (\rho q^2\ \mathrm{tr}(\mathbf{D}_t\mathbf{G}_t^2) + \rho^2 q\ \mathrm{tr}(\mathbf{G}_t^2)), &(20)\\
g_1(t) &\triangleq \left(2q^2\ \mathrm{tr}(\mathbf{D}_t^2\mathbf{G}_t^2) + 3\rho q\ \mathrm{tr}(\mathbf{D}_t\mathbf{G}_t^2) + \rho^2\ \mathrm{tr}(\mathbf{G}_t^2)\right) + \phi\Delta_\gamma, &(21)\\
g_2(t) &\triangleq \phi\Delta_1 + \gamma\Delta_1 + (N-2)\Delta_1 + \Delta_\gamma, &(22)
\end{aligned}
$$

where

$$
\mathbf{J}_i = \mathbf{I}\otimes\mathbf{X}_i^T,\ \Delta_\gamma \triangleq \Delta_{ik}|_{\gamma_{ik}=\gamma} = -\frac{\gamma}{\sigma^2(q\gamma+\sigma^2)},\ \Delta_1 \triangleq \Delta_\gamma|_{\gamma=1} = -\frac{1}{\sigma^2(q+\sigma^2)}. \qquad (23)
$$

Proof: see the Appendix. □

Observe that for a (non-uniform) horizontal scheme with B blocks, we have

$$
\mathbb{F}(\mathbf{\Gamma}, \mathbf{\Phi}) = \sum_{t=1}^{B}\mathbb{F}_t(\boldsymbol{\gamma}_t, \boldsymbol{\phi}_t) = g_0\mathbf{I} + g_1\mathbf{J}_1^T\mathbf{w}_1\mathbf{w}_1^T\mathbf{J}_1 + g_2\sum_{k=2}^{B}\mathbf{J}_1^T\mathbf{w}_k\mathbf{w}_k^T\mathbf{J}_1, \qquad (24)
$$

where

$$
g_i \triangleq \sum_{t=1}^{B} g_i(t).
$$

Due to the orthogonality property of the vectors $\{\mathbf{w}_i\}$ the inverse of the FIM can be easily computed if $N = m$. This leads to the following corollary.

Corollary 1: For a horizontal placement scheme with $N = m$, the CRLB is given by

$$
\mathbb{R} = d_0\mathbf{I} + d_1\mathbf{w}_1\mathbf{w}_1^T + d_2\sum_{k=2}^{N}\mathbf{w}_k\mathbf{w}_k^T, \qquad (25)
$$

where

$$
d_0 \triangleq \frac{1}{g_0},\ d_i \triangleq -\frac{g_i}{g_0(g_0+qg_i)}. \qquad (26)
$$

The trace of the CRLB matrix is given by

$$\mathrm{tr}(\mathbb{R}) \quad = \quad N\, d_0 + q(d_1 + (N-1)d_2). \tag{27}$$

It is especially interesting that, if $N = m$, the eigenvalues of the FIM depend only on the norm of the channel $||\mathbf{a}||$ and not the specific channel parameters. The same observation is valid for the trace of the CRLB matrix which means that the CRLB on the MSE of the channel estimator is channel independent.

IV. Behavior of FIM and CRLB

A. Equivalent Power Allocation Schemes

We present two theorems that reveal the equivalence between two classes of power allocation schemes.

Theorem 4: For $N \in \{2, 4\}$, if there is one symbol block that contains a single pilot, *i.e.,* for some block index t,

$$\boldsymbol{\phi}_t = \phi \mathbf{e}_k, \boldsymbol{\gamma}_t = \mathbf{1} - (1-\gamma)\mathbf{e}_k, \tag{28}$$

the position of that pilot symbol does not affect the FIM. See Fig. 4.

Proof: see the Appendix. □

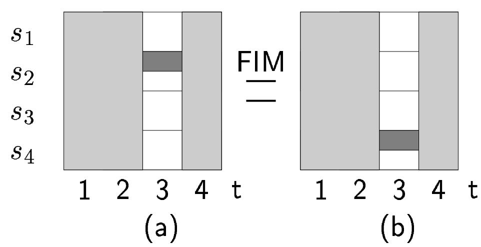

Fig. 4: Equivalent power allocation schemes in Theorem 4. The positions shaded by light grey are where power allocations can be arbitrary.

figure

An immediate consequence of the above theorem is that, if only one pilot symbol is transmitted within each block, it can be superimposed onto any data symbol (Fig. 5)(a),(b). Not so obvious is that this is not true in general for $N = 8$ and if there are more than one pilot symbol transmitted within each block. In those cases, it is possible that putting pilot symbols in different substreams gives different CRLBs.

The next theorem gives an equivalence between the uniform periodic horizontal and uniform periodic vertical placements as shown in Fig. 5(b)(c).

Theorem 5: If $N = 2$ or $N = 4$, the FIM is the same for the uniform horizontal and uniform periodic vertical placement with N blocks each and equal parameter γ.

Proof: see the Appendix. □

The equivalence of these two schemes does not seem to be obvious. Indeed, if there is no noise and each symbol is either a pilot or a data symbol, one block that contains N training symbols is sufficient for the

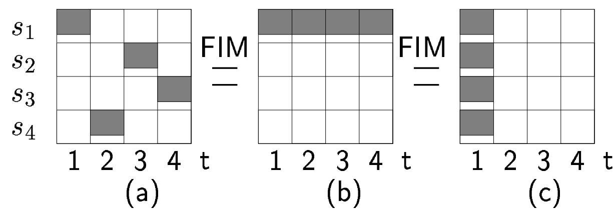

Fig. 5: Power allocation schemes in Theorems 4 and 5: (a) a periodic scheme that contains one training symbol in each block, all training symbols have equal power, (b) the uniform horizontal scheme, (c) the uniform periodic vertical scheme

figure

identification of the channel matrix. For the uniform horizontal placement, however, unknown data symbols always present in the observation. Nonetheless, it can be shown that the uniform horizontal placement of N blocks with full training symbols also leads to the identification of $\mathbf{A}$. This result is a consequence of the special properties of the orthogonal block codes, and it appears that this special code provides a symmetry in space and time; if we transpose the matrix of transmitted symbols the estimate of the channel does not change. Again, in general this result does not hold for $N = 8$ or if the placement is not uniform.

One may question the practical validity of the horizontal placement. By allowing the continuous transmission of data and pilot symbols, it seems that the the horizontal placement may offer better tracking capability for time varying channels.

B. The Convexity of FIM for Horizontal Placements

We now restrict ourselves to the horizontal placement with $N \in \{2, 4\}$. The convexity result is best illustrated in Fig. 6. Suppose that we start with a uniform horizontal placement where every pilot symbol has the same power. Now let us make the training power uneven by moving part of the training power from the second block to the first, and the same amount of data power from the first block to the second. How does the FIM vary?

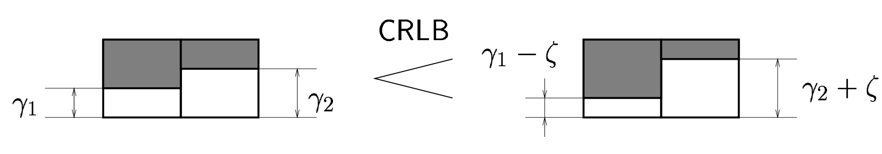

Fig. 6: Power allocation schemes compared in theorem 6 for $N = 4$.

figure

Theorem 6: Suppose that only one pilot symbol is transmitted in some block t under PPC, *i.e.*, $\boldsymbol{\gamma}_t = [\gamma, 1, \ldots, 1]^T$ and $\gamma \in (0, 1)$ and $\boldsymbol{\phi}_t = (1 - \gamma)\mathbf{e}_1$. Then $\forall \zeta \in (0, 1)$ such that $0 \leq \gamma - \zeta < \gamma + \zeta \leq 1$, the FIM $\mathbf{F}(\gamma) = \mathbb{F}_t(\boldsymbol{\gamma}_t)$

corresponding to block t satisfies

$$\frac{1}{2}(\mathbf{F}(\gamma-\zeta)+\mathbf{F}(\gamma+\zeta)) > \mathbf{F}(\gamma). \tag{29}$$

Proof: see the Appendix. □

If we apply this theorem to two blocks, it is then apparent that one should allocate training power unevenly as illustrated in Fig. 6. Please note that the sign $\{<\}$ between the schemes illustrates the fact that the RHS is "better" than the LHS, according to the CRLB criterion, which actually means that the relation between the traces of the two CRLB matrices is opposite. It is then a direct consequence of the convex property in Theorem 6 that one should avoid superimposing pilot with data symbol, and the optimal horizontal scheme is given by the following corollary.

Corollary 2: Under PPC with total training power P_{tr}, the optimal horizontal placement is given by making $\lfloor P_{tr} \rfloor$ symbols with full training and allocating the remaining training power $P_{tr} - \lfloor P_{tr} \rfloor$ to a single symbol. The rest of the symbols (if there are such symbols left) will contain only data.

If, however, the average power constraint (APC) is used, it is then possible that one symbol is transmitted with power greater than unity. In this case, one suspects that FIM can be increased further. The answer is affirmative as shown in the next theorem and illustrated in Fig. 7. In words, if there are two blocks, each with one pilot symbol. If the training power in one of the block is 100% and the other is not, the FIM increases by concentrating all the power to one block.

Theorem 7: For $N \in \{2,4\}$, consider two blocks t_1 and t_2 and denote

$$\boldsymbol{\phi}_{t_1} \triangleq [1,0,\ldots,0]^T,\ \boldsymbol{\gamma}_{t_1} \triangleq [0,1,\ldots,1]^T, \tag{30}$$

$$\boldsymbol{\phi}_{t_2} \triangleq [1-\zeta,0,\ldots,0]^T,\ \boldsymbol{\gamma}_{t_2} \triangleq [\gamma,1,\ldots,1]^T. \tag{31}$$

Then for any $\phi_1 < \phi_2$,

$$\mathbb{F}(\boldsymbol{\gamma}_{t_1},\boldsymbol{\gamma}_{t_2},\boldsymbol{\phi}_{t_1}+\phi_1\mathbf{e}_1,\boldsymbol{\phi}_{t_2}-\phi_1\mathbf{e}_1) \leq \mathbb{F}(\boldsymbol{\gamma}_{t_1},\boldsymbol{\gamma}_{t_2},\boldsymbol{\phi}_{t_1}+\phi_2\mathbf{e}_1,\boldsymbol{\phi}_{t_2}-\phi_2\mathbf{e}_1) \tag{32}$$

It follows immediately that

$$\mathrm{tr}(\mathbb{R}(\boldsymbol{\gamma}_{t_1},\boldsymbol{\gamma}_{t_2},\boldsymbol{\phi}_{t_1}+\phi_1\mathbf{e}_1,\boldsymbol{\phi}_{t_2}-\phi_1\mathbf{e}_1)) \geq \mathrm{tr}(\mathbb{R}(\boldsymbol{\gamma}_{t_1},\boldsymbol{\gamma}_{t_2},\boldsymbol{\phi}_{t_1}+\phi_2\mathbf{e}_1,\boldsymbol{\phi}_{t_2}-\phi_2\mathbf{e}_1)) \tag{33}$$

Proof: see the appendix. □

We note that, from the proof of the theorem (see Eq. (100),(102) and the comment that follows) that the FIM (and the CRLB) does not change if $\gamma = 0$. In other words, combining the energy of two full training symbols neither decreases nor increases the CRLB.

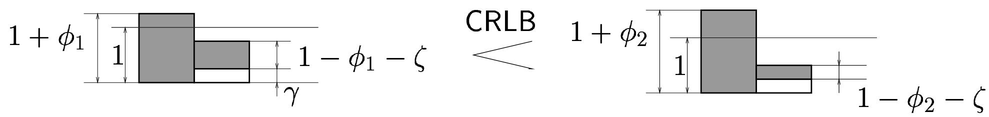

Fig. 7: Power Allocation schemes compared in theorem 7

figure

From Theorem 7, it follows that the power allocation scheme that is optimal under PPC can be improved by concentrating training power to a fewer number of pilot positions if one allows supraunitary power for some symbols. It also shows that superimposed horizontal training can not be optimal.

In order to find the optimal power allocation for APC constraints, one must now look at power allocation for data symbols. Here, unfortunately, we can only conjecture that it is also preferable to allocate data power unevenly.

Conjecture 1: Consider $N \in \{2,4\}$, $m = N$ and a horizontal power allocation scheme made up of two blocks defined as defined below :

$$\boldsymbol{\phi}_t = [\phi, 0, \ldots, 0]^T, \ t = 1, 2, \quad (34)$$

$$\boldsymbol{\gamma}_1 = [\gamma + \zeta, 1, \ldots, 1]^T, \quad (35)$$

$$\boldsymbol{\gamma}_2 = [\gamma - \zeta, 1, \ldots, 1]^T. \quad (36)$$

Then $\forall \phi, \gamma, \zeta$ such that $0 < \phi < 1, 0 < \gamma - \zeta \leq \gamma + \zeta < 1$

$$\frac{\partial \ \mathrm{tr}(\mathbb{R}(\gamma, \zeta, \phi))}{\partial \zeta} \leq 0. \quad (37)$$

□

In the conjecture above the assumption $m = N$ is important because this allows us to use the formula (27) of the trace of the CRLB matrix. The above conjecture can also be stated differently. Consider two power allocation schemes as in the conjecture with parameters (γ, ζ_1, ϕ) and (γ, ζ_2, ϕ) respectively, with $\zeta_1 < \zeta_2$. These schemes are represented in Fig. (8).Then, $\mathrm{tr}(\mathbb{R}(\gamma, \zeta_1, \phi)) \geq \mathrm{tr}(\mathbb{R}(\gamma, \zeta_2, \phi))$.

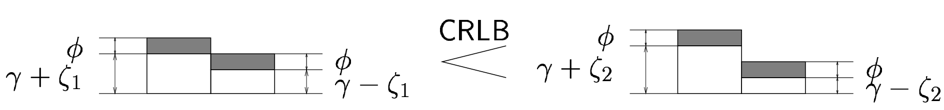

Fig. 8: Power Allocation schemes compared in conjecture 1

figure

C. Summary Scenarios

Fig. 9 summarizes our results graphically. Fig. 9(a) is the uniform horizontal scheme which can be improved by Theorem 6 to Fig. 9(c) via Fig. 9(b). This is the best under PPC. If APC is used, Fig. 9(c) can then be improved to Fig. 9(d) using Theorem 7. The same theorem says that the scheme in Fig. 9(e) and (d) have the same performance. Our conjecture suggests that Fig. 9(d),(e) are better than Fig. 9(f) and Fig. 9(g). In section V some numerical evaluations are given.

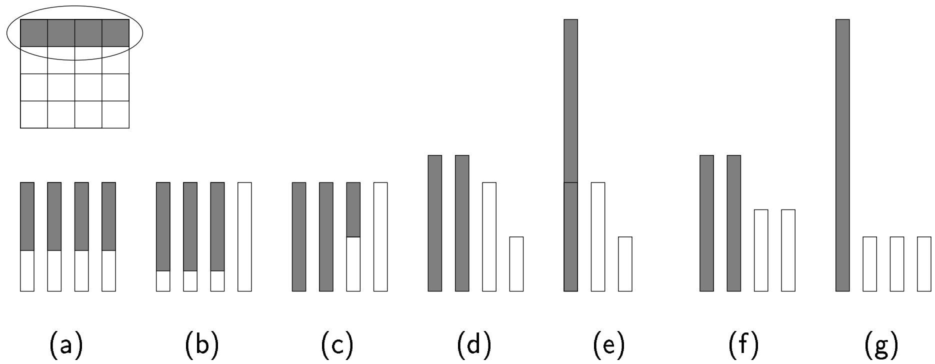

Fig. 9: Power allocation schemes compared - application of convexity theorem

figure

V. Simulations and Numerical Results

A. CRLB under Superimposed Training

In this section we want to see how does the amount of power allocated to training and the power allocation scheme influence the CRLB. Fig. 10 illustrates the variation of the trace of the CRLB matrix with the amount of power allocated to training for a system with $N = 4, m = 4, n = 4$ that uses a uniform horizontal placement scheme under PPC with $B = 32$ blocks. The channel parameters have been chosen randomly. As we expect, the CRLB decreases when the amount of training is modified. The figure reveals the behavior of the CRLB for the superimposed schemes with the SNR. At low SNR the influence of the amount of training is low, but it becomes large at high SNR. The performance of the superimposed schemes is limited by the data symbols that are unknown.

Fig. 11 compares the performance of schemes under PPC with the same amount of power allocated to training but in different ways. The system has $m = 3$, $N = 4$, $n = 5$, and number of blocks $B = 32$. The channel parameters have been chosen randomly. The uniform horizontal power allocation scheme was compared to the corresponding uniform power allocation scheme. We chose two values of the parameter γ of the horizontal schemes and then determine the allocation matrix of the uniform schemes such that the two schemes have the same power allocated to training. When the total power allocated to training is high ($\gamma = 0.1$) then the difference between the two schemes is significant, *i.e.,* the horizontal scheme (represented with continuous line) is much better. When the power allocated to training is low ($\gamma = 0.8$) then the performance of the two schemes is similar but (as we expect) lower than in the previous case.

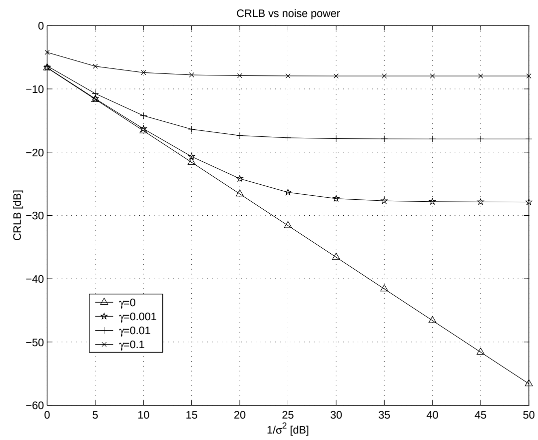

Fig. 10: $m = 4$, $N = 4$, $n = 4$

figure

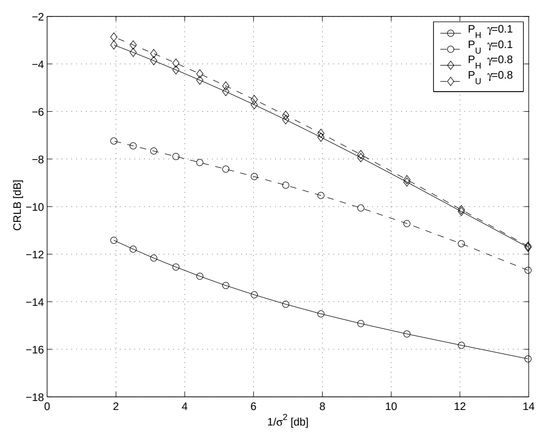

Fig. 11: $m = 3$, $N = 4$, $n = 5$

figure

B. Application of Convexity Theorems

We'll compare numerical some horizontal schemes with $N = m = n = 4$, $B = 32$, and $P_{training} = 2.5$. A comparison between these schemes has already been done in subsection IV-C that follow the convexity theorems. Also, the schemes have been represented in Fig. 9 from the same subsection. Note that in the figure is represented only one period of $N = 4$ blocks. The channel parameters have been chosen randomly. The variation of the CRLB with the SNR for different allocation schemes is plotted in Fig. 12. We'll refer to an allocation scheme by its index in Fig. 9.

It can be easy observed from Fig. 12 that the difference between the plots (a),(b) and the others is really big. This means that modifying the power allocation scheme under the conditions of theorem 6 leads to significant

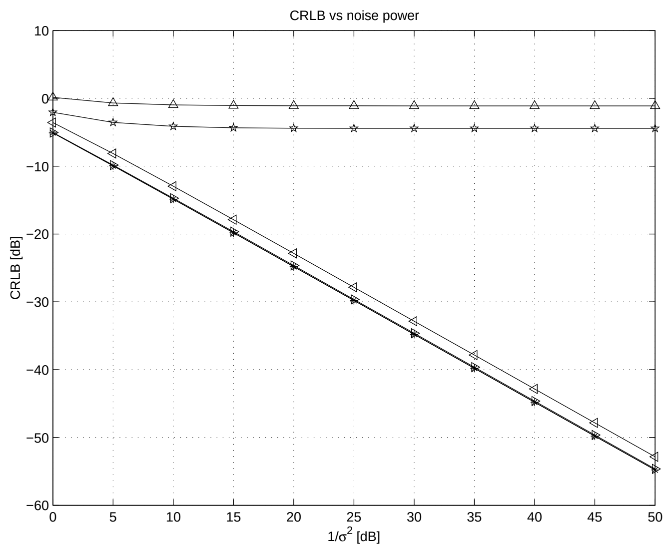

Fig. 12: Comparison of different placements: for the schemes depicted in Fig.9 the plots are from top to bottom : (a),(b),(g);-(f),(e),(c),(d) indistinguishable figure

changes of the CRLB. The modification of the scheme that is analyzed in theorem 7 produces a negligible effect on the CRLB, the performance plots for the schemes in Fig. 9(c),(d),(e) are indistinguishable. While the scheme in Fig. 9(f) is similar to the previous ones, the one represented in Fig. 9(g) is a little bit worse. Thus the modification of the CRLB under the conditions of conjecture is small.

C. Comparison between semiblind and training-based-only estimation techniques

In this section we want to compare the performance of semiblind and training-based only estimation schemes. However, such a comparation would not be fair for a superimposed scheme because the data will act like noise in the training-based-only case. Thus we have considered a uniform vertical power allocation scheme ($N = m = n = 4$, $B = 32$ blocks, channel chosen randomly) with $\gamma = 0$ (*i.e.,* full training symbols) and compared the performance of the two approaches. From theorem 5 we know that the performance of the vertical scheme used is the same as the performance of the corresponding horizontal scheme. From Fig. 13 it can be observed that the performance of the semi-blind algorithm is $4dB$ better than the performance of a training-based algorithm. However, it is easy to see that in order to have reasonable performance with training-based only estimation we need to have blocks that contain only training, *i.e.,* the training-based only estimation can't be applied efficiently if the power allocation scheme used is not uniform vertical with $\gamma = 0$.

D. The ML Algorithm

The channel estimation algorithm that was used was a semiblind Maximum Likelihood Algorithm using the scoring method. The numerical results that are presented were obtained using the following setup. The

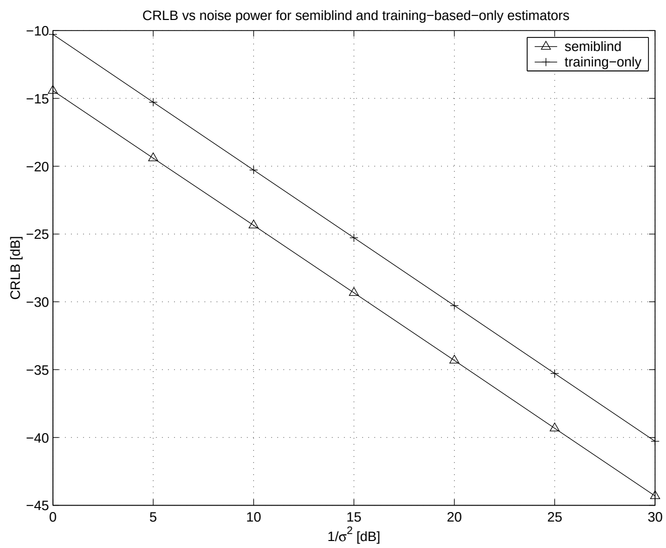

Fig. 13: CRLB for semiblind and training-based-only estimators.

figure

training symbols were binary, $\{\pm 1\}$, the parameters values were $N = 4$, $m = 3$, $n = 5$, the number of blocks considered $B = 32$. The channel coefficients were chosen randomly. 500 Monte Carlo simulations were performed. We evaluated the sum of the CRLBs for all parameters.

From Fig. 14 it can be observed that the average performance of the ML channel estimator is close to the CRLB. Similar results have been obtained using different system setups. This means that analyzing the CRLB is a good way to predict the behaviour of the performance of the channel estimation algorithms when the power allocation scheme is modified.

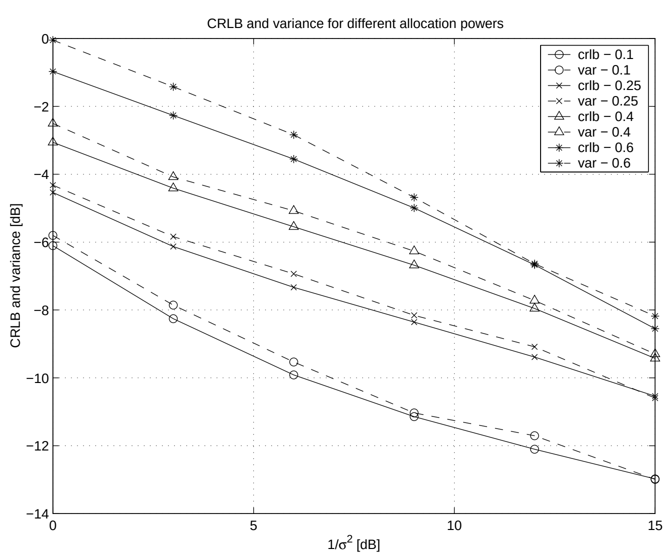

Fig. 14: CRLB and MSE of the ML estimator for $m = 3$, $N = 4$, $n = 5$.

figure

VI. Conclusions

This paper deals with channel estimation in communication systems that use a class of space-time orthogonal block codes. We have considered a general class of semiblind channel estimation techniques with superimposed data-pilot symbols. We have derived a closed-form expression for the FIM and further investigated the behaviour of CRLB for different channel estimation schemes. It was shown that for a subclass of codes the uniform horizontal scheme provides the same performance as the vertical one. Also we have characterized the behaviour of the CRLB for horizontal schemes when the power allocation parameters are varied.

What has not been achieved in this paper, unfortunately, is to find the optimal placement, which remains an open and challenging problem. What is clear though is that the placement that maximizes the mutual information is not the same as one that gives maximum Fisher information. To reach a sensible compromise, one must reformulate the problem in a different setting, allowing both channel estimation and detection errors be part of the overall consideration. Along this line, approaches like those of Hassibi and Hochwald [21], or Adireddy, Tong and Viswanathan [13] may be considered.

Appendix

A. Proof of theorem 1

If the channel is known, for Gaussian input symbols we have

$$I(\mathbf{y};\mathbf{s}) = \frac{1}{2}\log|\det(\mathsf{cov}(\mathbf{y}))| - \frac{1}{2}\log|\det(\sigma^2\mathbf{I})|,$$

where $\mathbf{s}\stackrel{\Delta}{=}[s_1(1),\ldots,s_N(1),\ldots,s_1(B),\ldots,s_N(B)]^T$. Using the properties of the space-time block codes we have :

$$\det(\mathsf{cov}(\mathbf{y})) = \det(\mathbf{C}(\mathbf{a})) = \sigma^{2B(nN-N)}\prod_{t=1}^{B}\prod_{k=1}^{N}(q\gamma_{kt}+\sigma^2), \tag{38}$$

because $q\gamma_{kt}+\sigma^2$ and σ^2 are the eigenvalues of $\mathbf{C}(\mathbf{a})$, σ^2 having multiplicity $B(nN-N)$. It is straightforward that under the constraint given in the theorem the determinant and thus the mutual information is maximized if γ_{kt} are equal.

Properties of the STBC matrices

Most of the theorems derived in this paper rely on the special properties of the STBC matrices used (2). In this section of the appendix we derive some extra properties of some STBC families of matrices, properties that will be used in the proof of theorems.

Lemma 1: Consider $N\in\{2,4\}$ and $\{\mathbf{X}_i,\ i=1,\ldots,N\}$ a family of $N\times N$ matrices satisfying conditions (2) and $\mathbf{X}_1=\mathbf{I}_N$. Construct the family $\mathbf{Z}_i\stackrel{\Delta}{=}\mathbf{X}_i\mathbf{X}_k^T$, where k is fixed and $i\in\{1,\ldots,N\}$. Then we have the following

:

$$\forall\, i \in \{1,\ldots,N\}\ \exists\, j \in \{1,\ldots,N\} \text{ such that } \mathbf{Z}_i = \pm\mathbf{X}_j$$

Proof:

If $k = 1$ then the statement is straightforward.

If $i = 1$ then $\mathbf{Z}_1 = \mathbf{X}_k^T = -\mathbf{X}_k$ thus we just choose $j = k$.

If $i = k$ then $\mathbf{Z}_k = \mathbf{X}_k\mathbf{X}_k^T = \mathbf{I} = \mathbf{X}_1$ so we choose $j = 1$.

Thus the statement follows for $N = 2$. For $N = 4$ we have to show that:

$$\forall\, i \neq k\ \exists\, j \in \{2,\ldots,N\} \text{ such that } \mathbf{X}_i\mathbf{X}_k^T = \pm\mathbf{X}_j.$$

Without loss of generality assume $k = 2$. We have the following :

$$\mathbf{X}_3\mathbf{X}_2^T \neq \pm\mathbf{X}_1;\ \mathbf{X}_3\mathbf{X}_2^T \neq \pm\mathbf{X}_2;\ \mathbf{X}_3\mathbf{X}_2^T \neq \pm\mathbf{X}_3 \tag{39}$$

Thus we must show that $\mathbf{X}_3\mathbf{X}_2^T = \pm\mathbf{X}_4$.

The family $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_2\mathbf{X}_3^T\}$ satisfies the conditions (2). We'll show that if the family $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{Z}\}$ satisfies (2) then $\mathbf{Z}$ is determined up to a sign. It follows that $\mathbf{X}_4 = \pm\mathbf{X}_2\mathbf{X}_3^T$.

We know that in any of the STBC matrices each row and each column has only one non-zero entry. The position of the nonzero entries of $\mathbf{Z}$ are determined by the other three matrices from the orthogonality conditions (2). We assumed $\mathbf{X}_1 = \mathbf{I}$ so the diagonal entries of the other three matrices are all 0. Also, because $\mathbf{X}_1 = \mathbf{I}$ we have $\mathbf{Z} = -\mathbf{Z}^T$ and the two nonzero entries below the main diagonal are determined by the other two nonzero entries.

Denote by $\mathbf{Z}(k,:)$ the k-th row of $\mathbf{Z}$. From $\mathbf{X}_2\mathbf{Z}^T = -\mathbf{Z}\mathbf{X}_2^T$ it follows that there are k_1 and k_2 such that

$$\mathbf{X}_2(k_1,:)\mathbf{Z}(k_2,:)^T = -\mathbf{Z}(k_1,:)\mathbf{X}_2(k_2,:)^T \neq 0. \tag{40}$$

This corresponds to a non-zero entry of $\mathbf{Z}\mathbf{X}_2^T$, so $k_1 \neq k_2$.

Each row vector $\mathbf{Z}(k_1,:)$ and $\mathbf{Z}(k_2,:)$ contains only one non-zero entry. If the nonzero entries that are contained in $\mathbf{Z}(k_1,:)$ and $\mathbf{Z}(k_2,:)$ are placed symmetrical with respect to the first diagonal of $\mathbf{Z}$, then $\mathbf{Z}(k_2,k_1) \neq 0$. This implies $\mathbf{X}_2(k_1,k_1) \neq 0$which is false because $\mathbf{X}_2 = -\mathbf{X}_2^T$.

Thus (40) provides a relation between two elements of $\mathbf{Z}$ which are not placed symmetrical with respect to the main diagonal. This means that $\mathbf{Z}$ is determined up to the sign, which shows that $\mathbf{X}_3\mathbf{X}_2^T = \pm\mathbf{X}_4$.

□

Remark : in Tarokh's paper [2] it is mentioned the connection between a family of 4×4 STBC matrices and the quaternionic algebra, however, the statement of the lemma does not follow immediately.

Lemma 2: Consider $N = 4$ and $\{\mathbf{X}_i,\ i = 1, \ldots, N\}$ a family of $m \times N$, $m < N$, matrices satisfying conditions (2). As in [7], denote by $\{\mathbf{Z}_i,\ i = 1, \ldots, N\}$ the generating family of matrices. Define the following family of matrices:

$$\mathbf{G}_k \stackrel{\Delta}{=} \begin{bmatrix} \mathbf{Z}_1(k,:) \\ \vdots \\ \mathbf{Z}_N(k,:) \end{bmatrix},\ k = 1, \ldots, N$$

Note that, by the definition of the family $\{\mathbf{G}_k\}$, $\mathbf{X}_k$ is made up of the first m rows of $\mathbf{G}_k$. Consider that $\mathbf{G}_1 = \mathbf{I}$. Then for any fixed k we have

$$\forall\, i \in \{1, \ldots, N\}\ \ \exists\, j \in \{1, \ldots, N\} \text{ such that } \mathbf{X}_i \mathbf{G}_k^T \mathbf{G}_1 = \pm \mathbf{X}_j. \tag{41}$$

We have to observe that if we fix i then j is different for different choices of k. It follows that if we fix j we have:

$$\forall\, i \in \{1, \ldots, N\}\ \ \exists\, k \in \{1, \ldots, N\} \text{ such that } \mathbf{X}_i \mathbf{G}_k^T \mathbf{G}_1 = \pm \mathbf{X}_j. \tag{42}$$

Proof:

It is easy to show that $\{\mathbf{G}_i, i \in \{1, \ldots, N\}\}$ satisfies (2). $\mathbf{X}_l \mathbf{G}_k^T \mathbf{G}_1 = \mathbf{X}_l \mathbf{G}_k^T$ are the first m rows of $\mathbf{G}_l \mathbf{G}_k^T$. Then (41) follows immediately from lemma 1.

□

Lemma 3: Consider the same family of matrices as in lemma 2. Then for any choice of 4 different indices k_1, k_2, k_3, k_4 we have :

$$\mathbf{X}_{k_1} \mathbf{X}_{k_2}^T = \pm \mathbf{X}_{k_3} \mathbf{X}_{k_4}^T$$

Proof:

Use (42) from 2 to choose k such that:

$$\mathbf{X}_{k_1} \mathbf{G}_k^T = \pm \mathbf{X}_{k_3}$$

Then by (41) with $i = k_2$ we have,

$$\mathbf{X}_{k_2} \mathbf{G}_k^T = \pm \mathbf{X}_{k_4}$$

because all the other three choices for the RHS matrix are not possible. The statement follows.

□

Proof of Theorem 2

Using the properties of the STBC matrices (2) the following properties can be checked:

$$\mathbf{w}_i^T \mathbf{w}_k \;=\; q\delta_{ik},\quad \mathbf{W}^T \mathbf{W} = q\mathbf{I}, \tag{43}$$

$$\begin{aligned}\mathbf{f}_{ik_1}^T \mathbf{f}_{jk_2} &= -\mathbf{f}_{ik_2}^T \mathbf{f}_{jk_1} \text{ if } k_1 \neq k_2, \quad \mathbf{f}_{ik_1}^T \mathbf{f}_{ik_2} = \delta_{k_1 k_2} \\ \mathbf{W}^T \mathbf{F}_i &= \mathbf{H}_i + a_i \mathbf{I}, \quad \mathbf{F}_i^T \mathbf{W} = -\mathbf{H}_i + a_i \mathbf{I} \\ \mathbf{H}_i^T &= -\mathbf{H}_i.\end{aligned}$$

It is convenient to separate the expression of the elements of the FIM in two parts

$$\begin{aligned}[\mathbb{F}_t(\boldsymbol{\phi}_t, \boldsymbol{\gamma}_t)]_{i,j} &= T_1(t) + T_2(t), \\ T_1(t) &= \left(\frac{\partial \boldsymbol{\mu}_t}{\partial a_i}\right)^T \mathbf{C}_{tt}^{-1} \left(\frac{\partial \boldsymbol{\mu}_t}{\partial a_j}\right), \\ T_2(t) &= \frac{1}{2} \mathrm{tr} \left(\mathbf{C}_{tt}^{-1} \frac{\partial \mathbf{C}_{tt}}{\partial a_i} \mathbf{C}_{tt}^{-1} \frac{\partial \mathbf{C}_{tt}}{\partial a_j} \right).\end{aligned}$$

In order to obtain the formula of the elements of the FIM in the theorem we write the following relations:

$$\frac{\partial \boldsymbol{\mu}_t}{\partial a_i} = \mathbf{F}_i \mathbf{P}_t \mathbf{v}_t, \ \mathbf{C}_{tt} = \mathbf{W} \mathbf{G}_t \mathbf{W}^T + \sigma^2 \mathbf{I} \tag{44}$$

$$\mathbf{C}_{tt}^{-1} = \mathbf{W} \mathbf{D}_t \mathbf{W}^T + \rho \mathbf{I}, \ \frac{\partial \mathbf{C}_{tt}}{\partial a_i} = \mathbf{F}_i \mathbf{G}_t \mathbf{W}^T + \mathbf{W} \mathbf{G}_t \mathbf{F}_i^T. \tag{45}$$

Plugging the subexpressions above in the formula of the FIM and taking into account the special properties previously listed we obtain the formula given in the theorem.

$T_1(t)$ is quite straightforward, as follows.

$$T_1(t) = \left(\frac{\partial \mu_t}{\partial a_i}\right)^T \mathbf{C}_{tt}^{-1} \left(\frac{\partial \mu_t}{\partial a_j}\right) \tag{46}$$

$$= \mathbf{v}^T \mathbf{P}_t \mathbf{F}_i^T \mathbf{W} \mathbf{D}_t \mathbf{W}^T \mathbf{F}_j \mathbf{P}_t \mathbf{v}_t + \rho \mathbf{v}^T \mathbf{P}_t \mathbf{F}_i^T \mathbf{F}_j \mathbf{P}_t \mathbf{v}_t \tag{47}$$

$$= \mathbf{v}^T \mathbf{P}_t (-\mathbf{H}_i + a_i \mathbf{I}) \mathbf{D}_t (\mathbf{H}_j + a_j \mathbf{I}) \mathbf{P}_t \mathbf{v}_t + \rho \ \mathrm{tr}(\mathbf{P}_t^2) \delta_{ij} \tag{48}$$

Denote

$$T_{11}(t) \triangleq \mathbf{v}^T \mathbf{P}_t (-\mathbf{H}_i + a_i \mathbf{I}) \mathbf{D}_t (\mathbf{H}_j + a_j \mathbf{I}) \mathbf{P}_t \mathbf{v}_t \tag{49}$$

$$T_{12}(t) \triangleq \rho \ \mathrm{tr}(\mathbf{P}_t^2) \delta_{ij} \tag{50}$$

For $T_2(t)$ we need to substitute the covariance matrix and its derivative and then write down all the sixteen terms. Then we simplify these terms exchanging the order of the terms under the trace and using the special properties of the matrices involved. The next steps are grouping the simplified terms in pairs of identical terms and substituting the formulas (44). Then we use once more the special properties of the matrices involved and we collect the terms in order to obtain the expression of $T_2(t)$ given in the theorem.

$$T_2(t) \triangleq \frac{1}{2}\text{tr}\left(\mathbf{C}_{tt}^{-1}\frac{\partial \mathbf{C}_{tt}}{\partial a_i}\mathbf{C}_{tt}^{-1}\frac{\partial \mathbf{C}_{tt}}{\partial a_j}\right)$$

With the notations previously introduced T_2 becomes:

$$T_2(t) = \frac{1}{2}\text{tr}\left((\mathbf{W}\mathbf{D}_t\mathbf{W}^T + \rho\mathbf{I})(\mathbf{F}_i\mathbf{G}_t\mathbf{W}^T + \mathbf{W}\mathbf{G}_t\mathbf{F}_i^T)(\mathbf{W}\mathbf{D}_t\mathbf{W}^T + \rho\mathbf{I})(\mathbf{F}_j\mathbf{G}_t\mathbf{W}^T + \mathbf{W}\mathbf{G}_t\mathbf{F}_j^T)\right)$$

In the following we write all the sixteen terms in the expression above, we simplify them and we see which of them are the same and/or cancel.

The simplified expressions of the terms are listed in the third column.

M_{1111}	$\text{tr}\left(\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{F}_i\mathbf{G}_t\mathbf{W}^T\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{F}_j\mathbf{G}_t\mathbf{W}^T\right)$	$q^2\ \text{tr}\left(\mathbf{D}_t\mathbf{W}^T\mathbf{F}_i\mathbf{G}_t\mathbf{D}_t\mathbf{W}^T\mathbf{F}_j\mathbf{G}_t\right)$
M_{1112}	$\text{tr}\left(\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{F}_i\mathbf{G}_t\mathbf{W}^T\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{W}\mathbf{G}_t\mathbf{F}_j^T\right)$	$q^2\ \text{tr}\left(\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{F}_i\mathbf{G}_t\mathbf{D}_t\mathbf{G}_t\mathbf{F}_j^T\right)$
M_{1121}	$\rho\ \text{tr}\left(\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{F}_i\mathbf{G}_t\mathbf{W}^T\mathbf{F}_j\mathbf{G}_t\mathbf{W}^T\right)$	$\rho q\ \text{tr}\left(\mathbf{D}_t\mathbf{W}^T\mathbf{F}_i\mathbf{G}_t\mathbf{W}^T\mathbf{F}_j\mathbf{G}_t\right)$
M_{1122}	$\rho\ \text{tr}\left(\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{F}_i\mathbf{G}_t\mathbf{W}^T\mathbf{W}\mathbf{G}_t\mathbf{F}_j^T\right)$	$\rho q\ \text{tr}\left(\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{F}_i\mathbf{G}_t^2\mathbf{F}_j^T\right)$
M_{1211}	$\text{tr}\left(\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{W}\mathbf{G}_t\mathbf{F}_i^T\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{F}_j\mathbf{G}_t\mathbf{W}^T\right)$	$q^2\ \text{tr}\left(\mathbf{D}_t\mathbf{G}_t\mathbf{F}_i^T\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{F}_j\mathbf{G}_t\right)$
M_{1212}	$\text{tr}\left(\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{W}\mathbf{G}_t\mathbf{F}_i^T\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{W}\mathbf{G}_t\mathbf{F}_j^T\right)$	$q^2\ \text{tr}\left(\mathbf{W}\mathbf{D}_t\mathbf{G}_t\mathbf{F}_i^T\mathbf{W}\mathbf{D}_t\mathbf{G}_t\mathbf{F}_j^T\right)$
M_{1221}	$\rho\ \text{tr}\left(\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{W}\mathbf{G}_t\mathbf{F}_i^T\mathbf{F}_j\mathbf{G}_t\mathbf{W}^T\right)$	$\rho q^2\ \text{tr}\left(\mathbf{D}_t\mathbf{G}_t\mathbf{F}_i^T\mathbf{F}_j\mathbf{G}_t\right)$
M_{1222}	$\rho\ \text{tr}\left(\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{W}\mathbf{G}_t\mathbf{F}_i^T\mathbf{W}\mathbf{G}_t\mathbf{F}_j^T\right)$	$\rho q\ \text{tr}\left(\mathbf{W}\mathbf{D}_t\mathbf{G}_t\mathbf{F}_i^T\mathbf{W}\mathbf{G}_t\mathbf{F}_j^T\right)$
M_{2111}	$\rho\ \text{tr}\left(\mathbf{F}_i\mathbf{G}_t\mathbf{W}^T\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{F}_j\mathbf{G}_t\mathbf{W}^T\right)$	$\rho q\ \text{tr}\left(\mathbf{F}_i\mathbf{G}_t\mathbf{D}_t\mathbf{W}^T\mathbf{F}_j\mathbf{G}_t\mathbf{W}^T\right)$
M_{2112}	$\rho\ \text{tr}\left(\mathbf{F}_i\mathbf{G}_t\mathbf{W}^T\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{W}\mathbf{G}_t\mathbf{F}_j^T\right)$	$\rho q^2\ \text{tr}\left(\mathbf{F}_i\mathbf{G}_t\mathbf{D}_t\mathbf{G}_t\mathbf{F}_j^T\right)$
M_{2121}	$\rho^2\ \text{tr}\left(\mathbf{F}_i\mathbf{G}_t\mathbf{W}^T\mathbf{F}_j\mathbf{G}_t\mathbf{W}^T\right)$	$\rho^2\ \text{tr}\left(\mathbf{F}_i\mathbf{G}_t\mathbf{W}^T\mathbf{F}_j\mathbf{G}_t\mathbf{W}^T\right)$
M_{2122}	$\rho^2\ \text{tr}\left(\mathbf{F}_i\mathbf{G}_t\mathbf{W}^T\mathbf{W}\mathbf{G}_t\mathbf{F}_j^T\right)$	$\rho^2 q\ \text{tr}\left(\mathbf{F}_i\mathbf{G}_t^2\mathbf{F}_j^T\right)$
M_{2211}	$\rho\ \text{tr}\left(\mathbf{W}\mathbf{G}_t\mathbf{F}_i^T\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{F}_j\mathbf{G}_t\mathbf{W}^T\right)$	$\rho q\ \text{tr}\left(\mathbf{G}_t\mathbf{F}_i^T\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{F}_j\mathbf{G}_t\right)$
M_{2212}	$\rho\ \text{tr}\left(\mathbf{W}\mathbf{G}_t\mathbf{F}_i^T\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{W}\mathbf{G}_t\mathbf{F}_j^T\right)$	$\rho q\ \text{tr}\left(\mathbf{W}\mathbf{G}_t\mathbf{F}_i^T\mathbf{W}\mathbf{D}_t\mathbf{G}_t\mathbf{F}_j^T\right)$
M_{2221}	$\rho^2\ \text{tr}\left(\mathbf{W}\mathbf{G}_t\mathbf{F}_i^T\mathbf{F}_j\mathbf{G}_t\mathbf{W}^T\right)$	$\rho^2 q\ \text{tr}\left(\mathbf{G}_t\mathbf{F}_i^T\mathbf{F}_j\mathbf{G}_t\right)$
M_{2222}	$\rho^2\ \text{tr}\left(\mathbf{W}\mathbf{G}_t\mathbf{F}_i^T\mathbf{W}\mathbf{G}_t\mathbf{F}_j^T\right)$	$\rho^2\ \text{tr}\left(\mathbf{W}\mathbf{G}_t\mathbf{F}_i^T\mathbf{W}\mathbf{G}_t\mathbf{F}_j^T\right)$

In the next table the terms from the third column of the previous table are grouped and then are further simplified substituting the formulas (44).

M_{1111}	T_{1212}	q^2 tr $\left(\mathbf{D}_t\mathbf{W}^T\mathbf{F}_i\mathbf{G}_t\mathbf{D}_t\mathbf{W}^T\mathbf{F}_j\mathbf{G}_t\right)$	q^2 tr$(\mathbf{D}_t\mathbf{H}_i\mathbf{G}_t\mathbf{D}_t\mathbf{H}_j\mathbf{G}_t) + q^2 a_i a_j$ tr$(\mathbf{D}_t^2\mathbf{G}_t^2)$
M_{1211}	T_{1112}	q^2 tr $\left(\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{F}_i\mathbf{G}_t\mathbf{D}_t\mathbf{G}_t\mathbf{F}_j^T\right)$	$-q^2$ tr$(\mathbf{D}_t\mathbf{H}_i\mathbf{G}_t^2\mathbf{D}_t\mathbf{H}_j) + q^2 a_i a_j$ tr$(\mathbf{D}_t^2\mathbf{G}_t^2)$
M_{1222} M_{2111}	T_{1121} T_{2211}	ρq tr$(\mathbf{D}_t\mathbf{W}^T\mathbf{F}_i\mathbf{G}_t\mathbf{W}^T\mathbf{F}_j\mathbf{G}_t)$	ρq tr$(\mathbf{D}_t\mathbf{H}_i\mathbf{G}_t\mathbf{H}_j\mathbf{G}_t) + \rho q a_i a_j$ tr$(\mathbf{D}_t\mathbf{G}_t^2)$
M_{2211}	T_{1122}	ρq tr $\left(\mathbf{W}\mathbf{D}_t\mathbf{W}^T\mathbf{F}_i\mathbf{G}_t^2\mathbf{F}_j^T\right)$	$-\rho q$ tr$(\mathbf{D}_t\mathbf{H}_i\mathbf{G}_t^2\mathbf{H}_j) + \rho q a_i a_j$ tr$(\mathbf{D}_t\mathbf{G}_t^2)$
M_{2112}	T_{1221}	ρq^2 tr $\left(\mathbf{D}_t\mathbf{G}_t\mathbf{F}_i^T\mathbf{F}_j\mathbf{G}_t\right)$	ρq^2 tr$(\mathbf{D}_t\mathbf{G}_t^2)\delta_{ij}$
M_{2222}	T_{2121}	$\rho^2 q$ tr $\left(\mathbf{G}_t\mathbf{F}_i^T\mathbf{F}_j\mathbf{G}_t\right)$	ρ^2 tr$(\mathbf{G}_t\mathbf{H}_i\mathbf{G}_t\mathbf{H}_j) + \rho^2 a_i a_j$ tr$(\mathbf{G}_t^2)$
M_{2221}	T_{2122}	$\rho^2 q$ tr $\left(\mathbf{F}_i\mathbf{G}_t^2\mathbf{F}_j^T\right)$	$\rho^2 q$ tr$(\mathbf{G}_t^2)\delta_{ij}$

Group together the terms that contain $a_i a_j$ the terms that contain δ_{ij} and the rest of the terms that contain the matrices $\mathbf{H}$.

$$\begin{aligned} T_2(t) &= T_{21}(t) + T_{22}(t) + T_{23}(t) && (51)\\ T_{21}(t) &= a_i a_j \left(2q^2 \text{ tr}(\mathbf{D}_t^2\mathbf{G}_t^2) + 3\rho q \text{ tr}(\mathbf{D}_t\mathbf{G}_t^2) + \rho^2 \text{ tr}(\mathbf{G}_t^2)\right) && (52)\\ T_{22}(t) &= \delta_{ij}(\rho q^2 \text{ tr}(\mathbf{D}_t\mathbf{G}_t^2) + \rho^2 q \text{ tr}(\mathbf{G}_t^2)) && (53)\\ T_{23}(t) &= q^2 \text{ tr}(\mathbf{D}_t\mathbf{H}_i\mathbf{G}_t\mathbf{D}_t\mathbf{H}_j\mathbf{G}_t) - q^2 \text{ tr}(\mathbf{D}_t\mathbf{H}_i\mathbf{G}_t^2\mathbf{D}_t\mathbf{H}_j) && (54)\\ &\quad +2\rho q \text{ tr}(\mathbf{D}_t\mathbf{H}_i\mathbf{G}_t\mathbf{H}_j\mathbf{G}_t) - \rho q \text{ tr}(\mathbf{D}_t\mathbf{H}_i\mathbf{G}_t^2\mathbf{H}_j)\\ &\quad +\rho^2 \text{ tr}(\mathbf{G}_t\mathbf{H}_i\mathbf{G}_t\mathbf{H}_j) \end{aligned}$$

The expression given in the theorem follows.

□

Proof of Theorem 3

First we'll present some properties that allow the simplification of the FIM formula for the horizontal case.

We'll exploit the structure of the matrices $\{\mathbf{H}_i\}$ by defining the vectors $\{\mathbf{h}_i\}$ and the matrices $\{\mathbf{B}_i\}$ by the relation below ($\mathbf{h}_i$ is the first column of the matrix $\mathbf{H}_i$ and $\mathbf{B}_i$ is the block of $\mathbf{H}_i$ obtained by deleting the first row and the first column)

$$\mathbf{H}_i = [[0, \mathbf{h}_i^T]^T, [-\mathbf{h}_i, \mathbf{B}_i^T]^T]. \tag{55}$$

It is easy to observe that the matrices $\{\mathbf{B}_i\}$ satisfy $\{\mathbf{B}_i^T = -\mathbf{B}_i\}$.

Proposition 1: For any two diagonal matrices of the form

$$\mathbf{C}_k \stackrel{\Delta}{=} = diag(c_{0k}, c_k, \ldots, c_k), \quad k \in \{1, 2\} \tag{56}$$

we have :

$$\mathrm{tr}(\mathbf{C}_1\mathbf{H}_i\mathbf{C}_2\mathbf{H}_j) = -(c_{01}c_2 + c_{02}c_1)\mathbf{h}_i^T\mathbf{h}_j - c_1c_2\ \mathrm{tr}(\mathbf{B}_i\mathbf{B}_j). \quad (57)$$

□

Proposition 2: We have :

$$[\mathbf{H}_i\mathbf{H}_j]_{k_1,k_2} = -[\mathbf{H}_i\mathbf{H}_j]_{k_2,k_1}\ \ k_1 \neq k_2. \quad (58)$$

Proof:

Taking into account that $\mathbf{H}_{ik,k} = 0$, we can write

$$[\mathbf{H}_i\mathbf{H}_j]_{k_1k_2} = \sum_{k_3 \neq k_1,k_2} [\mathbf{H}_i]_{k_1,k_3}[\mathbf{H}_j]_{k_3,k_2}. \quad (59)$$

For $N = 2$ the statement is clear (both sides are 0). For $N = 4$ the sum has only two terms

$$[\mathbf{H}_i\mathbf{H}_j]_{k_1k_2} = [\mathbf{H}_i]_{k_1k_3}[\mathbf{H}_j]_{k_3k_2} + [\mathbf{H}_i]_{k_1k_4}[\mathbf{H}_j]_{k_4k_2} \quad (60)$$

$$[\mathbf{H}_j\mathbf{H}_i]_{k_1k_2} = [\mathbf{H}_j]_{k_1k_3}[\mathbf{H}_i]_{k_3k_2} + [\mathbf{H}_j]_{k_1k_4}[\mathbf{H}_i]_{k_4k_2}. \quad (61)$$

From lemma 1 and lemma 3 and the expression of the elements of $\{\mathbf{H}_i\}$we know that $\forall\, i \in \{1, \ldots, N\}$

$$[\mathbf{H}_i]_{k_1k_4} = \pm[\mathbf{H}_i]_{k_3k_2},\ [\mathbf{H}_i]_{k_4k_2} = \pm[\mathbf{H}_i]_{k_1k_3}.$$

The statement follows.

□

Proposition 3: We have

$$\mathrm{tr}(\mathbf{B}_i\mathbf{B}_j) = -(N-2)\mathbf{h}_i^T\mathbf{h}_j. \quad (62)$$

Proof:

For $N = 2$ is clear. For $N = 4$ express the LHS using $\mathbf{w}_k$ and $\mathbf{f}_{ik}$ and then apply lemma 1 if $m = N$ or lemma 3 if $m < N$.

□

Now we return to the main part of the proof. For the horizontal placement scheme the matrices $\mathbf{G}_t$, $\mathbf{D}_t$ and $\mathbf{P}_t$, with t fixed, previously defined for the general case, become

$$\mathbf{G}_t = diag(\gamma, 1, \ldots, 1);\ \mathbf{D}_t = diag(\Delta_\gamma, \Delta_1, \ldots, \Delta_1);\ \mathbf{P}_t = diag(\sqrt{\phi}, 0, \ldots, 0).$$

Using the special structure of the matrices above and the properties previously derived, the general formula of the elements of the FIM can be simplified. We'll use the terms as derived in the proof of theorem 2.

$$
\begin{aligned}
T_{11H}(t) &= \mathbf{v}^T\mathbf{P}_t(-\mathbf{H}_i + a_i\mathbf{I})\mathbf{D}_t(\mathbf{H}_j + a_j\mathbf{I})\mathbf{P}_t\mathbf{v}_t && (63)\\
&= -\mathbf{v}^T\mathbf{P}_t\mathbf{H}_i\mathbf{D}_t\mathbf{H}_j\mathbf{P}_t\mathbf{v}_t + a_ia_j\mathbf{v}^T\mathbf{P}_t\mathbf{D}_t\mathbf{P}_t\mathbf{v}_t && (64)\\
&= \phi\Delta_1\mathbf{h}_i^T\mathbf{h}_j + a_ia_j\phi\Delta_\gamma && (65)
\end{aligned}
$$

We have used $\mathbf{v}^T\mathbf{P}_t\mathbf{H}_i\mathbf{D}_t\mathbf{P}_t\mathbf{v}_t = 0$.

$$
T_{12H}(t) = \rho\phi\delta_{ij}. \qquad (66)
$$

$$
\begin{aligned}
T_{21H}(t) &= a_ia_j(2q^2(\gamma^2{\Delta_\gamma}^2 + (N-1){\Delta_1}^2) + 3\rho q(\gamma^2\Delta_\gamma && (67)\\
&\quad +(N-1)\Delta_1) + \rho^2(\gamma^2 + (N-1)))
\end{aligned}
$$

$$
T_{22H}(t) = \delta_{ij}(\rho q^2(\gamma^2\Delta_\gamma + (N-1)\Delta_1) + \rho^2 q(\gamma^2 + (N-1))) \qquad (68)
$$

Now compute T_{23} for the horizontal placement scheme.

$$
\begin{aligned}
T_{23H}(t) &= q^2(-2\gamma\Delta_\gamma\Delta_1\mathbf{h}_i^T\mathbf{h}_j + {\Delta_1}^2\ \mathrm{tr}(\mathbf{B}_i\mathbf{B}_j)) && (69)\\
&\quad -q^2(-(\Delta_\gamma\Delta_1 + \gamma^2\Delta_\gamma\Delta_1)\mathbf{h}_i^T\mathbf{h}_j + {\Delta_1}^2\ \mathrm{tr}(\mathbf{B}_i\mathbf{B}_j))\\
&\quad +2\rho q(-(\gamma\Delta_\gamma + \gamma\Delta_1)\mathbf{h}_i^T\mathbf{h}_j + \Delta_1\ \mathrm{tr}(\mathbf{B}_i\mathbf{B}_j))\\
&\quad -\rho q(-(\Delta_\gamma + \gamma^2\Delta_1)\mathbf{h}_i^T\mathbf{h}_j + \Delta_1\ \mathrm{tr}(\mathbf{B}_i\mathbf{B}_j))\\
&\quad +\rho^2(-2\gamma\mathbf{h}_i^T\mathbf{h}_j + \ \mathrm{tr}(\mathbf{B}_i\mathbf{B}_j))
\end{aligned}
$$

$$
\begin{aligned}
&= (q^2\Delta_1\Delta_\gamma(\gamma-1)^2\mathbf{h}_i^T\mathbf{h}_j + (\rho q\Delta_1 + \rho^2)\ \mathrm{tr}(\mathbf{B}_i\mathbf{B}_j) && (70)\\
&\quad +(\rho q(\Delta_\gamma + \gamma^2\Delta_1 - 2\gamma\Delta_\gamma - 2\gamma\Delta_1) - 2\rho^2\gamma)\mathbf{h}_i^T\mathbf{h}_j\\
&= (\gamma\Delta_1 + \Delta_\gamma)\mathbf{h}_i^T\mathbf{h}_j - \Delta_1\ \mathrm{tr}(\mathbf{B}_i\mathbf{B}_j)\\
&= (\gamma\Delta_1 + \Delta_\gamma)\mathbf{h}_i^T\mathbf{h}_j + (N-2)\Delta_1\mathbf{h}_i^T\mathbf{h}_j && (71)
\end{aligned}
$$

$$
[FIM]_{ij} \triangleq T_{11H}(t) + T_{12H}(t) + T_{21H}(t) + T_{22H}(t) + T_{23H}(t) \qquad (72)
$$

$$\begin{aligned} T_{12H}(t) &= \rho \ \mathrm{tr}(\mathbf{P}_t^2)\delta_{ij} \\ T_{21H}(t) &= a_i a_j \left(2q^2 \ \mathrm{tr}(\mathbf{D}_t^2\mathbf{G}_t^2) + 3\rho q \ \mathrm{tr}(\mathbf{D}_t\mathbf{G}_t^2) + \rho^2 \ \mathrm{tr}(\mathbf{G}_t^2)\right) \\ T_{22H}(t) &= \delta_{ij}(\rho q^2 \ \mathrm{tr}(\mathbf{D}_t\mathbf{G}_t^2) + \rho^2 q \ \mathrm{tr}(\mathbf{G}_t^2)) \\ T_{11H}(t) + T_{23H}(t) &= \phi\Delta_\gamma a_i a_j + (\phi\Delta_1 + \gamma\Delta_1 + (N-2)\Delta_1 + \Delta_\gamma)\mathbf{h}_i^T\mathbf{h}_j \end{aligned} \tag{73}$$

The formula in the theorem follows from the following observations.

We have by definition,

$$\mathbf{h}_i = \begin{bmatrix} \mathbf{w}_2^T\mathbf{J}_1\mathbf{e}_i \\ \vdots \\ \mathbf{w}_N^T\mathbf{J}_1\mathbf{e}_i \end{bmatrix}, \tag{74}$$

which implies

$$\mathbf{h}_i^T\mathbf{h}_j = \mathbf{e}_i^T\mathbf{J}_1^T\sum_{k=2}^{N}\mathbf{w}_k\mathbf{w}_k^T\mathbf{J}_1\mathbf{e}_j. \tag{75}$$

Also, since $\mathbf{a} = \mathbf{J}_1^T\mathbf{w}_1$, it is clear that

$$a_i a_j = [\mathbf{J}_1^T\mathbf{w}_1\mathbf{w}_1^T\mathbf{J}_1]_{ij}. \tag{76}$$

The formulas above allow us to express the FIM in closed form instead of expressing each of its elements.

Proof of Theorem 4

Consider the received signal for one block

$$\mathbf{Y}(t) = \mathbf{A}\sum_{i=1}^{N}\mathbf{X}_i s_i(t) + \mathbf{N}(t). \tag{77}$$

Consider first the case $m = N$, *i.e.*, the code matrices are square.

We saw that we can assume *wlog* that $\mathbf{X}_1 = \mathbf{I}_N$. In order to show that the FIM does not change when we change the symbol in which the pilot is inserted, we consider the following two signals :

$$\begin{aligned} \mathbf{Y}^{(1)}(t) &= \mathbf{A}\mathbf{X}_1 s_0(t) + \mathbf{A}\sum_{i=2}^{N}\mathbf{X}_i u_i(t) + \mathbf{N}(t) \\ \mathbf{Y}^{(k)}(t) &= \mathbf{A}\mathbf{X}_k s_0(t) + \mathbf{A}\sum_{i=1, i\neq k}^{N}\mathbf{X}_i u_i(t) + \mathbf{N}(t). \end{aligned}$$

Observe that

$$\begin{aligned} \mathbf{Y}^{(k)}(t)\mathbf{X}_k^T &= \mathbf{A}s_0(t) + \mathbf{A}\sum_{i=1, i\neq k}^{N}\mathbf{X}_i\mathbf{X}_k^T u_i(t) + \mathbf{N}(t) \\ &= \mathbf{A}\mathbf{X}_1 s_0(t) + \mathbf{A}\sum_{i=1, i\neq k}^{N}\mathbf{X}_i\mathbf{X}_k^T u_i(t) + \mathbf{N}(t) \end{aligned} \tag{78}$$

From lemma 1, the family $\{\mathbf{Z}_i = \mathbf{X}_i \mathbf{X}_k^T, \, , i \neq k\}$ is the same as $\{\mathbf{X}_i, i \in \{2, \ldots, N\}\}$, up to the sign of matrices. Since the distribution of u_i is symmetric with respect to zero, the sign does not affect the FIM of the parameters. Also, it is easy to check that the transformation applied preserves the covariance matrix of the noise. This proves the theorem for $m = N$.

In the case $m < N$, using the same arguments as above we can consider the matrices $\mathbf{X}_i$, $i \in \{1, \ldots, N\}$ such that $\mathbf{G}_1 = \mathbf{I}_N$, see lemma 2. The theorem follows by multiplying $\mathbf{Y}^{(k)}(t)$ by $\mathbf{G}_k^T \mathbf{G}_1$ and applying lemma 2.

Proof of Theorem 5

The proof of the theorem is based on the formulas derived in theorems 2 and 3. In this subsection the terms that have the subindex H and V are for the horizontal and vertical placement scheme respectively.

It is easy to observe that $T_{12H} = T_{12V}$, $T_{21H} = T_{21V}$ and $T_{22H} = T_{22V}$. We need to show that $T_{11H} + T_{23H} = T_{11V} + T_{23V}$. This last statement follows if we compute separately each of the two sides of the relation. Below are the details.

In the special case of uniform placement scheme, we substitute $\phi = 1 - \gamma$ in formula (72) to get:

$$T_{11H}(t) = (1-\gamma)\Delta_1 \mathbf{h}_i^T \mathbf{h}_j + a_i a_j (1-\gamma)\Delta_\gamma$$

$$T_{11H} = N\, T_{11H}(t)$$

$$T_{23H} = N\, T_{23H}(t)$$

and further,

$$\begin{aligned} T_{11H} + T_{23H} &= N(1-\gamma)\Delta_\gamma a_i a_j + N\Delta_1 \mathbf{h}_i^T \mathbf{h}_j \\ &\quad + N\Delta_\gamma \mathbf{h}_i^T \mathbf{h}_j + N(N-2)\Delta_1 \mathbf{h}_i^T \mathbf{h}_j \\ &= N(1-\gamma)\Delta_\gamma a_i a_j + N(N-1)\Delta_1 \mathbf{h}_i^T \mathbf{h}_j + N\Delta_\gamma \mathbf{h}_i^T \mathbf{h}_j. \end{aligned} \tag{79}$$

Now we have to compute the FIM for the uniform vertical transmission scheme.

$$\mathbf{G}_1 = \gamma \mathbf{I}\,,\; \mathbf{G}_t|_{t>1} = \mathbf{I}\,, \tag{80}$$

$$\mathbf{D}_1 = \Delta_\gamma \mathbf{I}\,,\; \mathbf{D}_t|_{t>1} = \Delta_1 \mathbf{I}\,, \tag{81}$$

$$\mathbf{P}_1 = \sqrt{1-\gamma}\mathbf{I}\,,\; \mathbf{P}_t|_{t>1} = 0. \tag{82}$$

$$
\begin{aligned}
T_{11V} &= T_{11V}(1) \\
&= \mathbf{v}_t^T \mathbf{P}_t(-\mathbf{H}_i + a_i\mathbf{I})\mathbf{D}_t(-\mathbf{H}_j + a_j\mathbf{I})\mathbf{P}_t\mathbf{v}_t \\
&= -\mathbf{v}_t^T \mathbf{P}_t\mathbf{H}_i\mathbf{D}_t\mathbf{H}_j\mathbf{P}_t\mathbf{v}_t + a_i a_j \mathbf{v}_t^T\mathbf{P}_t\mathbf{D}_t\mathbf{P}_t\mathbf{v}_t \\
&= -(1-\gamma)\Delta_\gamma \mathbf{v}_t^T\mathbf{H}_i\mathbf{H}_j\mathbf{v}_t + (1-\gamma)\Delta_\gamma a_i a_j N \\
&= -(1-\gamma)\Delta_\gamma \sum (\mathbf{H}_i)_{k_1,k_2}(\mathbf{H}_j)_{k_1,k_2} + (1-\gamma)\Delta_\gamma a_i a_j N \\
&= N(1-\gamma)\Delta_\gamma \mathbf{h}_i^T\mathbf{h}_j + N(1-\gamma)\Delta_\gamma a_i a_j \qquad (83)
\end{aligned}
$$

The relation (83) follows from proposition (2).

Calculate the expression of T_{23} in the vertical case.

$$
\begin{aligned}
T_{23V} &= T_{23V}(1) + (N-1)T_{23V}(t>1) \\
&= \gamma^2(\Delta_\gamma \rho q + \rho^2)\ \mathrm{tr}(\mathbf{H}_i\mathbf{H}_j) + (N-1)(\Delta_1\rho q + \rho^2)\ \mathrm{tr}(\mathbf{H}_i\mathbf{H}_j) \\
&= (-\gamma\Delta_\gamma - (N-1)\Delta_1)(-2\mathbf{h}_i^T\mathbf{h}_j + \ \mathrm{tr}(\mathbf{B}_i\mathbf{B}_j)) \\
&= -N(-\gamma\Delta_\gamma - (N-1)\Delta_1)\mathbf{h}_i^T\mathbf{h}_j \qquad (84)
\end{aligned}
$$

$$
T_{11V} + T_{23V} = N(1-\gamma)\Delta_\gamma a_i a_j + N(N-1)\Delta_1 \mathbf{h}_i^T\mathbf{h}_j + N\Delta_\gamma \mathbf{h}_i^T\mathbf{h}_j \qquad (85)
$$

Proof of Theorem 6

Consider first the case $m = N$ for which $\mathbf{J}_1 = \mathbf{I}$ so that the formula of the FIM simplifies considerably. We'll show that the eigenvalues $\{g_0, g_0 + q * g_1, g_0 + q * g_2\}$ of the matrix $\mathbf{F}(\gamma)$ are convex functions (of γ). For each of the functions we'll separate the terms that are linear in γ (we'll denote the coefficients with θ_i) and calculate the second derivative of the nonlinear part.

Introduce the following notation :

$$
\xi \triangleq \frac{q}{\sigma^2} \qquad (86)
$$

Note that ξ can be interpreted as the average SNR at the receiver. With this notation Δ_γ and Δ_1 be come

$$
\Delta_\gamma = -\frac{\gamma}{\sigma^2(q\gamma + \sigma^2)} = -\rho^2\frac{\gamma}{\xi\gamma + 1};\ \Delta_1 = \Delta_\gamma|_{\gamma=1} = -\rho^2\frac{1}{\xi + 1} \qquad (87)
$$

With the notation above, taking into account that Δ_1 does not depend on γ, we obtain :

$$\begin{aligned} g_0 &= (N - \text{tr}(\mathbf{G}_1)) + (\rho q^2 \text{ tr}(\mathbf{D}_1\mathbf{G}_1^2) + \rho^2 q \text{ tr}(\mathbf{G}_1^2)) \\ &= \theta_1\gamma + \theta_0 + \rho q^2 \Delta_\gamma \gamma^2 + \rho^2 q\gamma^2 \\ &= \theta_1\gamma + \theta_0 - \rho q^2 \rho^2 \frac{\gamma^3}{\xi\gamma+1} + \rho^2 q\gamma^2 \\ &= \theta_1\gamma + \theta_0 + \rho^2 q \left(\gamma^2 - \xi\frac{\gamma^3}{\xi\gamma+1}\right) \\ &= \theta_1\gamma + \theta_0 + \rho^2 q \frac{\gamma^2}{\xi\gamma+1} \end{aligned} \tag{88}$$

$$\frac{\text{d}^2 g_0}{\text{d}\gamma^2} = \rho^2 q \frac{2}{(\xi\gamma+1)^3} \tag{89}$$

Taking into account the range of the variables for our problem, it follows that g_0 is convex.

$$\begin{aligned} g_1 &= \left(2q^2 \text{ tr}(\mathbf{D}_t^2\mathbf{G}_t^2) + 3\rho q \text{ tr}(\mathbf{D}_t\mathbf{G}_t^2) + \rho^2 \text{ tr}(\mathbf{G}_t^2)\right) \\ &\quad +(1-\gamma)\Delta_\gamma \\ &= \left(2q^2\rho^4 \frac{\gamma^4}{(\xi\gamma+1)^2} - 3\rho^2 q\frac{\gamma^3}{\xi\gamma+1} + \rho^2\gamma^2 - \frac{(1-\gamma)\gamma}{\xi\gamma+1}\right) \\ &\quad +\theta_0\gamma + \theta_1 \\ &= \rho^2\left(2\xi^2 \frac{\gamma^4}{(\xi\gamma+1)^2} - 3\xi\frac{\gamma^3}{\xi\gamma+1} + \gamma^2 - \frac{(1-\gamma)\gamma}{\xi\gamma+1}\right) \\ &\quad +\theta_0\gamma + \theta_1 \end{aligned} \tag{90}$$

$$\frac{\text{d}^2 g_1}{\text{d}\gamma^2} = 2\frac{\gamma(\xi^2 - 4\xi) + 2 + \xi}{(\xi\gamma+1)^4} \tag{91}$$

Usually, even for high noise powers $\xi > 4$ which makes the function g_1 convex wrt γ.

However, we'll show that the eigenvalue $h_1 = g_0 + qg_1$ is a convex function in γ for any value of ξ.

$$\begin{aligned} \frac{\text{d}^2 h_1}{\text{d}\gamma^2} &= q\frac{\text{d}^2 g_1}{\text{d}\gamma^2} + \frac{\text{d}^2 g_0}{\text{d}\gamma^2} \\ &= 2\rho^2 q^2 \frac{\gamma(\xi^2 - 4\xi) + 2 + \xi}{(\xi\gamma+1)^4} + 2\rho^2 q\frac{1}{(\xi\gamma+1)^3} \\ &= 2\xi^2 \frac{\gamma(\xi^2 - 3\xi) + 3 + \xi}{(\xi\gamma+1)^4} \end{aligned} \tag{92}$$

It is easy to observe that the denominator of the expression above is positive for any value of $\gamma \in [0, 1]$.

$$g_2 = ((N-1)\Delta_1 + \Delta_\gamma) \tag{93}$$

$$\frac{\mathsf{d}^{\,2} g_2}{\mathsf{d}\ \gamma^2} = \rho^2 \frac{2\xi}{(\xi\gamma+1)^3} \tag{94}$$

$$\tag{95}$$

g_2 is clearly convex wrt γ thus h_2 is the same.

If $m < N$ the relations above hold but $\{g_0, g_0 + q * g_1, g_0 + q * g_2\}$ are not the eigenvalues of the matrix $\mathbf{F}(\gamma)$ anymore. Instead, we can write,

$$\mathbf{F}(\gamma) = \mathbf{J}_1^T \tilde{\mathbf{F}}(\gamma)\mathbf{J}_1,$$

and then $\{g_0, g_0 + q * g_1, g_0 + q * g_2\}$ are the eigenvalues of $\tilde{\mathbf{F}}(\gamma)$. Thus we have

$$0 < \tilde{\mathbf{F}}(\gamma - \zeta) + \tilde{\mathbf{F}}(\gamma + \zeta) - 2\tilde{\mathbf{F}}(\gamma). \tag{96}$$

Since the matrix $\mathbf{J}_1$ is tall and full column rank for any positive definite matrix $\mathbf{U}$ the matrix $\mathbf{J}_1^T \mathbf{U} \mathbf{J}_1$ is also positive definite (see [14]). The formula in the theorem follows from (96).

Proof of Theorem 7

We need to how the FIM matrix for the horizontal power allocation scheme described below varies with ϕ :

$$\boldsymbol{\phi}_1 \triangleq [1+\phi, 0, \ldots, 0]^T,\ \boldsymbol{\gamma}_1 \triangleq [0, 1, \ldots, 1]^T, \tag{97}$$

$$\boldsymbol{\phi}_2 \triangleq [1-\zeta-\phi, 0, \ldots, 0]^T,\ \boldsymbol{\gamma}_2 \triangleq [\gamma, 1, \ldots, 1]^T. \tag{98}$$

Like in the proof of theorem 6, consider first that $m = N$ so that $\mathbf{J}_1 = \mathbf{I}$.

Unlike the previous theorem, here we don't have a symmetry in ϕ so we need to analyze the functions $g_k = g_k(1) + g_k(2)$, that are the eigenvalues of the FIM for two blocks.

$$g_0 = g_0(1) + g_0(2) = \rho(1 + \phi + 1 - \phi - \zeta) + \theta \tag{99}$$

$$\frac{\mathsf{d}\ g_0}{\mathsf{d}\ \phi} = 0 \tag{100}$$

This term is constant in ϕ.

$$g_1 = g_1(1) + g_1(2) = (1+\phi)\Delta_\gamma|_{\gamma=0} + (1 - \phi - \zeta + \theta)\Delta_\gamma$$

$$= (1 - \phi - \zeta + \theta)\Delta_\gamma \tag{101}$$

$$\frac{\mathsf{d}\ g_1}{\mathsf{d}\ \phi} = -\Delta_\gamma \tag{102}$$

Thus g_1 is an increasing function of ϕ.

It is easy to see that $g_2 = g_2(1) + g_2(2)$ does not depend on ϕ. Thus any increase of ϕ increases one of the eigenvalues of the FIM which implies that the CRLB is improved. Observe that $1 - \phi - \zeta > 0$ which implies that this theorem applies only to the cases in which $0 < \phi < 1$.

Also observe that $\gamma = 0$ implies $\Delta_\gamma = 0$, thus the eigenvalues are constant in this case.

For $m < N$ the statement in the theorem follows straightforward, like in the proof of theorem 6.

References

[1] A.F. Naguib, V. Tarokh, N. Seshadri and A.R. Calderbank, "A space-time coding modem for high-data-rate wireless communications". *IEEE Journal on Selected Areas in Communications*, 16(8):1459 -1478,1998.

[2] V. Tarokh; H. Jafarkhani and A.R. Calderbank, "Space-time block codes from orthogonal designs". *IEEE Trans. on Information Theory*, 45(5):1456-1467, 1999.

[3] T.P. Holden and K. Feher, "A spread spectrum based system technique for synchronization of digital mobile communication systems". *IEEE Trans. Broadcasting,Proceedings-Communications*, 36(3):185-194, 1990.

[4] P. Hoeher and F. Tufvesson, "Channel estimation with superimposed pilot sequence". In *GLOBECOM 99*, pp 2162 -2166, vol.4, Rio de Janeiro, Brazil, Dec 1999.

[5] J.H. Manton; I.Y. Mareels and Y. Hua, "Affine precoders for reliable communications ". In *ICASSP 00*, Vol. 5, pp. 2749 -2752, Istanbul, Turkey, June 200.

[6] D. Makrakis and K. Feher, "A novel pilot insertion-extraction technique based on spread spectrum Affine precoders for reliable communications ". presented at *Miami Technicon*, Miami, 1987.

[7] G.Ganesan and P. Stoica, "Space-Time Diversity". in *Signal Processing Advances in Wireless and Mobile Communications*, Prentice Hall, 2001, pp 59-88.

[8] S. Ohno and G.B. Giannakis, "Optimal Training and Redundant Precoding for Block Transmissions with Application to Wireless OFDM" *IEEE Transactions on Communications*, submitted November 2000.

[9] S. Adireddy and L. Tong "Optimal Placement of Known Symbols for Nonergodic Broadcast Channels" submitted to *IEEE Transactions on Information Theory*, August 28, 2001, downloadable at *http://www.ee.cornell.edu/∼ltong/pubJ.html.*

[10] S. Adireddy and L. Tong "Detection with Embedded Known Symbols: Optimal Symbol Placement and Equalization" *Proceedings of the 2000 Intl Conf. Acoust. Speech and Sig. Proc.*, vol. V, pp. 2541-2543, Istanbul, Turkey, June 5-9, 2000.

[11] R. Negi and J. Cioffi "Pilot Tone Selection for Channel Estimation in a Mobile OFDM System" *IEEE Transactions on Consumer Electronics*, 11(3):1122-1128, Aug. 1998

[12] M. Dong and L. Tong "Optimal Design and Placement of Pilot Symbols for Channel Estimation" submitted to *IEEE Trans. on Signal Processing*, April, 2001, downloadable at *http://www.ee.cornell.edu/∼ltong/pubJ.html.*

[13] S. Adireddy, L. Tong and H. Viswanathan "Optimal Placement of Training for Unknown Channels", submitted to *IEEE Transactions on Information Theory*, March, 2001. downloadable at *http://www.ee.cornell.edu/∼ltong/pubJ.html.*

[14] R.A. Horn and C.R. Johnson *Matrix Analysis*, Cambridge University Press, 1985.

[15] T.L. Marzetta "BLAST Training: Estimating Channel Characteristics for High Capacity Space-Time Wireless" *Proceedings 37-th Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, Sept. 22-24, 1999.

[16] Liu, Z.; Giannakis, G.B.; Barbarossa, S.; Scaglione, A. "Transmit antennae space-time block coding for generalized OFDM in the presence of unknown multipath" *IEEE Journal on Selected Areas in Communications*, Volume: 19 Issue: 7 , July 2001, Page(s): 1352 -1364

[17] Muquet, B.; Zhou, S.; Giannakis, G.B. "Subspace-based estimation of frequency-selective channels for space-time block precoded transmissions" *Conference Record of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computers*, Volume: 2 , 2000,Page(s): 975 -979

[18] Swindlehurst, A.L. Simultaneous channel estimation and decoding for diagonal space-time codes *Proceedings of the 2000 IEEE Sensor Array and Multichannel Signal Processing Workshop*. 2000, Page(s): 173 -177.

[19] De Carvalho, E.; Slock, D.T.M. Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation *First IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, 1997 Page(s): 129 -132.

[20] Min Dong; Lang Tong Optimal design and placement of pilot symbols for channel estimation *Proceedings of 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Volume: 4 , 2001 Page(s): 2109 -2112.

[21] Hassibi, B.; Hochwald, B.M. How much training is needed in multiple antenna wireless links ? *IEEE Transactions on Information Theory*,submitted August 2000

[22] Marzetta, T.L.; Hochwald, B.M. Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading *IEEE Transactions on Information Theory*, Volume: 45 Issue: 1 , Jan. 1999, Page(s): 139 -157.

[23] I.E. Telatar, Capacity of multi antenna Gaussian channels *European Transactions on Telecommunications*, 10(6), Nov/Dec. 1999, downloadable at *http://mars.bell-labs.com/cm/ms/what/mars/index.html*.

[24] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, 1993.

Contents

List of Figures