

Channel Estimation for Space-Time Orthogonal Block Codes

Cristian Budianu, *Student Member, IEEE*, and Lang Tong, *Senior Member, IEEE*

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 the 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—Channel estimation, Cramér–Rao bound, orthogonal designs, pilot symbols, power allocation, space-time codes.

I. INTRODUCTION

A MAJOR challenge in wireless space-time communications is coping with channel uncertainties. While Shannon theory does not mandate channel estimation [1], 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 [2]–[6]. 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 that 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 [7] that the channel estimation errors can be reduced significantly by using semiblind techniques that 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 and designing pilot symbols optimally provides gain in channel capacity [8]–[10] and reduction of symbol and channel estimation errors [11]–[14].

A more general form of training that allows the superposition of pilot and data symbols has attracted attention recently [10], [15]–[17]. Such schemes, proposed earlier in [18] and [19], 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 *et al.* [20]. 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

Manuscript received October 1, 2001; revised June 10, 2002. 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. The associate editor coordinating the review of this paper and approving it for publication was Dr. Thomas L. Marzetta.

The authors are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: ltong@ece.cornell.edu).

Publisher Item Identifier 10.1109/TSP.2002.803340.

in general. For the orthogonal codes presented in [20], 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 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 the maximum likelihood (ML) technique lends support for the use of the 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 the paper in Section VI. The proofs of the theorems are presented in the Appendix.

The notations used in this paper are as follows. 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, and $\Re\{\cdot\}$ and $\Im\{\cdot\}$ 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. $\text{cov}(\mathbf{y})$ 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 \mathbf{B} square matrices means that their difference $\mathbf{A} - \mathbf{B}$ is positive definite. Similarly, $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite. $[\mathbf{A}]_{ij}$ means the element of \mathbf{A} with coordinates (i, j) , and the vector \mathbf{e}_k is the k th vector of the standard basis.

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 k th

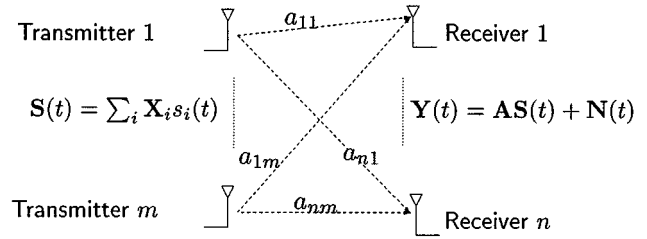


Fig. 1. m -transmitter n -receiver space-time system.

column of $\mathbf{S}(t)$ corresponds to the transmitted vector in the k th symbol interval. The space-time code proposed by Tarokh *et al.* [20] has the form

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

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

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

The theory of orthogonal designs also shows that (see [20]) for rate one codes and real symbols, the family $\{\mathbf{X}_i, i = 1, \dots, N\}$ exists if and only if N is 2, 4, or 8. It is shown in [21] that using N single-user detectors in parallel, the choice of the matrices $\{\mathbf{X}_i, i = 1, \dots, N\}$, as above, provides the best SNR.

Under the *quasistatic flat fading* model with coherence time of B blocks, the received signal matrix for the t th block is given by

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

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

$$\mathbf{y}(t) \triangleq \text{vec}(\mathbf{Y}^T(t)), \quad \mathbf{n}(t) \triangleq \text{vec}(\mathbf{N}^T(t)) \quad (4)$$

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

The received signal in one block can then be written as

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

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

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

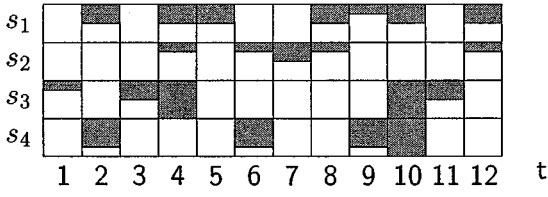


Fig. 2. General power allocation scheme under the power constraint (11).

Thus, a system with n receiving antennas and complex channel coefficients is equivalent to a system with $2n$ receiving antennas with Gaussian noise $\mathcal{N}(\mathbf{0}, (\sigma^2/2)\mathbf{I})$. 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 i th transmitted symbol $s_i(t)$ of block t is expressed as a linear combination of a pilot symbol and a data symbol

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

where $v_i(t)$ is the known pilot taking values from $\{\pm 1\}$, and $u_i(t)$ is 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 coherence time of B blocks can be completely specified by the two $N \times B$ power allocation matrices

$$\Phi = [\phi_1, \dots, \phi_B], \quad \phi_t = [\phi_{1t}, \dots, \phi_{Nt}]^T \quad (8)$$

$$\Gamma = [\gamma_1, \dots, \gamma_B], \quad \gamma_t = [\gamma_{1t}, \dots, \gamma_{Nt}]^T. \quad (9)$$

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.,

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

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

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

Power allocation schemes that specify training schemes can be illustrated graphically. Fig. 2 shows 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 three-dimensional (3-D) bar-diagram may be necessary.

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 $\{\Phi, \Gamma\}$ satisfy

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

The *periodic horizontal placement* is a horizontal placement that repeats itself every N blocks; see Fig. 3(b). The *uniform horizontal placement*, which is 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* [see Fig. 3(d)] and *uniform periodic vertical* placements are defined similarly. The *periodic uniform vertical placement* has the pilot symbols 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 $\gamma = 0$.

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., $u_i(t) \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$.
- A4) The 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 $\mathbf{X}_1 = [\mathbf{I}_m; \mathbf{0}_{m \times N-m}]$.

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

$$\boldsymbol{\mu} = [\boldsymbol{\mu}_1^T, \dots, \boldsymbol{\mu}_B^T]^T$$

$$\boldsymbol{\mu}_t \triangleq \mathbb{E}\{\mathbf{y}(t)\} = \sum_{i=1}^N \mathbf{w}_i \sqrt{\phi_{it}} v_i(t) \quad (12)$$

$$\mathbf{C} = \text{diag}[\mathbf{C}_{11}, \dots, \mathbf{C}_{BB}]$$

$$\mathbf{C}_{tt} \triangleq \text{cov}\{\mathbf{y}(t)\} = \sum_{i=1}^N \mathbf{w}_i \mathbf{w}_i^T \gamma_{it} + \sigma^2 \mathbf{I}_{nN}. \quad (13)$$

The assumption that the 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 [22]. 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.

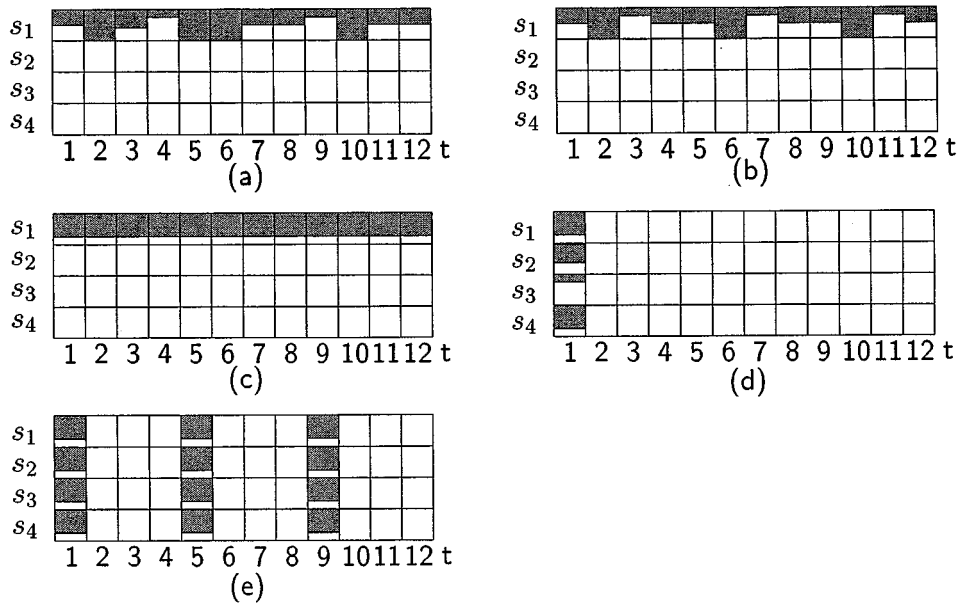


Fig. 3. Power allocation schemes defined previously. (a) (Nonuniform) horizontal scheme. (b) (Nonuniform) periodic horizontal scheme. (c) Uniform horizontal scheme. (d) (Nonuniform) vertical. (e) Periodic uniform vertical.

D. 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 as 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 $I(\mathbf{y}; s_1(1), \dots, s_N(1), \dots, s_1(B), \dots, s_N(B))$ 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

$$\gamma_{kt} = 1 - \frac{P_{tr}}{NB} \quad \forall k = 1 \dots N, t = 1 \dots B \quad (14)$$

maximizes the mutual information between the input and the output.

Proof: See the Appendix. \square

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 CRLB

For our problem, the Fisher information matrix (FIM) is defined as

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

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

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

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 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

$$\mathcal{F}(\mathbf{\Phi}, \mathbf{\Gamma}) = \sum_{t=1}^B \mathcal{F}_t(\phi_t, \gamma_t) \quad (15)$$

where $\mathcal{F}_t(\phi_t, \gamma_t)$ is the FIM defined for block t . Under the Gaussian assumption [A1) and A3)], the FIM has a well-known special form given by [23]

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

where a_i is the i th component of vector \mathbf{a} .

TABLE I
 NOTATIONS IN THE CRLB EXPRESSIONS

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

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.

Theorem 2: Consider one block (of index t) of a power allocation scheme under APC. The FIM for estimating the channel from the received vector $\mathbf{y}(t)$ is given by

$$[\mathcal{F}_t(\boldsymbol{\phi}_t, \boldsymbol{\gamma}_t)]_{ij} = [\mathbf{T}_1(t)]_{ij} + [\mathbf{T}_2(t)]_{ij} \quad (16)$$

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

$$[\mathbf{T}_1(t)]_{ij} \triangleq \mathbf{v}_t^T \mathbf{P}_t (-\mathbf{H}_i + a_i \mathbf{I}) \mathbf{D}_t (\mathbf{H}_j + a_j \mathbf{I}) \mathbf{P}_t \mathbf{v}_t + \rho \text{tr}(\mathbf{P}_t^2) \delta_{ij} \quad (17)$$

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

$$\begin{aligned} [\mathbf{T}_2(t)]_{ij} &= a_i a_j (2q^2 \text{tr}(\mathbf{D}_t^2 \mathbf{G}_t^2) + 3\rho q \text{tr}(\mathbf{D}_t \mathbf{G}_t^2) + \rho^2 \text{tr}(\mathbf{G}_t^2)) \\ &+ \delta_{ij} (\rho q^2 \text{tr}(\mathbf{D}_t \mathbf{G}_t^2) + \rho^2 q \text{tr}(\mathbf{G}_t^2)) \\ &+ q^2 \text{tr}(\mathbf{D}_t \mathbf{H}_i \mathbf{G}_t \mathbf{D}_t \mathbf{H}_j \mathbf{G}_t) - q^2 \text{tr}(\mathbf{D}_t \mathbf{H}_i \mathbf{G}_t^2 \mathbf{D}_t \mathbf{H}_j) \\ &+ 2\rho q \text{tr}(\mathbf{D}_t \mathbf{H}_i \mathbf{G}_t \mathbf{H}_j \mathbf{G}_t) - \rho q \text{tr}(\mathbf{D}_t \mathbf{H}_i \mathbf{G}_t^2 \mathbf{H}_j) \\ &+ \rho^2 \text{tr}(\mathbf{G}_t \mathbf{H}_i \mathbf{G}_t \mathbf{H}_j). \end{aligned} \quad (18)$$

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. Key steps are given in the Appendix with detailed derivation in [24]. \square

The formula in (18) 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 one block (of index t) of a horizontal power allocation scheme under APC, where $N \in \{2, 4\}$, 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

$$\begin{aligned} \mathcal{F}_t(\boldsymbol{\gamma}_t, \boldsymbol{\phi}_t) &= g_0(t) \mathbf{I} + g_1(t) \mathbf{J}_1^T \mathbf{w}_1 \mathbf{w}_1^T \mathbf{J}_1 \\ &+ g_2(t) \sum_{k=2}^N \mathbf{J}_1^T \mathbf{w}_k \mathbf{w}_k^T \mathbf{J}_1 \end{aligned} \quad (19)$$

$$g_0(t) \triangleq \rho \phi + (\rho q^2 \text{tr}(\mathbf{D}_t \mathbf{G}_t^2) + \rho^2 q \text{tr}(\mathbf{G}_t^2)) \quad (20)$$

$$g_1(t) \triangleq (2q^2 \text{tr}(\mathbf{D}_t^2 \mathbf{G}_t^2) + 3\rho q \text{tr}(\mathbf{D}_t \mathbf{G}_t^2) + \rho^2 \text{tr}(\mathbf{G}_t^2)) + \phi \Delta_\gamma \quad (21)$$

$$g_2(t) \triangleq \phi \Delta_1 + \gamma \Delta_1 + (N-2) \Delta_1 + \Delta_\gamma \quad (22)$$

where

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

Proof: See the Appendix. \square

Observe that for a (nonuniform) horizontal scheme with B blocks, we have

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

where

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

It can be observed that the formula of the FIM for the horizontal placement scheme does not depend on the training symbols \mathbf{v}_t . Since the analysis that follows is about horizontal placement schemes, the influence of the training sequence is not considered.

One might ask why the theorem above does not hold for $N = 8$. The proof of the theorem is based on the key property of the family of code matrices that is stated in Lemma 1, which is not true for $N = 8$. This property can be interpreted as the connection between the STBC families of matrices for $N = \{2, 4\}$ and the complex numbers and quaternionic numbers, respectively.

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

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

where

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

The trace of the CRLB matrix is given by

$$\text{tr}(\mathcal{R}) = N d_0 + q(d_1 + (N-1)d_2). \quad (27)$$

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

IV. BEHAVIOR OF FIM AND CRLB

A. Equivalent Power Allocation Schemes

We present two theorems that reveal some equivalent power allocation schemes.

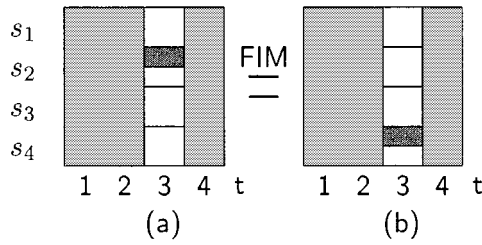


Fig. 4. Equivalent power allocation schemes in Theorem 4. The positions shaded by light gray are where power allocations can be arbitrary.

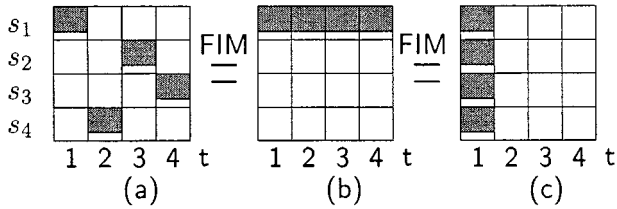


Fig. 5. Power allocation schemes in Theorems 4 and 5. (a) Periodic scheme that contains one training symbol in each block, where all training symbols have equal power. (b) Uniform horizontal scheme. (c) Uniform periodic vertical scheme.

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

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

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

Proof: See the Appendix. \square

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) and (b)]. Not so obvious is that this is not true in general for $N = 8$ and if more than one pilot symbol is transmitted within each block. In those cases, it is possible that inserting 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) and (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. \square

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 identification of the channel matrix. For the uniform horizontal placement, however, unknown data symbols are 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 considered, and it appears that these special codes provide 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. A direct consequence of the theorem stated previously is that for the uniform vertical placement scheme, the FIM does not depend on the training symbols used.

B. 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?

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

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

Proof: See the Appendix. \square

If we apply this theorem to two blocks, it is then apparent that one should allocate training power unevenly, as illustrated in Fig. 6. It is then a direct consequence of the convex property in Theorem 6 that one should avoid superimposing pilot with data symbol and that 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 the average power constraint (APC) is used, it is 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 blocks is 100% and in 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 of a power allocation scheme under APC, and denote

$$\phi_{t_1} \triangleq [1, 0, \dots, 0]^T, \quad \boldsymbol{\gamma}_{t_1} \triangleq [0, 1, \dots, 1]^T \quad (30)$$

$$\phi_{t_2} \triangleq [1 - \zeta, 0, \dots, 0]^T, \quad \boldsymbol{\gamma}_{t_2} \triangleq [\gamma, 1, \dots, 1]^T \quad (31)$$

where $0 \leq \gamma \leq \zeta \leq 1$. Then, $\forall \phi_1, \phi_2$ such that $0 < \phi_1 < \phi_2 < 1 - \zeta$, we have

$$\begin{aligned} \mathcal{F}(\boldsymbol{\gamma}_{t_1}, \boldsymbol{\gamma}_{t_2}, \phi_{t_1} + \phi_1 \mathbf{e}_1, \phi_{t_2} - \phi_1 \mathbf{e}_1) \\ \leq \mathcal{F}(\boldsymbol{\gamma}_{t_1}, \boldsymbol{\gamma}_{t_2}, \phi_{t_1} + \phi_2 \mathbf{e}_1, \phi_{t_2} - \phi_2 \mathbf{e}_1). \end{aligned} \quad (32)$$

It follows immediately that

$$\begin{aligned} \text{tr}(\mathcal{R}(\boldsymbol{\gamma}_{t_1}, \boldsymbol{\gamma}_{t_2}, \phi_{t_1} + \phi_1 \mathbf{e}_1, \phi_{t_2} - \phi_1 \mathbf{e}_1)) \\ \geq \text{tr}(\mathcal{R}(\boldsymbol{\gamma}_{t_1}, \boldsymbol{\gamma}_{t_2}, \phi_{t_1} + \phi_2 \mathbf{e}_1, \phi_{t_2} - \phi_2 \mathbf{e}_1)). \end{aligned} \quad (33)$$

Proof: See the Appendix. \square

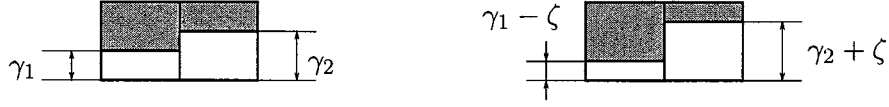

 Fig. 6. Power allocation schemes compared in Theorem 6 for $N = 4$.


Fig. 7. Power allocation schemes compared in Theorem 7.



Fig. 8. Power allocation schemes compared in Conjecture 1.

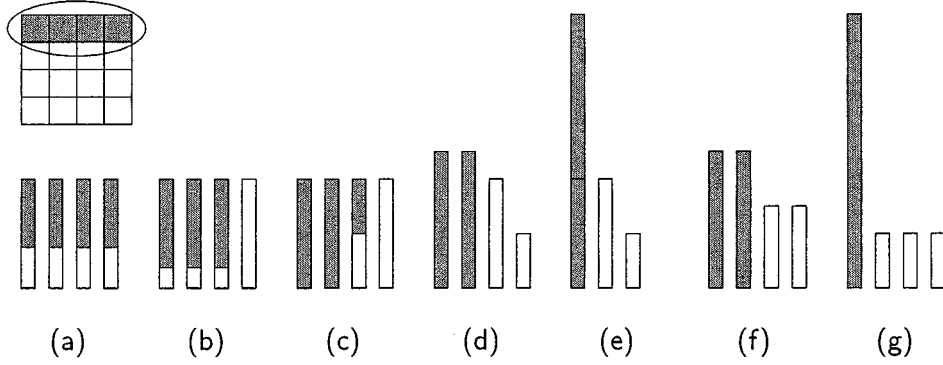


Fig. 9. Power allocation schemes compared. Application of convexity theorem.

We note that from the proof of the theorem [see (74) and (76) and the comment that follows], 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.

From Theorem 7, it follows that the power allocation scheme that is optimal under PPC can be improved by concentrating the training power to fewer pilot symbols if one allows supraunitary power for some symbols. It also shows that superimposed horizontal training cannot 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 two blocks t_1 and t_2 of a power allocation scheme under APC:

$$\phi_t = [\phi, 0, \dots, 0]^T, \quad t = 1, 2 \quad (34)$$

$$\gamma_1 = [\gamma + \zeta, 1, \dots, 1]^T \quad (35)$$

$$\gamma_2 = [\gamma - \zeta, 1, \dots, 1]^T \quad (36)$$

where $\gamma + \zeta + \phi \leq 1$. Then, $\forall \phi, \gamma, \zeta$ such that $0 < \phi \leq 1 - \gamma - \zeta$ and $0 < \gamma - \zeta \leq \gamma + \zeta < 1$, we have

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

In the conjecture above, the assumption $m = N$ is important because this allows us to use (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, where $\zeta_1 < \zeta_2$. These schemes are represented in Fig. 8. Then $\text{tr}(\mathcal{R}(\gamma, \zeta_1, \phi)) \geq \text{tr}(\mathcal{R}(\gamma, \zeta_2, \phi))$.

C. Summary Scenarios

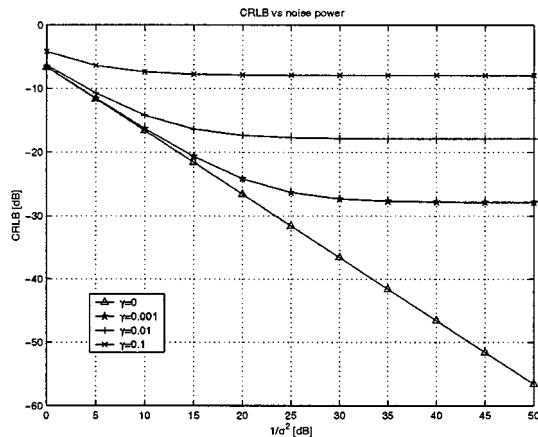
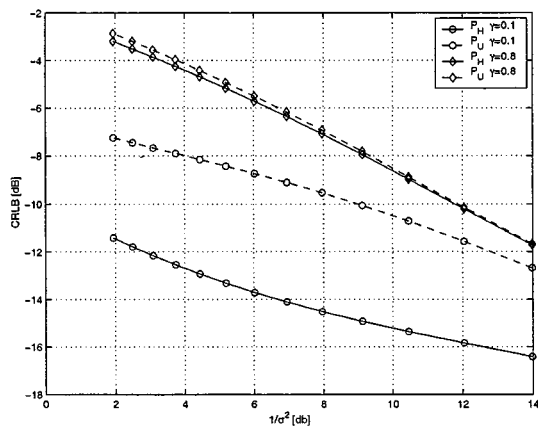
Fig. 9 summarizes our results graphically. Fig. 9(a) is the uniform horizontal scheme that can be improved by Theorem 6 to Fig. 9(c) via Fig. 9(b). This is the best under PPC. If APC is used, Fig. 9(c) can then be improved to Fig. 9(d) using Theorem 7. The same theorem says that the scheme in Fig. 9(e)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

V. SIMULATIONS AND NUMERICAL RESULTS

A. CRLB Under Superimposed Training

In this section, we want to investigate the effect of the amount of power allocated to training and the power allocation scheme on 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

□

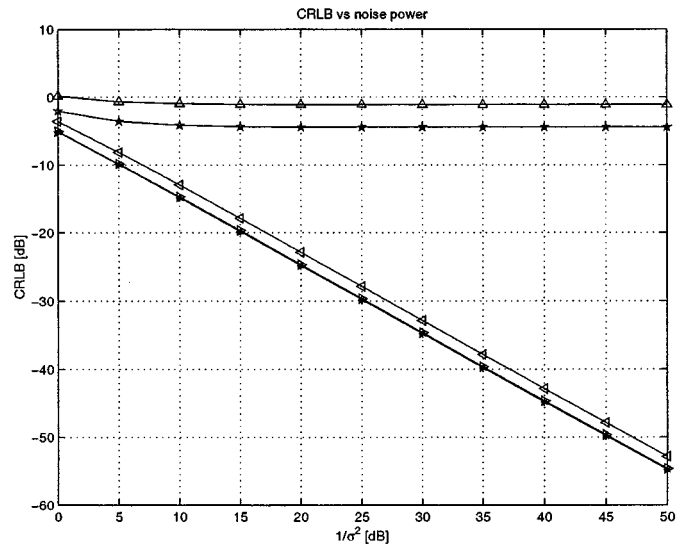
Fig. 10. $m = 4$, $N = 4$, and $n = 4$.Fig. 11. $m = 3$, $N = 4$, and $n = 5$.

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 $B = 32$. The channel parameters have been chosen randomly. The uniform horizontal power allocation scheme was compared with the uniform power allocation scheme given by (14), which has the same total amount of power allocated to training. 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.

B. Application of Convexity Theorems

We compare numerically some horizontal schemes with $N = m = n = 4$, $B = 32$, and $P_{\text{training}} = 2.5$. A comparison be-

Fig. 12. Comparison of different placements. For the schemes depicted in Fig. 9, the plots are, from top to bottom, (a) \triangle , (b) $*$, (g) \triangleleft ; (c)–(f) indistinguishable.

tween these schemes has already been done in Section IV-C; the schemes have been represented in Fig. 9 from the same subsection. Note that only one period of $N = 4$ blocks is represented in the figure. 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 will refer to an allocation scheme by its index in Fig. 9.

It can be easily observed from Fig. 12 that the difference between the plots (a) and (b) and the others is really substantial. This means that modifying the power allocation scheme under the conditions of Theorem 6 leads to significant changes in the CRLB. This is due to the fact that the data symbols act as noise in the case of superimposed training schemes. 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)–(e) are indistinguishable. It appears that the improvement in the SNR obtained by making the two training-only symbols supraunitary is compensated almost perfectly by the loss of information from the third symbol. 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. One would have expected this since the information about the channel embedded in the data symbols is much less than the information carried by the training symbols. It is interesting to see that the fact observed for the training symbols (that the power should be concentrated) holds for the data part as well. The information derived from the covariance matrix of the received signal is optimal if we concentrate the power in the data symbols.

C. Comparison Between Semiblind and Training-Based-Only Estimation Techniques

In this section, we compare the performance of semiblind and training-based only estimation schemes. However, such a comparison 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

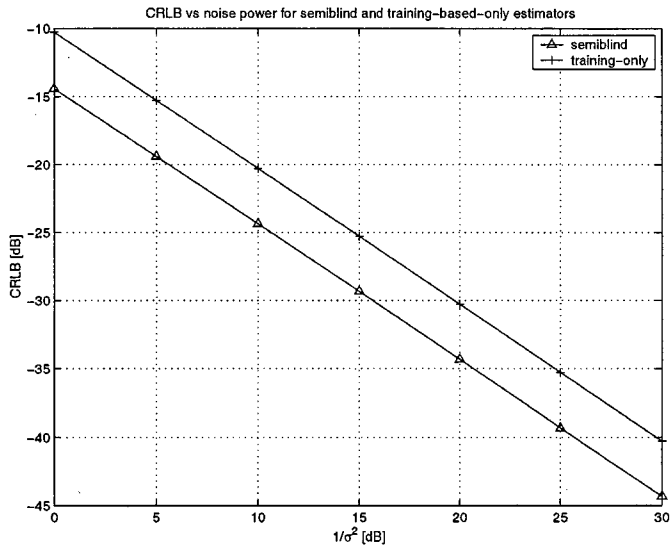


Fig. 13. CRLB for semiblind and training-based-only estimators.

($N = m = n = 4$, $B = 32$ 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 if a semiblind algorithm is used, 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 4 dB 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 cannot be applied efficiently if the power allocation scheme used is not uniform vertical with $\gamma = 0$. More exactly, the performance of a training-based-only algorithm for a horizontal scheme with full training symbols cannot exceed the one of an algorithm using the corresponding vertical scheme..

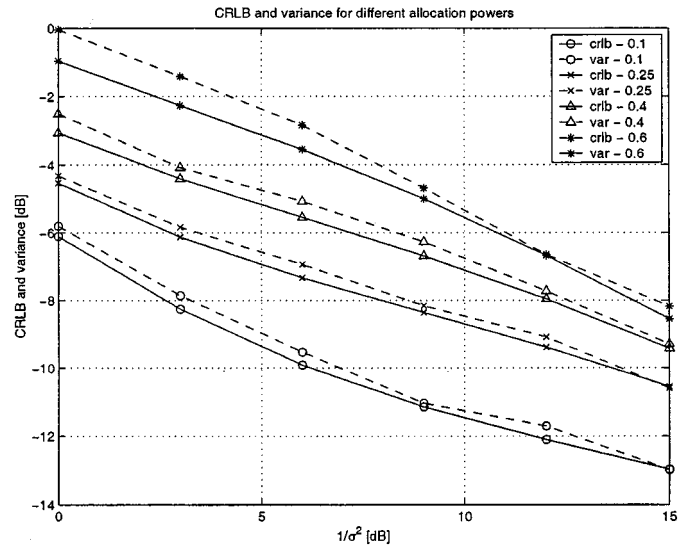
D. ML Algorithm

The channel estimation algorithm used was a semiblind maximum likelihood algorithm using the scoring method. The numerical results that are presented were obtained using the following setup. The training symbols were binary, $\{\pm 1\}$, the parameter values were $N = 4$, $m = 3$, $n = 5$, and the number of blocks considered $B = 32$. The channel coefficients were chosen randomly. We performed 500 Monte Carlo simulations. 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 behavior of the performance of the channel estimation algorithms when the power allocation scheme is modified.

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 esti-


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

mation techniques with superimposed data-pilot symbols. We have derived a closed-form expression for the FIM and further investigated the behavior 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. One may question the practical validity of the horizontal placement. By allowing continuous transmission of data and pilot symbols, the horizontal placement may offer better tracking capability for time-varying channels and more robustness to bursty interference. In addition, we have characterized the behavior 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 [25], or Adireddy *et al.* [8], may be considered.

APPENDIX

Proof of Theorem 1

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

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

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

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

because $q\gamma_{kt} + \sigma^2$ and σ^2 are the eigenvalues of $\mathbf{C}(\mathbf{a})$, where σ^2 has 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. \square

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 $\{\mathbf{X}_i, i = 1, \dots, N\}$ to be a family of $N \times N$ matrices satisfying conditions (2) and $\mathbf{X}_1 = \mathbf{I}_N$. Construct the family $\mathbf{Z}_i \triangleq \mathbf{X}_i \mathbf{X}_k^T$, where k is fixed, and $i \in \{1, \dots, N\}$. Then

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

Proof: If $k = 1$, then the statement is straightforward.

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

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

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

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

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

$$\mathbf{X}_3 \mathbf{X}_2^T \neq \pm \mathbf{X}_1; \quad \mathbf{X}_3 \mathbf{X}_2^T \neq \pm \mathbf{X}_2; \quad \mathbf{X}_3 \mathbf{X}_2^T \neq \pm \mathbf{X}_3. \quad (39)$$

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 will 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 nonzero entry. The position of the nonzero entries of \mathbf{Z} are determined by the other three matrices from the orthogonality conditions (2). We assumed $\mathbf{X}_1 = \mathbf{I}$; therefore, the diagonal entries of the other three matrices are all 0. In addition, 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

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

This corresponds to a nonzero entry of $\mathbf{Z} \mathbf{X}_2^T$; therefore, $k_1 \neq k_2$.

Each row vector $\mathbf{Z}(k_1, :)$ and $\mathbf{Z}(k_2, :)$ contains only one nonzero 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 $\mathbf{X}_3 \mathbf{X}_2^T = \pm \mathbf{X}_4$. \square

Remark: In Tarokh's paper [20], the connection between a family of 4×4 STBC matrices and the quaternionic algebra is mentioned; however, the statement of the lemma does not follow immediately.

Lemma 2: Consider $N = 4$ and $\{\mathbf{X}_i, i = 1, \dots, N\}$ to be a family of $m \times N$, $m < N$ matrices satisfying conditions (2). As in [21, form. 2.7.13, p. 77], denote by $\{\mathbf{Z}_i, i = 1, \dots, N\}$ the generating family of matrices. Define the following family of matrices:

$$\mathbf{G}_k \triangleq \begin{bmatrix} \mathbf{Z}_1(k, :) \\ \vdots \\ \mathbf{Z}_N(k, :) \end{bmatrix}, \quad k = 1, \dots, N.$$

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

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

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

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

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

Lemma 3: Consider the same family of matrices as in Lemma 2. Then, for any choice of four different indexes 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 right-hand side matrix are not possible. The statement follows. \square

Proof of Theorem 2

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

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

$$\mathbf{f}_{ik_1}^T \mathbf{f}_{jk_2} = -\mathbf{f}_{ik_2}^T \mathbf{f}_{jk_1} \text{ if } k_1 \neq k_2, \quad \mathbf{f}_{ik_1}^T \mathbf{f}_{ik_2} = \delta_{k_1 k_2}$$

$$\mathbf{W}^T \mathbf{F}_i = \mathbf{H}_i + a_i \mathbf{I}, \quad \mathbf{F}_i^T \mathbf{W} = -\mathbf{H}_i + a_i \mathbf{I}$$

$$\mathbf{H}_i^T = -\mathbf{H}_i.$$

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

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

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

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

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

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.

Here is a sketch of the steps done in order to obtain the expression of the elements of the FIM. $T_1(t)$ is quite straightforward. For $T_2(t)$, we need to substitute the covariance matrix and its derivative and then write down all 16 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 (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:

$$T_2(t) = T_{21}(t) + T_{22}(t) + T_{23}(t) \quad (46)$$

$$\begin{aligned} T_{21}(t) &= a_i a_j (2q^2 \text{tr}(\mathbf{D}_t^2 \mathbf{G}_t^2) + 3\rho q \text{tr}(\mathbf{D}_t \mathbf{G}_t^2) \\ &\quad + \rho^2 \text{tr}(\mathbf{G}_t^2)) \end{aligned} \quad (47)$$

$$T_{22}(t) = \delta_{ij} (\rho q^2 \text{tr}(\mathbf{D}_t \mathbf{G}_t^2) + \rho^2 q \text{tr}(\mathbf{G}_t^2)) \quad (48)$$

$$\begin{aligned} T_{23}(t) &= q^2 \text{tr}(\mathbf{D}_t \mathbf{H}_i \mathbf{G}_t \mathbf{D}_t \mathbf{H}_j \mathbf{G}_t) - q^2 \text{tr}(\mathbf{D}_t \mathbf{H}_i \mathbf{G}_t^2 \mathbf{D}_t \mathbf{H}_j) \\ &\quad + 2\rho q \text{tr}(\mathbf{D}_t \mathbf{H}_i \mathbf{G}_t \mathbf{H}_j \mathbf{G}_t) - \rho q \text{tr}(\mathbf{D}_t \mathbf{H}_i \mathbf{G}_t^2 \mathbf{H}_j) \\ &\quad + \rho^2 \text{tr}(\mathbf{G}_t \mathbf{H}_i \mathbf{G}_t \mathbf{H}_j). \end{aligned} \quad (49)$$

The complete proof can be found in the technical report [24].

Proof of Theorem 3

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

We will exploit the structure of the matrices $\{\mathbf{H}_i\}$ by defining the vectors $\{\mathbf{h}_i\}$ and the matrices $\{\mathbf{B}_i\}$ by the following relation (\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):

$$\mathbf{H}_i = \begin{bmatrix} 0, & \mathbf{h}_i^T \end{bmatrix}^T, \quad \begin{bmatrix} -\mathbf{h}_i, & \mathbf{B}_i^T \end{bmatrix}^T. \quad (50)$$

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

$$\mathbf{C}_k \triangleq \text{diag}(c_{0k}, c_k, \dots, c_k), \quad k \in \{1, 2\} \quad (51)$$

we have

$$\text{tr}(\mathbf{C}_1 \mathbf{H}_i \mathbf{C}_2 \mathbf{H}_j) = -(c_{01} c_2 + c_{02} c_1) \mathbf{h}_i^T \mathbf{h}_j - c_1 c_2 \text{tr}(\mathbf{B}_i \mathbf{B}_j). \quad (52)$$

□

Proposition 2: We have

$$[\mathbf{H}_i \mathbf{H}_j]_{k_1 k_2} = -[\mathbf{H}_i \mathbf{H}_j]_{k_2, k_1} \quad k_1 \neq k_2. \quad (53)$$

Proof: Taking into account that $[\mathbf{H}_i]_{kk} = 0$, we can write

$$[\mathbf{H}_i \mathbf{H}_j]_{k_1 k_2} = \sum_{k_3 \neq k_1, k_2} [\mathbf{H}_i]_{k_1 k_3} [\mathbf{H}_j]_{k_3 k_2}. \quad (54)$$

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

$$[\mathbf{H}_i \mathbf{H}_j]_{k_1 k_2} = [\mathbf{H}_i]_{k_1 k_3} [\mathbf{H}_j]_{k_3 k_2} + [\mathbf{H}_i]_{k_1 k_4} [\mathbf{H}_j]_{k_4 k_2} \quad (55)$$

$$[\mathbf{H}_j \mathbf{H}_i]_{k_1 k_2} = [\mathbf{H}_j]_{k_1 k_3} [\mathbf{H}_i]_{k_3 k_2} + [\mathbf{H}_j]_{k_1 k_4} [\mathbf{H}_i]_{k_4 k_2}. \quad (56)$$

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

$$[\mathbf{H}_i]_{k_1 k_4} = \pm [\mathbf{H}_i]_{k_3 k_2}, \quad [\mathbf{H}_i]_{k_4 k_2} = \mp [\mathbf{H}_i]_{k_1 k_3}.$$

The statement follows. □

Proposition 3: We have

$$\text{tr}(\mathbf{B}_i \mathbf{B}_j) = -(N-2) \mathbf{h}_i^T \mathbf{h}_j. \quad (57)$$

Proof: For $N = 2$, this is clear. For $N = 4$, express the left-hand side 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, which were previously defined for the general case, become

$$\mathbf{G}_t = \text{diag}(\gamma, 1, \dots, 1); \quad \mathbf{D}_t = \text{diag}(\Delta_\gamma, \Delta_1, \dots, \Delta_1)$$

$$\mathbf{P}_t = \text{diag}(\sqrt{\phi}, 0, \dots, 0).$$

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 [24] as

$$[\text{FIM}]_{ij} \triangleq T_{11}(t) + T_{12}(t) + T_{21}(t) + T_{22}(t) + T_{23}(t) \quad (58)$$

$$T_{12}(t) = \rho \text{tr}(\mathbf{P}_t^2) \delta_{ij} \quad (59)$$

$$\begin{aligned} T_{21}(t) &= a_i a_j (2q^2 \text{tr}(\mathbf{D}_t^2 \mathbf{G}_t^2) + 3\rho q \text{tr}(\mathbf{D}_t \mathbf{G}_t^2) \\ &\quad + \rho^2 \text{tr}(\mathbf{G}_t^2)) \end{aligned} \quad (60)$$

$$T_{22}(t) = \delta_{ij} (\rho q^2 \text{tr}(\mathbf{D}_t \mathbf{G}_t^2) + \rho^2 q \text{tr}(\mathbf{G}_t^2)) \quad (61)$$

$$\begin{aligned} T_{11}(t) + T_{23}(t) &= \phi \Delta_\gamma a_i a_j + (\phi \Delta_1 + \gamma \Delta_1 + (N-2) \Delta_1 \\ &\quad + \Delta_\gamma) \mathbf{h}_i^T \mathbf{h}_j. \end{aligned} \quad (62)$$

The formula in the theorem follows from the following observations.

We have, by definition

$$\mathbf{h}_i = \begin{bmatrix} \mathbf{w}_2^T \mathbf{J}_1 \mathbf{e}_i \\ \vdots \\ \mathbf{w}_N^T \mathbf{J}_1 \mathbf{e}_i \end{bmatrix} \quad (63)$$

which implies

$$\mathbf{h}_i^T \mathbf{h}_j = \mathbf{e}_i^T \mathbf{J}_1^T \sum_{k=2}^N \mathbf{w}_k \mathbf{w}_k^T \mathbf{J}_1 \mathbf{e}_j. \quad (64)$$

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

$$a_i a_j = [\mathbf{J}_1^T \mathbf{w}_1 \mathbf{w}_1^T \mathbf{J}_1]_{ij}. \quad (65)$$

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

$$\mathbf{Y}(t) = \mathbf{A} \sum_{i=1}^N \mathbf{X}_i s_i(t) + \mathbf{N}(t). \quad (66)$$

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

We saw that we can assume w.l.o.g. 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

$$\begin{aligned} \mathbf{Y}^{(k)}(t) \mathbf{X}_k^T &= \mathbf{A} s_0(t) + \mathbf{A} \sum_{i=1, i \neq k}^N \mathbf{X}_i \mathbf{X}_k^T u_i(t) + \mathbf{N}(t) \\ &= \mathbf{A} \mathbf{X}_1 s_0(t) + \mathbf{A} \sum_{i=1, i \neq k}^N \mathbf{X}_i \mathbf{X}_k^T u_i(t) + \mathbf{N}(t). \end{aligned} \quad (67)$$

From Lemma 1, the family $\{\mathbf{Z}_i = \mathbf{X}_i \mathbf{X}_k^T, i \neq k\}$ is the same as $\{\mathbf{X}_i, i \in \{2, \dots, N\}\}$, 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. In addition, 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 $\mathbf{X}_i, i \in \{1, \dots, N\}$ 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. We obtain

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

The details are provided in the technical report [24].

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 will 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 will separate the terms that are linear in γ (we will denote the coefficients with θ_i) and calculate the second derivative of the nonlinear part.

Introduce the following notation:

$$\xi \triangleq \frac{q}{\sigma^2}. \quad (68)$$

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

$$\begin{aligned} \Delta_\gamma &= -\frac{\gamma}{\sigma^2(q\gamma + \sigma^2)} = -\rho^2 \frac{\gamma}{\xi\gamma + 1} \\ \Delta_1 &= \Delta_\gamma|_{\gamma=1} = -\rho^2 \frac{1}{\xi + 1}. \end{aligned} \quad (69)$$

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

$$\begin{aligned} g_0 &= \theta_1 \gamma + \theta_0 + \rho^2 q \frac{\gamma^2}{\xi\gamma + 1} \\ \frac{d^2 g_0}{d\gamma^2} &= \rho^2 q \frac{2}{(\xi\gamma + 1)^3}. \end{aligned}$$

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

$$\begin{aligned} g_1 &= \rho^2 \left(2\xi^2 \frac{\gamma^4}{(\xi\gamma + 1)^2} - 3\xi \frac{\gamma^3}{\xi\gamma + 1} + \gamma^2 - \frac{(1 - \gamma)\gamma}{\xi\gamma + 1} \right) \\ &\quad + \theta_0 \gamma + \theta_1 \\ \frac{d^2 g_1}{d\gamma^2} &= 2 \frac{\gamma(\xi^2 - 4\xi) + 2 + \xi}{(\xi\gamma + 1)^4}. \end{aligned}$$

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

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

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

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

$$\begin{aligned} g_2 &= (N - 1)\Delta_1 + \Delta_\gamma \\ \frac{d^2 g_2}{d\gamma^2} &= \rho^2 \frac{2\xi}{(\xi\gamma + 1)^3}. \end{aligned}$$

g_2 is clearly convex with respect to 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

$$0 < \tilde{\mathbf{F}}(\gamma - \zeta) + \tilde{\mathbf{F}}(\gamma + \zeta) - 2\tilde{\mathbf{F}}(\gamma). \quad (70)$$

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 [26]). The formula in the theorem follows from (70).

The proof of the theorem is complete. Intermediate steps of the calculations are provided in [24].

Proof of Theorem 7

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

$$\phi_1 \triangleq [1 + \phi, 0, \dots, 0]^T, \quad \gamma_1 \triangleq [0, 1, \dots, 1]^T \quad (71)$$

$$\phi_2 \triangleq [1 - \zeta - \phi, 0, \dots, 0]^T, \quad \gamma_2 \triangleq [\gamma, 1, \dots, 1]^T. \quad (72)$$

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 do not have a symmetry in ϕ ; therefore, we need to analyze the functions $g_k = g_k(1) + g_k(2)$ that are the eigenvalues of the FIM for two blocks

$$g_0 = g_0(1) + g_0(2) = \rho(1 + \phi + 1 - \phi - \zeta) + \theta \quad (73)$$

$$\frac{d g_0}{d \phi} = 0. \quad (74)$$

This term is constant in ϕ

$$\begin{aligned} g_1 &= g_1(1) + g_1(2) = (1 + \phi)\Delta_\gamma|_{\gamma=0} \\ &\quad + (1 - \phi - \zeta + \theta)\Delta_\gamma \\ &= (1 - \phi - \zeta + \theta)\Delta_\gamma \end{aligned} \quad (75)$$

$$\frac{d g_1}{d \phi} = -\Delta_\gamma. \quad (76)$$

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 $0 < \phi < 1$.

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

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

REFERENCES

- [1] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.
- [2] N. Seshadri, A. F. Naguib, V. Tarokh, and A. R. Calderbank, "A space-time coding modem for high-data-rate wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1459–1478, Oct. 1998.
- [3] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high capacity space-time wireless," in *Proc. 37th Annu. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Sept. 1999.
- [4] Z. Liu, G. B. Giannakis, S. Barbarossa, and A. Scaglione, "Transmit antennae space-time block coding for generalized OFDM in the presence of unknown multipath," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1352–1364, July 2001.
- [5] B. Muquet, S. Zhou, and G. B. Giannakis, "Subspace-based estimation of frequency-selective channels for space-time block precoded transmissions," in *Proc. Conf. Rec. Thirty-Fourth Asilomar Conf. Signals, Syst., Comput.*, vol. 2, 2000, pp. 975–979.
- [6] A. L. Swindlehurst, "Simultaneous channel estimation and decoding for diagonal space-time codes," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop*, 2000, pp. 173–177.
- [7] E. De Carvalho and D. T. M. Slock, "Cramer–Rao bounds for semi-blind, blind and training sequence based channel estimation," in *Proc. First IEEE Signal Process. Workshop Signal Process. Adv. Wireless Commun.*, 1997, pp. 129–132.
- [8] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for unknown channels," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2338–2353, Mar. 2001. [Online]. Available: <http://www.ee.cornell.edu/~ltong/pubJ.html>, submitted for publication.
- [9] S. Adireddy and L. Tong, "Optimal placement of known symbols for nonergodic broadcast channels," *IEEE Trans. Inform. Theory*, Aug. 2001. [Online]. Available: <http://www.ee.cornell.edu/~ltong/pubJ.html>, submitted for publication.
- [10] S. Ohno and G. B. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," *IEEE Trans. Commun.*, Nov. 2000, submitted for publication.
- [11] S. Adireddy and L. Tong, "Detection with embedded known symbols: Optimal symbol placement and equalization," in *Proc. Int. Conf. Acoust. Speech, Signal Process.*, vol. V, Istanbul, Turkey, June 2000, pp. 2541–2543.
- [12] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, pp. 2109–2112, May 2001.
- [13] —, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Processing*, Apr. 2001. [Online]. Available: <http://www.ee.cornell.edu/~ltong/pubJ.html>, to be published.
- [14] P. Spasojevic and C. N. Georghades, "Complementary sequences for ISI channel estimation," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1145–1152, Mar. 2001.
- [15] J. H. Manton, I. Y. Mareels, and Y. Hua, "Affine precoders for reliable communications," in *Proc. ICASSP*, vol. 5, Istanbul, Turkey, June 2000, pp. 2749–2752.
- [16] M. Guenach and L. Vandendorpe, "Performance analysis of joint EM/SAGE estimation and multistage detection in UTRA-WCDMA uplink," in *Proc. IEEE Int. Conf. Commun.*, vol. 2, 2000, pp. 638–640.
- [17] S. Ohno and G. B. Giannakis, "Superimposed training on redundant precoding for low-complexity recovery of block transmissions," in *Proc. IEEE International Conf. Commun.*, vol. 5, Helsinki, Finland, June 2001, pp. 1501–1505.

- [18] D. Makrakis and K. Feher, "A novel pilot insertion-extraction technique based on spread spectrum techniques," in *Proc. Miami Technicon*, Miami, FL, 1987, pp. 129-132.
- [19] P. Hoehner and F. Tufvesson, "Channel estimation with superimposed pilot sequence," in *GLOBECOM*, vol. 4, Rio de Janeiro, Brazil, Dec. 1999, pp. 2162-2166.
- [20] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456-1467, July 1999.
- [21] G. Ganesan and P. Stoica, "Space-time diversity," in *Signal Processing Advances in Wireless and Mobile Communications*, P. G. Giannakis, Y. Hua, P. Stoica, and L. Tong, Ed. Englewood Cliffs, NJ: Prentice-Hall, 2001.
- [22] I. E. Telatar, "Capacity of multi antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, Nov./Dec. 1999.
- [23] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [24] C. Budianu and L. Tong. (2001) Channel estimation for space-time block coding systems. Cornell Univ., Dept. Elect. Comput. Eng., Tech. Rep.. [Online]. Available: <http://www.ee.cornell.edu/~ltong/pubJ.html>.
- [25] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple antenna wireless links?," *IEEE Trans. Information Theory*, Aug. 2000, submitted for publication.
- [26] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.



Lang Tong (S'87-M'91) received the B.E. degree from Tsinghua University, Beijing, China, in 1985 and the M.S. and Ph.D. degrees in electrical engineering in 1987 and 1990, respectively, from the University of Notre Dame, Notre Dame, IN.

After being a Postdoctoral Research Associate with the Information Systems Laboratory, Stanford University, Stanford, CA, he joined the Department of Electrical and Computer Engineering, West Virginia University, Morgantown, and was also with the University of Connecticut, Storrs. Since the

Fall of 1998, he has been with the School of Electrical Engineering, Cornell University, Ithaca, NY, where he is an Associate Professor. He also held a Visiting Assistant Professor position at Stanford University in the Summer of 1992. His research interests include statistical signal processing, wireless communications, and system theory.

Dr. Tong received the Young Investigator Award from the Office of Naval Research in 1996 and the Outstanding Young Author Award from the IEEE Circuits and Systems Society.



Cristian Budianu (S'97) was born in Bucharest, Romania, in 1976. He received the B.S. degree (with highest honors), in electronics and telecommunications from the University "Politehnica" of Bucharest in 1999. He has been pursuing the Ph.D. degree in electrical engineering at Cornell University, Ithaca, NY, since August 1999.

His interests lie in the areas of statistical signal processing, communications, information theory, and applied probability.