# OPTIMAL TRAINING DESIGNS FOR BAYESIAN CHANNEL ESTIMATORS WITH APPLICATION IN CDMA SYSTEMS

*Azadeh Vosoughi and Anna Scaglione*

Cornell University, Ithaca, NY 14853
av78@cornell.edu, anna@ece.cornell.edu

**ABSTRACT**

In this paper, we address the problem of data transmission over a block fading frequency selective multi-input multi-output (MIMO) channel. The encoded data symbols are passed through an affine precoder and sent over multiple transmit antennas. Affine precoding is a unifying model of the existing training schemes, namely, preamble based training, pilot symbol assisted modulation (PSAM), and superimposed training. Within this general model, we show that the channel Cramer-Rao bound (CRB) is minimized if the precoded data and training symbols satisfy a special form of orthogonality. Under the power constraint on the training and the precoded symbols, we provide optimal and suboptimal training design guidelines.

## 1. INTRODUCTION

One of the greatest difficulties which system designers face with in wireless communication is estimating the fading channel. The fading channel coefficients can be obtained by transmitting training symbols. Considering the transmission power and bandwidth constraints, the challenge is to design training that introduces minimum overhead, in terms of dimensions and power, and still provides an acceptable level of performance, in the sense of estimation error and transmission rate. Three major classes of training designs are $(i)$ preamble based training, in which a training sequence is included at the beginning of data burst; $(ii)$ PSAM technique in which training symbols are inserted in the data stream and are separated from the data symbols either in frequency or in time (so called time division multiplexed (TDM) training); $(iii)$ superimposed training, in which the training sequence is added to the data sequence. The current wireless communication systems exploit any of these schemes to acquire fading channel. For instance, Orthogonal Frequency Division Multiplexing (OFDM) scheme, adopted in the IEEE 802.11a standard, utilizes frequency multiplexed training. In GSM, which uses TDMA scheme, the training scheme is TDM. In IMT2000 standard, which uses wideband CDMA (W-CDMA) scheme, channel estimation in downlink is facilitated with common pilots shared by all users, which are code multiplexed[1], and dedicated pilot symbols devoted to each individual user, which are TDM.

The advantage of preamble based and PSAM training schemes is that they reduce the receiver complexity by decoupling the symbol detection and channel estimation. However, it has not been proven whether the best performance with limited resources can be achieved only if the data and the training symbols are spanned on non-overlapping subspaces. If the training has to use dimensions that are not occupied by data signals, this means that the time-bandwidth product (i.e., the product of the transmission bandwidth and the duration of the transmission block) has to be increased accordingly to accommodate sufficient dimensions for both training and data. Hence, the mere observation that training and data overlap in time does not necessarily lead to the conclusion that the system achieves greater bandwidth efficiency, unless the signal dimensions are shared among both training and data.

A number of results have been obtained in the context of optimal preamble training, PSAM, and superimposed training designs, where the optimization has relied on either information theoretic bounds [1], or estimation theoretic bounds [2, 3]. Despite the vast literature on this topic, there are still not unique guidelines for training design. This work is an attempt to illustrate how the optimality of certain designs is tied to the constraints imposed on the training design. As we have discussed in [4], to optimally design the training one has to specify the receiver architecture. In this paper we assume that upon observing the received signal, the receiver forms the minimum mean squared error (MMSE) estimate of the channel. Although the MMSE channel estimator always exist, in general it is difficult to obtain it in a closed form. Moreover, the calculation of its MSE is cumbersome. Hence, it is useful to establish lower bounds on the attainable MSE to which the performance of the optimal channel estimator or any suboptimal channel estimator can be com-

This work is supported in part by the National Science Foundation under grant CCR-0133635.

[1] Code multiplexed training can be viewed as superimposed training, since training and data are separated neither in time nor in frequency.

pared. The Bayesian channel CRB, defined in [5], provides a lower bound on the MSE of any Bayesian channel estimator. Choosing the channel CRB as the optimality criterion and adopting affine precoding as the transmission strategy, we investigate the best training design for data transmission over a block fading frequency selective MIMO channel. Note that the affine precoding scheme introduced in [6] is a general framework in which preamble based training, PSAM schemes or superimposed training can be treated as special cases. We show that a lower channel CRB is obtained when the precoded data and the training satisfy a special form of orthogonality. Under the power constraint on the training and the precoded symbols, we provide training design guidelines.

**Notation**: Boldface upper and lower cases denote matrices and column vectors, respectively. Complex conjugate, Hermitian, transpose and expectation operations are represented by $(.)^H, (.)^T$ and $E\{.\}$, respectively. The $(i,j)$ entry of matrix $\boldsymbol{A}$ is indicated with $\boldsymbol{A}_{ij}$. $tr(\boldsymbol{A})$ and $\mathcal{N}(\boldsymbol{A})$ are trace and the null space of $\boldsymbol{A}$. The column vector formed by stacking vertically the columns of $\boldsymbol{A}$ is $\boldsymbol{a} = vec(\boldsymbol{A})$. The probability density function (pdf) is presented as $p(.)$. $\otimes$ is the Kronecker product. For any real valued number $x$, $(x)^+ = \max(x,0)$. $\mathcal{P}$ denotes the total average transmit power per block, whereas $\mathcal{P}_t$ and $\mathcal{P}_s$ indicate the portion dedicated to the training and data symbols, respectively.

## 2. SYSTEM MODEL

The system model we consider here is a single input multiple antenna channel. The input bit stream is encoded and modulated to generate a symbol stream. The coded symbols are passed through a linear affine precoder and transmitted over the MIMO fading channel with $K$ transmit and $R$ receive antennas. Using the observation vector, corresponding to the discrete time complex equivalent model for the received signal at all receive antennas, the receiver obtains the MMSE estimate of the channel, which will be utilized later to reconstruct the transmitted data symbols. Assuming that the antennas are deployed sufficiently apart, the $KR$ established links between the transmit and receive antennas are uncorrelated. We assume that the overall impulse response of the channel between the $k$-th transmitter and the $r$-th receiver has finite length $L$ and that the channel coefficients are independent and identically distributed (i.i.d) with pdf $\sim \mathcal{CN}(0, \sigma_{hh}^2)$. The fading coefficients are constant for $P := M + L$ symbol periods and change to independent values over the next transmission block.

Let $\boldsymbol{x}[n]$ denote the discrete time $K \times 1$ transmitted signal vector and $\boldsymbol{z}[n]$ represent the discrete time $R \times 1$ received signal vector. The received signal is the transmitted signal faded by the random channel and corrupted by the additive noise. We write $\boldsymbol{z}[n] = \sum_{l=0}^{L} \mathbf{H}[l]\boldsymbol{x}[n-l] + \boldsymbol{n}[n]$ where the $(r,k)$ element of the $R \times K$ matrix $\mathbf{H}[l]$ characterizes the $l$-th tap of the discrete time overall channel between the $k$-th transmit and the $r$-th receive antennas. We model the noise samples $\boldsymbol{n}[n]$ as i.i.d with pdf $\sim \mathcal{CN}(0, \sigma_{nn}^2)$. Stacking $P$ transmit snapshots in a $PK \times 1$ vector $\boldsymbol{x}_i := vec([\boldsymbol{x}[iP], \ldots, \boldsymbol{x}[iP+P-1]])$ and $M$ received snapshots in an $MR \times 1$ vector $\boldsymbol{z}_i := vec([\boldsymbol{z}[iP+L], \ldots, \boldsymbol{z}[iP+P-1]])$, where we eliminated the first $L$ vectors at the receiver to cancel the inter-block interference (IBI), we obtain $\boldsymbol{z}_i = \boldsymbol{H}\boldsymbol{x}_i + \boldsymbol{n}_i$ where $\boldsymbol{H}$ is an $RM \times KP$ block sylvester matrix:

$$\boldsymbol{H} = \begin{bmatrix} \mathbf{H}[L] & \cdots & \mathbf{H}[0] & 0 & \ldots & 0 \\ 0 & \mathbf{H}[L] & \cdots & \mathbf{H}[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{H}[L] & \cdots & \mathbf{H}[0] \end{bmatrix}$$

The information sequence is parsed into blocks of size $N$, $\boldsymbol{s}_i$, where $N \leq \text{rank}(\boldsymbol{H}) \leq \min(RM, KP)$. We assume $\boldsymbol{s}_i$ is Gaussian with the covariance $\boldsymbol{R}_{ss} = \sigma_{ss}^2 \boldsymbol{I}_N$ and $\boldsymbol{n}_i$ and $\boldsymbol{s}_i$ are uncorrelated. Each $N \times 1$ block $\boldsymbol{s}_i$ is precoded by a tall $KP \times N$ precoding matrix $\boldsymbol{F}$ with $N \leq KP$. A $KP \times 1$ training vector $\boldsymbol{t}$, which is known to the receiver, is added to the precoding block $\boldsymbol{F}\boldsymbol{s}_i$ to obtain the transmitted data block $\boldsymbol{x}_i = \boldsymbol{F}\boldsymbol{s}_i + \boldsymbol{t}$. For the mapping from $\boldsymbol{s}_i$ to $\boldsymbol{x}_i$ to be invertible, we require $\boldsymbol{F}$ to be full column rank. We constraint the total average transmit power $\mathcal{P} = E\{||\boldsymbol{x}_i||^2\}$ such that:

$$\mathcal{P} = E\{||\boldsymbol{x}_i||^2\} = \sigma_{ss}^2 tr(\boldsymbol{F}\boldsymbol{F}^H) + ||\boldsymbol{t}||^2 = \mathcal{P}_s + \mathcal{P}_t \quad (1)$$

Combining all, we obtain:

$$\boldsymbol{z}_i = \boldsymbol{H}\boldsymbol{F}\boldsymbol{s}_i + \boldsymbol{H}\boldsymbol{t} + \boldsymbol{n}_i \quad (2)$$

Without IBI, we assume that the resulting channel estimator operates on a block-by-block basis and we omit the block index $i$. Finally, let $\boldsymbol{h} \in \mathcal{C}^{KR(L+1)}$ be the complex vector containing the channel parameters to be estimated $\boldsymbol{h} := vec\left([\ \mathbf{H}[0]\ \ldots\ \mathbf{H}[L]\ ]^T\right)$. To simplify the further derivation we rewrite $\boldsymbol{H}\boldsymbol{x}$ explicitly as a function of $\boldsymbol{h}$. To this end we introduce a mapping $\boldsymbol{\Phi} : \boldsymbol{x} \rightarrow \mathbf{X}$ such that $\boldsymbol{H}\boldsymbol{x} = \mathbf{X}\boldsymbol{h}$ where $\mathbf{X} := \boldsymbol{\Phi}(\boldsymbol{x})$ is defined as:

$$\mathbf{X} := \begin{bmatrix} \boldsymbol{I}_R \otimes \boldsymbol{\mathcal{X}}_{(1,:)} \\ \boldsymbol{I}_R \otimes \boldsymbol{\mathcal{X}}_{(2,:)} \\ \vdots \\ \boldsymbol{I}_R \otimes \boldsymbol{\mathcal{X}}_{(M,:)} \end{bmatrix} \quad (3)$$

$$\boldsymbol{\mathcal{X}} := \begin{bmatrix} \boldsymbol{x}^T[L] & \cdots & \boldsymbol{x}^T[0] \\ \boldsymbol{x}^T[L+1] & \cdots & \boldsymbol{x}^T[1] \\ \vdots & \ddots & \vdots \\ \boldsymbol{x}^T[P-1] & \cdots & \boldsymbol{x}^T[M-1] \end{bmatrix} \quad (4)$$

in which $\boldsymbol{\mathcal{X}}_{(i,:)}$ is the $i$-th row of the $M \times K(L+1)$ Toeplitz structure matrix $\boldsymbol{\mathcal{X}}$.

## 3. LOWER BOUND ON THE MSE OF BAYESIAN CHANNEL ESTIMATORS

Lemma 1 provides the expression for the complex Fisher information matrix (FIM) (proofs are omitted due to lack of space [4]).

**Lemma 1** *Under certain regularity conditions [5] which are satisfied by Gaussian random vectors [2], the FIM $\mathcal{J}$ is well defined [5] and is given by [4]:*

$$\mathcal{J} = E_h\{\mathcal{J}_c\} + \sigma_{hh}^{-2}\boldsymbol{I}_{KR(L+1)} \tag{5}$$

*in which [c.f.(3)-(4)]:*

$$\begin{aligned}
\mathcal{J}_c &= \mathbf{T}^H\boldsymbol{R}_z^{-1}\mathbf{T} + \boldsymbol{\Sigma} \\
\mathbf{T} &= \boldsymbol{\Phi}(\boldsymbol{t}) \\
\boldsymbol{R}_z &= \sigma_{ss}^2\boldsymbol{H}\boldsymbol{F}\boldsymbol{F}^H\boldsymbol{H}^H + \sigma_{nn}^2\boldsymbol{I}_{RM} \\
\boldsymbol{\Sigma} &= \sigma_{ss}^4\boldsymbol{\mathcal{E}}^H(\boldsymbol{D}^T \otimes \boldsymbol{R}_z^{-1})\boldsymbol{\mathcal{E}} \\
\boldsymbol{D} &= \boldsymbol{F}\boldsymbol{F}^H\boldsymbol{H}^H\boldsymbol{R}_z^{-1}\boldsymbol{H}\boldsymbol{F}\boldsymbol{F}^H \\
\boldsymbol{\mathcal{E}} &= \left[vec(\boldsymbol{E}_1), \ldots, vec(\boldsymbol{E}_{(L+1)RK})\right] \\
\boldsymbol{E}_i &= \frac{\partial \boldsymbol{H}}{\partial h_i} \quad i = 1, \ldots, (L+1)RK
\end{aligned}$$

## 4. TRAINING DESIGN GUIDELINES: CRB CRITERION

Motivated by the existing literature on preamble based training and PSAM optimization, which enforce the orthogonality between $\boldsymbol{F}$ and $\boldsymbol{t}$, we attempt to answer the following question: Given a precoder $\boldsymbol{F}$ satisfying $\sigma_{ss}^2 tr(\boldsymbol{F}\boldsymbol{F}^H) \leq \mathcal{P}_s$, and considering the channel CRB as the optimality criterion, what is the best superimposed training vector $\boldsymbol{t}$ with the constraint $||\boldsymbol{t}||^2 \leq \mathcal{P}_t$? The answer to this question is provided in lemmas 2 and 3. We assume that $\boldsymbol{F}$ is the solution of a different optimization problem. For instance, $\boldsymbol{F}$ achieves the best rate tradeoff of channel coding versus linear precoding for a given symbol error probability and $\mathcal{P}_s$. We believe that this issue deserves separate investigation. In the following lemma we answer the question raised above:

**Lemma 2** *Given a precoder $\boldsymbol{F}$ satisfying $\sigma_{ss}^2 tr(\boldsymbol{F}\boldsymbol{F}^H) \leq \mathcal{P}_s$ we define $\boldsymbol{\mathcal{V}}$ as the matrix of coefficients of an orthonormal basis that spans the intersection of null spaces of all $\mathbf{F}_i^H$, i.e., $\bigcap_{i=1}^N \mathcal{N}(\mathbf{F}_i^H)$, where $\mathbf{F}_i = \boldsymbol{\Phi}(\boldsymbol{f}_i)\ \forall\ i$. Assuming that $\boldsymbol{\mathcal{V}}$ is non-empty, the training matrices $\mathbf{T} = \boldsymbol{\Phi}(\boldsymbol{t})$ satisfying $tr(\mathbf{T}^H\mathbf{T}) = \overline{\mathcal{P}}_t$ [2] which minimize $tr(\mathcal{J}^{-1})$ admit the following orthogonality condition [4]:*

$$\mathbf{T}^H\mathbf{F}_i = \mathbf{0} \quad i = 1, 2, \ldots, N \tag{6}$$

*Under the orthogonality condition $\mathcal{J}$ reduces to:*

$$\mathcal{J} = \sigma_{nn}^{-2}\mathbf{T}^H\mathbf{T} + E_h\{\boldsymbol{\Sigma}\} + \sigma_{hh}^{-2}\boldsymbol{I}_{RK(L+1)} \tag{7}$$

[2]For $M \gg L$ $\overline{\mathcal{P}}_t \simeq MRK(L+1)\mathcal{P}_t$.

*Remark*: Lemma 2 states that PSAM scheme is the optimum transmission strategy. The constraint in (6) is analogous to the affine precoding design constraint in [7]. To reduce the complexity of the receiver structure, the authors in [7] enforce the orthogonality constraint to be able to convert the nonlinear estimation problem in model (2) to two low complexity, albeit suboptimal, linear estimation problems. In their model, the receiver obtains a channel estimate that is based on the training only. Interestingly, as $\mathcal{J}$ expression in (7) indicates, the efficient channel estimator incorporates the statistics of the data symbols as well as the training to measure the channel [3].

As prescribed in proof of lemma 2 [4] the training matrix $\mathbf{T}$ satisfying (6) is in the form of $\mathbf{T} = \boldsymbol{\mathcal{V}}\boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is an arbitrary matrix satisfying $tr(\boldsymbol{\Theta}^H\boldsymbol{\Theta}) = \overline{\mathcal{P}}_t$. In [4] we described the design for $\boldsymbol{t}$ such that $\mathbf{T} = \boldsymbol{\Phi}(\boldsymbol{t})$ is compatible with lemma 3 for the class of affine precoders that incorporates cyclic prefix (CP). We showed that within this class, the columns of the precoder and training must be loaded on non-overlapping subcarriers in order to satisfy the orthogonality constraint in (6). It is of interest to see if there exists a matrix (or a class of matrices) $\mathbf{T}$ that minimizes (7), when the search space is not restricted only to the matrices of the form $\boldsymbol{\Phi}(\boldsymbol{t})$. Lemma 4 describes the matrix $\mathbf{T}$ which admits (6) and achieves the minimum of the channel CRB, i.e., the minimum of $tr(\mathcal{J}^{-1})$ where $\mathcal{J}$ is given in (7).

**Lemma 3** *Let $E_h\{\boldsymbol{\Sigma}\} + \sigma_{hh}^{-2}\boldsymbol{I} := \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H$ where $\boldsymbol{U}$ is a unitary matrix and $\boldsymbol{\Lambda}$ is a diagonal matrix with non-decreasing diagonal entries. The solution of the optimization problem [4]*

$$\mathbf{T}_{opt} = \arg\min tr(\mathcal{J}^{-1}), \;\; tr(\mathbf{T}^H\mathbf{T}) = \overline{\mathcal{P}}_t$$

*is $\mathbf{T} = \boldsymbol{\mathcal{V}}\boldsymbol{\Delta}\boldsymbol{U}^H$ where $\boldsymbol{\Delta}$ is an $n \times RK(L+1)$ matrix with zero entries, except for:*

$$|\boldsymbol{\Delta}_{ii}|^2 = \left(\frac{\overline{\mathcal{P}}_t + \sigma_{nn}^2\sum_{j=1}^{\overline{n}}\boldsymbol{\Lambda}_{jj}}{\overline{n}} - \sigma_{nn}^2\boldsymbol{\Lambda}_{ii}\right)^+ \quad i = 1, \ldots, \overline{n}$$

*in which $\overline{n} := rank(\mathbf{T}^H\mathbf{T}) \leq min(n, RK(L+1))$ and $n$ is the number of columns of $\boldsymbol{\mathcal{V}}$.*

Lemma 3 is in general incompatible with the fact that $\mathbf{T}_{opt}$ can be written as $\boldsymbol{\Phi}(\boldsymbol{t})$ for a vector $\boldsymbol{t}$. A suboptimal $\boldsymbol{t}$ could be obtained, for example, by minimizing the error between $\mathbf{T}_{opt}$ provided by lemma 3 and $\boldsymbol{\Phi}(\boldsymbol{t})$, i.e., $||\mathbf{T}_{opt} - \boldsymbol{\Phi}(\boldsymbol{t})||^2$, under the constraint that $||\boldsymbol{t}||^2 = \mathcal{P}_t$. The quadratic cost-function will result in a linear gradient that will be send to zero if $\boldsymbol{t}$ belongs to the null space of a structured matrix constructed with the element of $\mathbf{T}_{opt}$. If the

[3]An efficient Bayesian channel estimator exists if and only if $p(\boldsymbol{h}|\boldsymbol{z})$ is Gaussian [5], which does not hold in our set up. However, we may argue that (7) indicates we should exploit the statistics of data symbols as well as training to achieve a performance close to the CRB.

null space is empty the suboptimal $\boldsymbol{t}$ can be given by the minimum singular vector of the same matrix.

## 5. ORTHOGONALITY CONSTRAINT IN W-CDMA

The results presented in lemmas 2 and 3 can be utilized to provide guidelines for training design in downlink of a W-CDMA system. Let $\boldsymbol{z}_i$ denote the vector received by the $i$-th user:

$$\boldsymbol{z}_i = \boldsymbol{H}_i\boldsymbol{A}\boldsymbol{s} + \boldsymbol{H}_i\boldsymbol{B}\boldsymbol{t} + \boldsymbol{n}_i$$

where $\boldsymbol{A}$ is $KP \times N_s$ and $\boldsymbol{B}$ is $KP \times N_t$ and $N_s + N_t \leq KP$. The matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are the code matrices associated with the data and the training symbols, respectively and they satisfy $\boldsymbol{A}^H\boldsymbol{A} = \boldsymbol{I}_{N_s}$, $\boldsymbol{B}^H\boldsymbol{B} = \boldsymbol{I}_{N_t}$, and $\boldsymbol{A}^H\boldsymbol{B} = \boldsymbol{0}_{N_s \times N_t}$. Let $\boldsymbol{a}_i$ be the $i$-th column of the code matrix $\boldsymbol{A}$ and define $\mathbf{A}_i := \boldsymbol{\Phi}(\boldsymbol{a}_i)$ and $\overline{\boldsymbol{B}} := \boldsymbol{\Phi}(\boldsymbol{B}\boldsymbol{t})$ [c.f.(3)]. The orthogonality constraint requires:

$$\mathbf{A}_i^H\overline{\boldsymbol{B}} = \mathbf{0} \;\; i = 1, \cdots, N_s \qquad (8)$$

Following an analogous approach to the one suggested in Section 4 we choose $\boldsymbol{t}$ such that under the constraint $||\boldsymbol{t}||^2 = \mathcal{P}_t$ the error $||\mathbf{T}_{opt} - \boldsymbol{\Phi}(\boldsymbol{B}\boldsymbol{t})||^2$ is minimized, where $\mathbf{T}_{opt}$ is obtained using lemma 3 for a given $\boldsymbol{A}$.

## 6. MMSE CHANNEL ESTIMATOR

We may view the received vector $\boldsymbol{z}$ in (2) as $\boldsymbol{z} = \mathbf{T}\boldsymbol{h} + \boldsymbol{w}$ where $\boldsymbol{w} := \boldsymbol{H}\boldsymbol{F}\boldsymbol{s} + \boldsymbol{n}$ is a zero mean non-Gaussian noise. We treat the symbols as the nuisance parameters and we find the MMSE channel estimate $\hat{\boldsymbol{h}} = E\{\boldsymbol{h}|\boldsymbol{z}\} = E\{E\{\boldsymbol{h}|\boldsymbol{z}, \boldsymbol{s}\}\}$. Given $\boldsymbol{s}$, $\boldsymbol{w}$ is a zero mean Gaussian noise with covariance $\boldsymbol{R}_w = \overline{\boldsymbol{F}}(\boldsymbol{s}\boldsymbol{s}^H \otimes \sigma_{hh}^2\boldsymbol{I})\overline{\boldsymbol{F}}^H + \sigma_{nn}^2\boldsymbol{I}_{RM}$ where $\overline{\boldsymbol{F}} := [\mathbf{F}_1 \; \mathbf{F}_2 \; \ldots \; \mathbf{F}_N]$ and $\mathbf{F}_i = \boldsymbol{\Phi}(\boldsymbol{f}_i)$. Now we write [8]:

$$E\{\boldsymbol{h}|\boldsymbol{z}, \boldsymbol{s}\} = \sigma_{hh}^2\mathbf{T}^H(\sigma_{hh}^2\mathbf{T}\mathbf{T}^H + \boldsymbol{R}_w)^{-1}\boldsymbol{z}$$

which immediately follows:

$$\hat{\boldsymbol{h}} = \sigma_{hh}^2\mathbf{T}^H E_s\left\{(\sigma_{hh}^2\mathbf{T}\mathbf{T}^H + \boldsymbol{R}_w)^{-1}\right\}\boldsymbol{z}$$

In Section 7, we compare numerically the performance of this estimator against the CRB.

## 7. NUMERICAL RESULTS

In our simulation we set $K = 2$, $R = 2$, $L = 3$, $\sigma_{hh}^2 = 1/(L+1)$, and $\sigma_{ss}^2 = 1$. The simulation results are averaged over 100 sets of independent Rayleigh fading channels. Without loss of generality, we assume that $\mathcal{P} = 1$ and therefore SNR is defined as $SNR := -10log_{10}\sigma_{nn}^2$. We define $\zeta = ||\boldsymbol{t}||^2/\mathcal{P},\; 0 \leq \zeta \leq 1$ as the normalized power

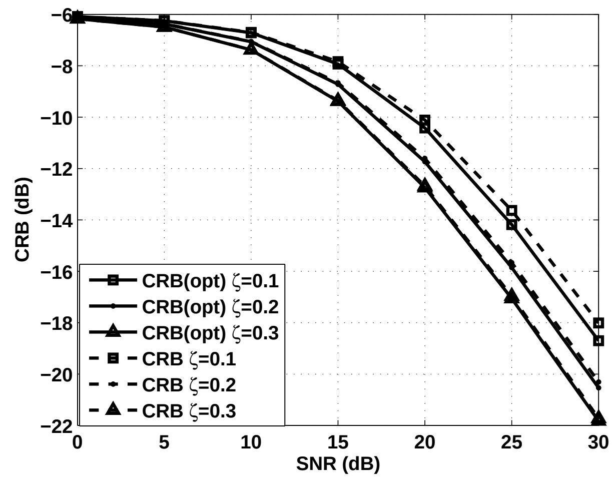


**Fig. 1**. channel CRB: comparison between the semi-unitary training design and the design in lemma 4.

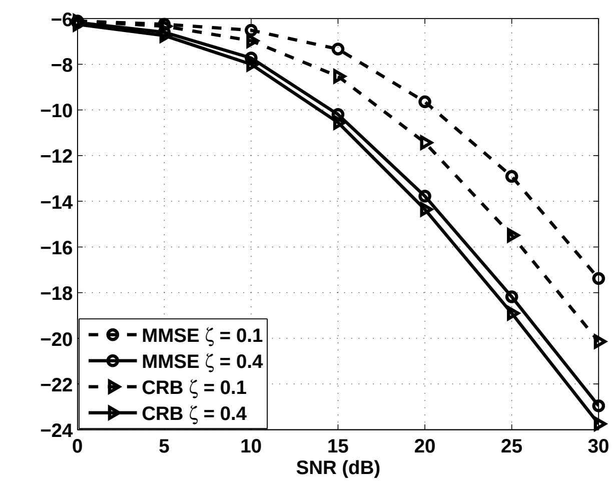


**Fig. 2**. comparison between the MSE of MMSE channel estimate and channel CRB

allocated to training. At each SNR level, $\boldsymbol{t}$ and $\boldsymbol{F}$ are scaled such that the power constraint in (1) is satisfied. The CRB values shown in the plots are normalized by the number of diagonal elements of the CRB matrix.

We consider two pairs of affine precoders, namely $(\boldsymbol{F}_i, \boldsymbol{t}_i)$ for $i = 1, 2$. The pairs belong to the class of affine precoders incorporating CP and the columns of the precoder and the training are loaded on non-overlapping tones. The selected precoders $\boldsymbol{F}_i$ $i = 1, 2$ provide us adequate dimension so that the training matrix $\mathbf{T}_i = \boldsymbol{\Phi}(\boldsymbol{t}_i)$ can be potentially designed as a full column rank matrix. The vector $\boldsymbol{t}_i$ is designed such that $\mathbf{T}_i^H\mathbf{T}_i = \overline{\mathcal{P}}_t\boldsymbol{I}$ and each pilot tone is transmitted only by one of the two transmitters. The columns of $\boldsymbol{F}_1$ are loaded on non-overlapping tones and each tone is transmitted by one antenna. Whereas the columns of $\boldsymbol{F}_2$ share tones and each tone is sent by both transmitters. For $(\boldsymbol{F}_1, \boldsymbol{t}_1)$ we let $M = 16$, $N = 8$ and the set of ordered indices of pilot tones $\mathcal{I}_0 = \{2i, i = 0, \ldots, 7\}$. For $(\boldsymbol{F}_2, \boldsymbol{t}_2)$ we let $M = 24$, $N = 4$ and $\mathcal{I}_0 = \{3i, i = 0, \ldots, 7\}$.

Figure 1 compares the channel CRB given the precoder $\boldsymbol{F}_1$. The curves labelled as CRB(opt) and CRB are plotted using the CRB optimal training matrix provided by lemma 4 and $\mathbf{T}_1 = \mathbf{\Phi}(\boldsymbol{t}_1)$, respectively. We observe that CRB(opt) is slightly better than CRB, while the gap between the two decreases as $\zeta$ increases. As we mentioned, the optimal training matrix in lemma 3 may not be in the form of $\mathbf{T} = \mathbf{\Phi}(\boldsymbol{t})$ for a vector $\boldsymbol{t}$. The gain of CRB(opt) over CRB would be less if we constraint the CRB optimal training matrix to have a particular structure. So, in general, lemma 3 allows to assess if the quest for the optimum constrained training design is worthwhile. It can be shown that under the uncorrelated uniform scattering model we adopted and for MIMO-OFDM type precoder $\boldsymbol{F}$, $E_h\{\mathbf{\Sigma}\}$ is approximately proportional to an identity matrix and hence the CRB optimal training is semi-unitary[4].

Figure 2 compares the performance of the MMSE channel estimator against the channel CRB, when we choose fix the precoder as $\mathbf{F}_2$ and we use the unconstrained CRB optimal training matrix given in lemma 3. The performance of the MMSE estimate becomes reasonably close to the CRB as $\zeta$ increases.

In a nutshell, we showed that a necessary condition to achieve the minimum of the channel CRB is that the training and the precoded data symbols are spanned on a non-overlapping space, in the sense defined in lemma 2. We also derived the optimal and a suboptimal training design in the sense of minimizing the channel CRB. The design guidelines we proposed can be utilized in the downlink of the W-CDMA systems for channel estimation.

## 8. REFERENCES


[1] B. Hassibi and B. Hochwald, " How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, Vol.49, No.4 , pp. 951 -963, April 2003.

[2] M. Dong and L. Tong, " Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Sig. Proc.*, Vol. 50, Issue 12, pp. 3055-3069, December 2002.

[3] Cristian Budianu and Lang Tong, " Channel estimation for space-time orthogonal block codes," *IEEE Trans. Sig. Proc.*, Vol.50, No.10, pp. 2515-2528, October 2002.

[4] A. Vosoughi and A. Scaglione " Everything you always wanted to know about training: guidelines derived using the affine precoding framework, " accepted for publication in *IEEE Trans. Sig. Proc.*

[5] H. L. Van Trees, *Detection Estimation and modulation Theory*, part 1, New York Wiley 1968.

[6] J. H. Manton, I. Y. Mareels and Y.Hua, " Affine precoders for reliable communications," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Sig. Proc.*, Vol.5, pp. 2741-2744, Istanbul, Turkey, June 2000.

[7] S.Ohno and G. B. Giannakis, " Optimal training and redundant precoding for block transmissions with application to wirlesss OFDM," *IEEE Trans. Comm.*, Vol.50, No.12, pp. 2113-2123, December 2002.

[8] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*, vol. 1, Princton Hall 1993.


[4] We assume that $\boldsymbol{F}$ provides adequate dimension so that $\mathbf{T}$ can be designed as a full column rank matrix.