# THE BEST TRAINING DEPENDS ON THE RECEIVER ARCHITECTURE

*Azadeh Vosoughi and Anna Scaglione*

School of Electrical and Computer Engineering
Cornell University, Ithaca, NY 14853
av78@cornell.edu; anna@ece.cornell.edu

## ABSTRACT

We consider a block fading frequency selective multi-input multi-output (MIMO) channel in additive white Gaussian noise (AWGN). The channel input is a training vector superimposed on a linearly precoded vector of Gaussian symbols. This form of precoding is referred to as *affine precoding*. We derive the Cramer-Rao bound (CRB) under two circumstances: the random parameter vector to be estimated contains $(i)$ only fading channel coefficients, $(ii)$ unknown data symbols as well as the channel coefficients. While case $(i)$ corresponds to the decoding schemes in which the channel is estimated first and the channel measurement is utilized to recover the data symbols, case $(ii)$ corresponds to methods in which channel and symbol estimation is performed jointly. The interesting outcome of our investigation is that minimizing trace of the channel CRB for cases $(i)$ and $(ii)$ under a total transmit power constraint leads to different affine precoder design guidelines.

## 1. INTRODUCTION

One of the greatest challenges in wireless communications is channel estimation. To acquire the channel state information (CSI) without ambiguity training is required. Two major classes of training design are $(i)$ preamble based training, in which a training sequence is included at the beginning of data burst; $(ii)$ Pilot Symbol Assisted Modulation (PSAM) technique [1] in which training symbols are inserted in the data stream and are separated from the information symbols either in frequency (pilot tones) or in time (time division multiplexed training). While the preamble based and PSAM reduce the receiver complexity by decoupling the symbol detection and channel estimation, their optimality has not been established.

The potential benefits of *superimposed training* technique, modelling training as a known sequence added to the unknown data sequence at the transmitter, is utilized in [2] for the purpose of channel estimation. By construction, there is no increase in the bandwidth. Affine precoding scheme [3] enjoys the benefits of both linear precoding [4] and training sequence. It can be viewed as a general framework in which PSAM with pilot tones or superimposed training sequences can be treated as special cases.

Although the literature devoted to optimal preamble and PSAM design is extensive, there are still not unique guidelines for the design, due to the modelling assumptions used that make the results often difficult to generalize or compare. In this work we utilize affine precoding as the transmission strategy and CRB as the optimality criterion and we study the relationship between the optimal affine precoders and the assumptions made on the receiver structure (e.g., decoupled vs. joint channel and symbol estimation) and the constraints on the training design (e.g., preamble, PSAM or superimposed training).

**Notation**: Boldface upper and lower cases denote matrices and column vectors respectively. The $tr(\boldsymbol{A})$ is trace 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 is presented as $p(.)$. $diag(\boldsymbol{a})$ is a diagonal matrix whose diagonal elements are the components of $\boldsymbol{a}$. Complex conjugate, Hermitian, transpose, pseudoinverse and expectation operations are represented by $(.)^*$, $(.)^H$, $(.)^T$, $(.)^\dagger$, $E\{.\}$ respectively.

## 2. SYSTEM MODEL

The system considered has $K$ transmit and $R$ receive antennas. We assume a block fading model where $P$ is the coherence time of the channel, i.e., the equivalent discrete-time impulse response of the channel does not change during the transmission of $P$ snapshots. In our setup The channel has finite memory $L$. The information sequence $\boldsymbol{s}[n]$ is parsed into blocks of size $N$, namely $\boldsymbol{s}_i$. Each block is precoded by a tall $KP \times N$ precoding matrix $\boldsymbol{F}$. An $KP \times 1$ 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 $\boldsymbol{x}_i = \boldsymbol{F}\boldsymbol{s}_i + \boldsymbol{t}$. The $PK \times 1$ vector $\boldsymbol{x}_i$ is obtained by stacking $P$ transmit snapshots $\boldsymbol{x}_i := vec([\boldsymbol{x}[iP], \ldots, \boldsymbol{x}[iP + P - 1]])$, where $\boldsymbol{x}[iP + p]$ is the $K \times 1$ (coded) symbol vector, emitted by the $K$ transmit antennas. For the mapping from $\boldsymbol{s}_i$ to $\boldsymbol{x}_i$ to be invertible, we require $\boldsymbol{F}$ to be full column rank. Stacking $M = P - L$ received snapshots in an $MR \times 1$ vector $\boldsymbol{y}_i := vec([\boldsymbol{y}[iP + L], \ldots, \boldsymbol{y}[iP + P - 1]])$, in which we eliminated the first $L$ vectors to cancel the inter-block interference (IBI), we obtain $\boldsymbol{y}_i = \boldsymbol{H}\boldsymbol{x}_i$, where $\boldsymbol{H}$ is an $RM \times KP$ block Toeplitz matrix:

$$\boldsymbol{H} = \begin{bmatrix} \mathbf{H}[L] & \cdots & \mathbf{H}[0] & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}[L] & \cdots & \mathbf{H}[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{H}[L] & \cdots & \mathbf{H}[0] \end{bmatrix} \tag{1}$$

$\{\mathbf{H}[l]\}_{r,k}$ is the $l$-th sample of the impulse response characterizing the channel between the $k$-th transmitter and the $r$-th receiver. We require $N \le \text{rank}(\boldsymbol{H}) \le \min(KP, RM)$. The received signal $\boldsymbol{z}_i = \boldsymbol{y}_i + \boldsymbol{n}_i$ where $\boldsymbol{n}_i \sim \mathcal{N}(\mathbf{0}, \sigma_{nn}^2 \boldsymbol{I}_{RM})$. Furthermore, we assume that $\boldsymbol{s}_i \sim \mathcal{N}(\mathbf{0}, \sigma_{ss}^2 \boldsymbol{I}_N)$. and $\boldsymbol{n}_i$ and $\boldsymbol{s}_i$ are uncorrelated. Combining all we obtain:

$$\boldsymbol{z}_i = \boldsymbol{H}\boldsymbol{F}\boldsymbol{s}_i + \boldsymbol{H}\boldsymbol{t} + \boldsymbol{n}_i \tag{2}$$

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

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

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

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

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

## 3. DECOUPLED SYMBOL AND CHANNEL ESTIMATION

### 3.1. CRB Expression

Let $\boldsymbol{h}$ be the complex random parameter vector we wish to estimate. The CRB is the inverse of the complex Fisher information matrix (FIM) $\boldsymbol{\mathcal{J}}$:

$$\boldsymbol{\mathcal{J}} := E\left\{\frac{\partial \ln p(\boldsymbol{z}, \boldsymbol{h})}{\partial \boldsymbol{h}^*}\left(\frac{\partial \ln p(\boldsymbol{z}, \boldsymbol{h})}{\partial \boldsymbol{h}^*}\right)^H\right\} \tag{5}$$

where the expectation is taken over $p(\boldsymbol{z}, \boldsymbol{h})$. (proofs are omitted due to lack of space [5]).

**Lemma 1** *$\boldsymbol{\mathcal{J}}$ is given by $\boldsymbol{\mathcal{J}} = E\{\boldsymbol{\mathcal{J}}_c\} + \sigma_{hh}^{-2}\boldsymbol{I}$ where $\boldsymbol{\mathcal{J}}_c$ is the FIM for deterministic channel estimation:*

$$\boldsymbol{\mathcal{J}}_c = \mathbf{T}^H \boldsymbol{R}_z^{-1}\mathbf{T} + \sigma_{ss}^4 \boldsymbol{\mathcal{E}}^H(\boldsymbol{D}^T \otimes \boldsymbol{R}_z^{-1})\boldsymbol{\mathcal{E}} \tag{6}$$

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

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

Interestingly, $\boldsymbol{\mathcal{J}}_c$ is decomposed into two terms, where the first term $\mathbf{T}^H\boldsymbol{R}_z^{-1}\mathbf{T}$ depends on both $\boldsymbol{F}$ and $\boldsymbol{t}$ through $\boldsymbol{R}_z^{-1}$ and $\mathbf{T}$ respectively, whereas the second term $\boldsymbol{\mathcal{E}}^H(\boldsymbol{D}^T \otimes \boldsymbol{R}_z^{-1})\boldsymbol{\mathcal{E}}$ depends on $\boldsymbol{F}$ but not on $\boldsymbol{t}$.

### 3.2. Affine Precoder Design Guidlines: CRB Criterion

We look for pairs $(\boldsymbol{F}, \boldsymbol{t})$ which provide us a lower $tr(\boldsymbol{\mathcal{J}}^{-1})$ than the others, independent of the underlying FIR MIMO channel. Our interesting observation is that $\boldsymbol{\mathcal{J}}_c$ has higher eigenvalues if $\boldsymbol{F}$ and $\boldsymbol{t}$ satisfy a form of *orthogonality*, in the sense stated in the lemma 2. Since this constraint is independent of the channel, it also increases the eigenvalues of $\boldsymbol{\mathcal{J}}_c$ averaged over the channel and consequently the eigenvalues for $\boldsymbol{\mathcal{J}}$.

**Lemma 2** *The pair $(\boldsymbol{F}, \boldsymbol{t})$ which satisfy the following orthogonality constraint provides a lower $tr(\boldsymbol{\mathcal{J}}^{-1})$:*

$$\boldsymbol{\mathcal{T}}^H\boldsymbol{\mathcal{F}}_i = \mathbf{0} \quad i = 1, 2, \ldots, N \tag{7}$$

*where $\boldsymbol{\mathcal{T}}$ and $\boldsymbol{\mathcal{F}}_i$ are defined according to (3)-(4) using $\boldsymbol{t}$ and $\boldsymbol{f}_i$ (the $i$-th column of $\boldsymbol{F}$).*

The orthogonality constraint in (7) is analogous to the affine precoding design constraints in [6]. The authors found that the class of $(\boldsymbol{F}, \boldsymbol{t})$ which admit a form of orthogonality constraint similar to (7) converts the nonlinear estimation problem in model (2) to two low complexity, albeit suboptimal, linear estimation problems. Although, we did not enforce symbol detection and channel estimation to be decoupled *a priori*, using $tr(\text{CRB})$ as the optimality design criteria led us to a similar orthogonality constraint.

We wish to characterize the pairs $(\boldsymbol{F}, \boldsymbol{t})$ which satisfy (7). When $M \gg L$ for any pair $(\boldsymbol{F}, \boldsymbol{t})$ we can introduce $(\boldsymbol{F}_{CP}, \boldsymbol{t}_{CP})$ (where $\boldsymbol{F}_{CP}$ and $\boldsymbol{t}_{CP}$ have the cyclic prefix (CP)) such that $||\boldsymbol{F}_{CP} - \boldsymbol{F}|| \ll 1$ and $||\boldsymbol{t}_{CP} - \boldsymbol{t}|| \ll 1$. Hence, we select our general design such that the matrix $\boldsymbol{F}$ and the vector $\boldsymbol{t}$ incorporate CP. To this end, we define $\boldsymbol{F} := \boldsymbol{\Psi}\overline{\boldsymbol{F}}, \ \ \boldsymbol{t} := \boldsymbol{\Psi}\bar{\boldsymbol{t}}$ where $\boldsymbol{\Psi} := [\mathbf{0}_{KL \times K(M-L)} \ \ \boldsymbol{I}_{KL}; \boldsymbol{I}_{KM}]$ is the $KP \times KM$ CP-inducing matrix. Under the CP assumption, $\boldsymbol{\mathcal{T}}$ and $\boldsymbol{\mathcal{F}}_i$s will be block circulant matrices and can be diagonalized using the fast Fourier transform (FFT) matrix.

### 3.3. Design of Affine Precoding With CP

Let $W := \exp(j2\pi/M)$ and $\boldsymbol{U}$ be the $M \times M$ FFT matrix with $[\boldsymbol{U}]_{mn} = M^{-1/2}W^{-(m-1)(n-1)}$. We can express $\boldsymbol{\mathcal{T}}$ as $\boldsymbol{\mathcal{T}} = \boldsymbol{U}^H\boldsymbol{\Delta}_{\bar{t}}(\boldsymbol{U}_{0:L} \otimes I_K)$, where $\boldsymbol{U}_{0:L}$ denotes the first $L+1$ columns of $\boldsymbol{U}$, $\boldsymbol{\Delta}_{\bar{t}}$ is an $M \times MK$ block diagonal matrix defined as $\boldsymbol{\Delta}_{\bar{t}} := diag[\tilde{\boldsymbol{t}}^T(1), \tilde{\boldsymbol{t}}^T(W), \cdots, \tilde{\boldsymbol{t}}^T(W^{M-1})]$ and $\tilde{\boldsymbol{t}}(z) := \sum_{m=0}^{M-1}\bar{\boldsymbol{t}}[m]z^{-m}$ is the $\mathcal{Z}$ transform of $\bar{\boldsymbol{t}}$. Similarly, we can establish the FFT-based diagonalization of $\boldsymbol{\mathcal{F}}_i$ and define $\tilde{\boldsymbol{f}}_i(z) := \sum_{m=0}^{M-1}\overline{\boldsymbol{f}}_i[m]z^{-m}$.

**Lemma 3** *A sufficient condition for (7) to be satisfied is that $(\overline{\boldsymbol{F}}, \bar{\boldsymbol{t}})$ are selected from the family:*

$$\bar{\boldsymbol{t}} = \frac{1}{\sqrt{M}}(\boldsymbol{U}^H \otimes \boldsymbol{I}_K)\tilde{\boldsymbol{t}} \quad \overline{\boldsymbol{F}} = \frac{1}{\sqrt{M}}(\boldsymbol{U}^H \otimes \boldsymbol{I}_K)\tilde{\boldsymbol{\mathcal{F}}} \tag{8}$$

*where:*

$$\tilde{\boldsymbol{t}} := \begin{bmatrix} \tilde{\boldsymbol{t}}(1) \\ \vdots \\ \tilde{\boldsymbol{t}}(W^{M-1}) \end{bmatrix} \ \tilde{\boldsymbol{\mathcal{F}}} := \begin{bmatrix} \tilde{\boldsymbol{f}}_1(1) & \ldots & \tilde{\boldsymbol{f}}_N(1) \\ \vdots & \ddots & \vdots \\ \tilde{\boldsymbol{f}}_1(W^{M-1}) & \ldots & \tilde{\boldsymbol{f}}_N(W^{M-1}) \end{bmatrix} \tag{9}$$

*where $\tilde{\boldsymbol{\mathcal{F}}}$ is a full column rank matrix and $\tilde{\boldsymbol{t}}(W^m)\tilde{\boldsymbol{f}}_i^H(W^m) = \mathbf{0}$ for $m = 0, \ 1, \ \cdots, \ M-1$ and $i = 1, \ldots, N$.*

## 4. JOINT CHANNEL AND SYMBOL ESTIMATION

### 4.1. CRB Expression

Let $\boldsymbol{\theta} := [\boldsymbol{s}^H \;\; \boldsymbol{h}^H]^H$ be the $(N + (L+1)RK) \times 1$ complex random parameter vector we desire to estimate. The FIM $\overline{\boldsymbol{\mathcal{J}}}$ is defined as:

$$\overline{\boldsymbol{\mathcal{J}}} := E\left\{ \frac{\partial \ln p(\boldsymbol{z}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \left( \frac{\partial \ln p(\boldsymbol{z}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \right)^H \right\}$$

$$:= \begin{bmatrix} \overline{\boldsymbol{\mathcal{J}}}_{1,1} & \overline{\boldsymbol{\mathcal{J}}}_{1,2} \\ \overline{\boldsymbol{\mathcal{J}}}_{1,2}^H & \overline{\boldsymbol{\mathcal{J}}}_{2,2} \end{bmatrix} \quad (10)$$

where the expectation is taken over $p(\boldsymbol{z}, \boldsymbol{\theta})$.

**Lemma 4** *$\overline{\boldsymbol{\mathcal{J}}}$ is given by its block partitions:*

$$\overline{\boldsymbol{\mathcal{J}}}_{1,1} = \left(\frac{\sigma_{hh}}{\sigma_{nn}}\right)^2 \begin{bmatrix} tr(\mathbf{F}_1^H \mathbf{F}_1) & \cdots & tr(\mathbf{F}_1^H \mathbf{F}_N) \\ \vdots & \ddots & \vdots \\ tr(\mathbf{F}_N^H \mathbf{F}_1) & \cdots & tr(\mathbf{F}_N^H \mathbf{F}_N) \end{bmatrix} + \sigma_{ss}^{-2} \boldsymbol{I}_N$$

$$\overline{\boldsymbol{\mathcal{J}}}_{2,2} = \sigma_{nn}^{-2} \left( \sigma_{ss}^2 \sum_{i=1}^{N} \mathbf{F}_i^H \mathbf{F}_i + \mathbf{T}^H \mathbf{T} \right) + \sigma_{hh}^{-2} \boldsymbol{I}_{RR(L+1)}$$

$$\overline{\boldsymbol{\mathcal{J}}}_{1,2} = \mathbf{0} \quad (11)$$

*where [c.f.(3)-(4)] $\mathbf{T} := \boldsymbol{\Phi}(\boldsymbol{t})$ and $\mathbf{F}_i := \boldsymbol{\Phi}(\boldsymbol{f}_i)$, in which $\boldsymbol{f}_i$ is the $i$-th column of $\boldsymbol{F}$. The block inversion formula implies that the symbol and the channel CRBs are $\overline{\boldsymbol{\mathcal{J}}}_{1,1}^{-1}$ and $\overline{\boldsymbol{\mathcal{J}}}_{2,2}^{-1}$ respectively.*

The structure of $\overline{\boldsymbol{\mathcal{J}}}_{2,2}$ reveals the fact that the channel CRB does not distinguish the precoder $\boldsymbol{F}$ from the training $\boldsymbol{t}$. This is in contrast to what we observed in Section 3.2, where minimizing $tr(\boldsymbol{\mathcal{J}}^{-1})$ requires a special form of orthogonality between $\boldsymbol{F}$ and $\boldsymbol{t}$. In fact, let $\boldsymbol{G} := [\sigma_{ss}^2 \boldsymbol{F} \;\; \boldsymbol{t}]$ be the extended precoder whose $i$-th column is denoted as $\boldsymbol{g}_i$. We can rewrite $\overline{\boldsymbol{\mathcal{J}}}_{2,2}$ as:

$$\overline{\boldsymbol{\mathcal{J}}}_{2,2} = \frac{1}{\sigma_{nn}^2} \boldsymbol{I}_R \otimes \sum_{i=1}^{N+1} \boldsymbol{\mathcal{G}}_i^H \boldsymbol{\mathcal{G}}_i + \sigma_{hh}^{-2} \boldsymbol{I} \quad (12)$$

where $\boldsymbol{\mathcal{G}}_i$ is defined according to (3)-(4) and $\mathbf{G}_i := \boldsymbol{\Phi}(\boldsymbol{g}_i)$. On the other hand, the symbol CRB is independent of $\boldsymbol{t}$.

### 4.2. Affine Precoder Design Guidelines: CRB Criterion

We search for a family of affine precoders $(\boldsymbol{F}, \boldsymbol{t})$ which minimizes $tr(\overline{\boldsymbol{\mathcal{J}}}_{1,1}^{-1})$ and $tr(\overline{\boldsymbol{\mathcal{J}}}_{2,2}^{-1})$ [c.f. (11)] simultaneously. Schwartz's inequality states that for any positive definite matrix $\boldsymbol{A}$ we have $\boldsymbol{A}_{ii}^{-1} \geq \frac{1}{\boldsymbol{A}_{ii}}$ in which equality holds for diagonal $\boldsymbol{A}$. Here, $tr(\overline{\boldsymbol{\mathcal{J}}}_{2,2}^{-1})$ has the following lower bound:

$$tr(\overline{\boldsymbol{\mathcal{J}}}_{2,2}^{-1}) \geq \sum_{m=1}^{RK(L+1)} \frac{1}{\sigma_{hh}^{-2} + \sigma_{nn}^{-2} \left[ \boldsymbol{I}_R \otimes \sum_{i=1}^{N+1} \boldsymbol{\mathcal{G}}_i^H \boldsymbol{\mathcal{G}}_i \right]_{mm}} \quad (13)$$

where the equality holds if and only if $\sum_{i=1}^{N+1} \boldsymbol{\mathcal{G}}_i^H \boldsymbol{\mathcal{G}}_i = \alpha \boldsymbol{I}_{K(L+1)}$ for some nonzero constant $\alpha$. A simple design for $\boldsymbol{\mathcal{G}}_i$'s such that $\sum_{i=1}^{N+1} \boldsymbol{\mathcal{G}}_i^H \boldsymbol{\mathcal{G}}_i = \alpha \boldsymbol{I}_{K(L+1)}$ is $\boldsymbol{\mathcal{G}}_i^H \boldsymbol{\mathcal{G}}_i = |g_i|^2 \boldsymbol{I}_{K(L+1)}$ for some $|g_i|$ and $i = 1, \ldots, N+1$. This is feasible only and only if $\boldsymbol{\mathcal{G}}_i$'s are tall and full column rank. For $\boldsymbol{\mathcal{G}}_i$ to be tall, we require $M \geq K(L+1)$. We start with the following simple design:

$$\boldsymbol{g}_i = \boldsymbol{e}_j \otimes \overline{\boldsymbol{g}_i} \quad j = L+1, \; \cdots, \; M \quad i = 1, \ldots, N+1 \quad (14)$$

where $\boldsymbol{e}_j$ is an $P \times 1$ canonical vector and $\overline{\boldsymbol{g}_i}$ is an arbitrary $K \times 1$ vector. To have at least $N$ distinct canonical vector corresponding to the first $N$ linearly independent columns of $\boldsymbol{G}$ (i.e., the columns of $F$) we require $M - L \geq N$ (minimum redundancy corresponds to $M = L + N$). Whereas $\boldsymbol{g}_{N+1}$ (the training vector $\boldsymbol{t}$) can be identical to any other columns of the precoder $\boldsymbol{F}$. We assume that total transmitted power is limited,i.e., $\mathcal{P} = ||\boldsymbol{x}||^2 = \sigma_{ss}^2 tr(\boldsymbol{F}\boldsymbol{F}^H) + ||\boldsymbol{t}||^2$ which implies $\sum_{i=1}^{N+1} |g_i|^2 = \mathcal{P}$. Under (14) we obtain:

$$\sum_{i=1}^{N+1} \boldsymbol{\mathcal{G}}_i^H \boldsymbol{\mathcal{G}}_i = \boldsymbol{I}_{(L+1)} \otimes \sum_{i=1}^{N+1} \overline{\boldsymbol{g}_i}^* \overline{\boldsymbol{g}_i}^T \quad (15)$$

hence $\sum_{i=1}^{N+1} \boldsymbol{\mathcal{G}}_i^H \boldsymbol{\mathcal{G}}_i = \alpha \boldsymbol{I}_{K(L+1)}$ if and only if $\sum_{i=1}^{N+1} \overline{\boldsymbol{g}_i}^* \overline{\boldsymbol{g}_i}^T = \alpha \boldsymbol{I}_K$. For the case where $\kappa := (N+1)/K$ is an integer, a simple design can be accomplished by letting $\overline{\boldsymbol{g}_i}$ to be:

$$\overline{\boldsymbol{g}_i} = |g_i| \boldsymbol{e}_k \quad k = 1, \; \cdots, \; K \quad (16)$$

and divide the $N+1$ $\boldsymbol{g}_i$s vectors into $K$ groups, such that each group contains $\kappa$ vectors and the summation of $|g_i|^2$ for each group adds up to $\alpha = \mathcal{P}/K$.

**Lemma 5** *Under the design in (14) combined with the power constraint we obtain:*

$$tr(\overline{\boldsymbol{\mathcal{J}}}_{2,2}^{-1}) = \frac{R(L+1)}{\sigma_{hh}^2 + \sigma_{nn}^{-2} \mathcal{P}} \quad (17)$$

$$tr(\overline{\boldsymbol{\mathcal{J}}}_{1,1}^{-1}) = \frac{\sigma_{ss}^2 \sigma_{nn}^2}{\sigma_{hh}^2} \sum_{i=1}^{N} \frac{1}{\sigma_{ss}^2 R(L+1)|g_i|^2 + 1} \quad (18)$$

Interestingly, $tr(\overline{\boldsymbol{\mathcal{J}}}_{2,2}^{-1})$ does not depend on a specific power distribution between the symbols and the training, but on the total transmitted power $\mathcal{P}$. Whereas $tr(\overline{\boldsymbol{\mathcal{J}}}_{1,1}^{-1})$ is minimized if we distribute all the power only between the data symbols, without increasing $tr(\overline{\boldsymbol{\mathcal{J}}}_{2,2}^{-1})$. In other words, we can set $|g_{N+1}| = 0$ and modify $\kappa$ to $N/K$, while we maintain the minimum of $tr(\overline{\boldsymbol{\mathcal{J}}}_{2,2}^{-1})$. This observation indicates that not only there is no need of having specific orthogonality between symbols and training, but also training is absorbing the power that could have been used to estimate the symbols, without any improvement in channel estimate performance.

## 5. NUMERICAL RESULTS

We set $K = 2$, $R = 2$, $L = 3$, $\sigma_{ss}^2 = 1$ $\sigma_{hh}^2 = 1/(L+1)$. 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 the signal-to-noise ratio (SNR) is $SNR := -10log_{10}\sigma_{nn}^2$. We define $\zeta := ||\boldsymbol{t}||^2/\mathcal{P}$ as the training power fraction. For each SNR, $\boldsymbol{t}$ and $\boldsymbol{F}$ are scaled such that the power constraint is satisfied. We designed three pairs of $(\boldsymbol{F}, \boldsymbol{t})$ namely $(i)$,$(ii)$, and $(iii)$ which are the optimal designs for the first, the second, and both CRB expression respectively:

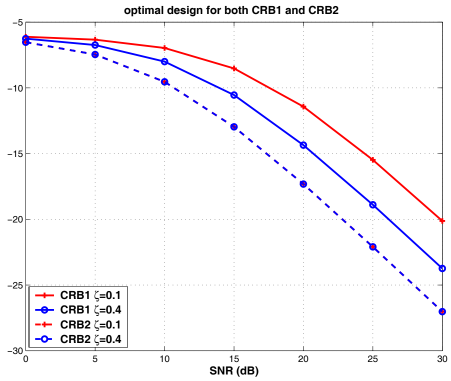


**Fig. 1**. $tr(CRB1)$ and $tr(CRB2)$ evaluated for design $(iii)$

- *design* $(i)$: We set $M = 16$, $N = 8$. To satisfy the orthogonality constraint in (7) we load $\boldsymbol{t}$ and $\boldsymbol{F}$ on non-overlapping subcarriers. In particular, we choose $\tilde{\boldsymbol{t}}(W^m) \neq \mathbf{0}$ for $m \in \{0, 2, 4, 6, 8, 10, 12, 14\}$, $\tilde{\boldsymbol{t}}(W^m) = \boldsymbol{e}_1$ for $m = 0, 4, 8, 12$ and $\tilde{\boldsymbol{t}}(W^m) = \boldsymbol{e}_2$ for $m = 2, 6, 10, 14$. $\tilde{\boldsymbol{\mathcal{F}}} = \mathbf{0}$ except for $\tilde{\boldsymbol{f}}_1(1) = \tilde{\boldsymbol{f}}_2(W^3) = \tilde{\boldsymbol{f}}_3(W^5) = \tilde{\boldsymbol{f}}_4(W^7) = \tilde{\boldsymbol{f}}_5(W^9) = \tilde{\boldsymbol{f}}_6(W^{11}) = \tilde{\boldsymbol{f}}_7(W^{13}) = \tilde{\boldsymbol{f}}_8(W^{15}) = \boldsymbol{e}_1$.
- *design* $(ii)$: We set $M = 10$, $N = 7$. $\boldsymbol{G} = [\sigma_{ss}^2 \boldsymbol{F} \ \ \boldsymbol{t}]$ is designed along the guidelines provided in (14) and (16), in which $|g_i| = \sqrt{1/8}$ for $i = 1, 2, 3, 4$, $|g_i| = \sqrt{(\zeta - 0.5)/3}$ $i = 5, 6, 7$, and $|g_8| = \sqrt{\zeta}$ where $\zeta < 0.5$.
- *design* $(iii)$: We set $M = 24$, $N = 4$. We choose $\tilde{\boldsymbol{t}}(W^m) \neq \mathbf{0}$ for $m \in \{0, 3, 6, 9, 12, 15, 18, 21\}$, $\tilde{\boldsymbol{t}}(W^m) = \boldsymbol{e}_1$ for $m \in \{0, 6, 12, 18\}$ and $\tilde{\boldsymbol{t}}(W^m) = \boldsymbol{e}_2$ for $m \in \{3, 9, 15, 21\}$. We choose $\tilde{\boldsymbol{f}}_i(W^m) \neq \mathbf{0}$ $i = 1, 2$ for $m \in \{1, 4, 7, 10, 13, 16, 19, 22\}$, $\tilde{\boldsymbol{f}}_1(W^m) = \boldsymbol{e}_1$ and $\tilde{\boldsymbol{f}}_2(W^m) = \boldsymbol{e}_2$ for $m \in \{1, 7, 13, 19\}$, whereas $\tilde{\boldsymbol{f}}_1(W^m) = \boldsymbol{e}_2$ and $\tilde{\boldsymbol{f}}_2(W^m) = \boldsymbol{e}_1$for $m \in \{4, 10, 16, 22\}$. We choose $\tilde{\boldsymbol{f}}_i(W^m) \neq \mathbf{0}$ $i = 3, 4$ for $m \in \{2, 5, 8, 11, 14, 17, 20, 23\}$, $\tilde{\boldsymbol{f}}_3(W^m) = \boldsymbol{e}_1$ and $\tilde{\boldsymbol{f}}_4(W^m) = \boldsymbol{e}_2$ for $m \in \{2, 8, 14, 20\}$, whereas $\tilde{\boldsymbol{f}}_3(W^m)$ $= \boldsymbol{e}_2$ and $\tilde{\boldsymbol{f}}_4(W^m) = \boldsymbol{e}_1$for $m \in \{5, 11, 17, 23\}$.

Fig. 1, compares trace of the channel CRB in dB, i.e., $tr(CRB1)$ and $tr(CRB2)$ for designs $(i)$, $(ii)$, while Fig.2 illustrates $tr(CRB1)$ and $tr(CRB2)$ for design $(iii)$ as a function of SNR at $\zeta = 0.1$ and $\zeta = 0.4$. The simulation result confirm the fact that $tr(\text{CRB2})$ for its optimal design does not depend on $\zeta$. Furthermore, in all figures $tr(\text{CRB2})$ is lower than $tr(\text{CRB1})$. This is reasonable because CRB1 considers the symbols as nuisance, while CRB2 includes them in the unknown parameter vector.

## 6. CONCLUSION

In summary, we showed that different assumptions on the receiver structure may lead us to different designs for the affine precoders. In particular, we showed that for receivers which *decouples* channel and symbol estimation, a special form of orthogonality between $\boldsymbol{F}$ and $\boldsymbol{t}$ is required to lower down $tr(\text{CRB})$. Whereas for receivers which performs joint channel and symbol estimation, no such a restrict is required. Furthermore, we found that in the latter case, for the optimal design, trace of the channel CRB does not depend on the training power, but depends on the total transmitted power; whereas trace of the symbol CRB depends on the symbol power. This indicates that the training is absorbing the power that could have been used to estimate the symbols, without any improvement in channel estimate performance.

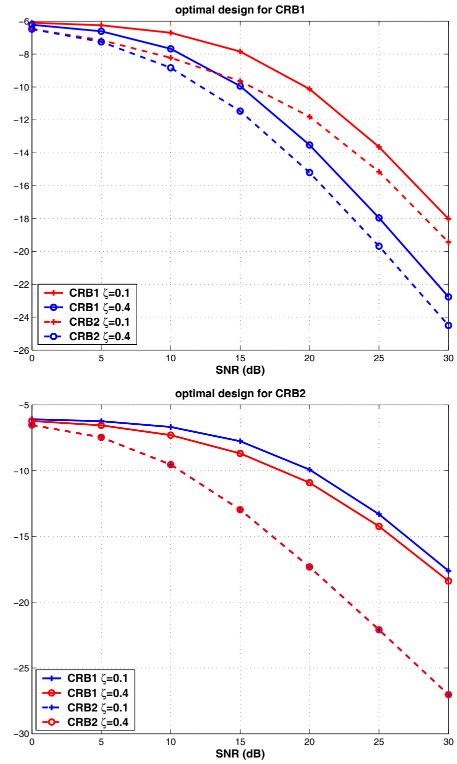


**Fig. 2**. $tr(CRB1)$ and $tr(CRB2)$ evaluated for design $(i)$ (top) and $(ii)$ (bottom)

## 7. REFERENCES


[1] J. K. Cavers," An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Vehicular Technology*, Vol.40, pp686-693, Nov. 1991.

[2] P. Hoeher and F. Tufvesson," Channel estimation with superimposed pilot sequence," *Proc. GLOBECOM Conf.* pp. 2162-2166, Rio de Janeiro, Brazil, 1999.

[3] J. H. Manton, I. Y. Mareels and Y.Hua, " Affine precoders for reliable communications," *Proc. ICASSP*, vol.5, pp. 2741-2744, Istanbul, Turkey, June 2000.

[4] G. B. Giannakis," Filterbanks for blind channel identification and equalization," *IEEE Sig Proc. Lett.*, vol.4, pp184-187, June 1997

[5] A. Vosoughi and A. Scaglione " Everything you always wanted to know about training: guidelines derived using the affine precoding framework, " *submitted to IEEE Trans. Signal Processing*

[6] S.Ohno and G. B. Giannakis, " Optimal training and redundant precodering for block transmissions with application to wirlesss OFDM," *IEEE Trans. Communications*, vol.50, No.12, pp. 2113 -2123, Dec. 2002.