Optimal Training for MIMO Frequency-Selective Fading Channels*

Xiaoli Ma, *Liuqing Yang*, and *Georgios B. Giannakis*

Dept. of ECE, Univ. of Minnesota, 200 Union St. SE, Minneapolis, MN 55455, USA

Abstract

High data rates give rise to frequency-selective channel effects. Space-time multiplexing and/or coding offer attractive means of combating fading, and boosting capacity of multi-antenna communications. As the number of antennas increases, channel estimation becomes challenging because the number of unknowns increases, and the power is split at the transmitter. In this paper, we design a low complexity optimal training scheme for block transmissions over frequency-selective channels with multiple antennas. The optimality in designing our training schemes consists of maximizing a lower bound on the average capacity that is shown to be equivalent to minimizing the mean-square error of the linear channel estimator. Simulation results confirm our theoretical analysis.

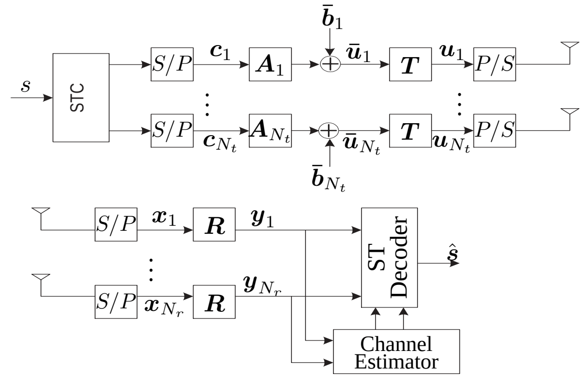


Figure 1: Discrete-time baseband equivalent MIMO model

1 Introduction

High rate wireless transmissions experience frequency-selective propagation effects. For channel state information (CSI) acquisition, three classes of methods are available: blind methods, which estimate CSI from the received symbols; differential ones that bypass CSI estimation by differential encoding; and input-output methods which rely on training symbols that are known to the receiver. Although the insertion of training symbols can be suboptimal and bandwidth consuming, training remains attractive especially when it decouples symbol from channel estimation and thus simplifies the receiver implementation, and relaxes the required identifiability conditions.

Training symbols can be placed either at the beginning of each burst (as a preamble), or, regularly throughout the burst. In rapidly fading or quasi-static fading channels, preamble-based training may not work well. This motivates embedding training symbols in each transmitted block, instead of concentrating them at the preamble. The so termed pilot symbol aided modulation (PSAM) originally developed for time-selective channels [2], has recently been extended to single-antenna frequency- and doubly-selective channels [6, 8], and also to *MIMO flat-fading* channels [4].

Training sequences for frequency-selective channels have been designed through exhaustive search for space-time (ST) trellis codes [3]. (Semi-) blind channel estimators have also been reported [9]. Here, we start from a general model, in which training symbols are superimposed with information symbols. In this paper, we design optimal PSAM schemes for linear minimum mean-square error (LMMSE) estimation of *MIMO frequency-selective* fading channels. Our design criteria jointly account for both channel estimation performance, and transmission rate. Due to lack of space, we will summarize our major results; detailed proofs can be found in [7].

Notation: Upper (lower) bold face letters will be used for matrices (column vectors). Superscript $^{\mathcal{H}}$ will denote Hermitian, * conjugate, ^{T} transpose. We will reserve $\otimes$ for the Kronecker product, E$[\cdot]$ for expectation, and tr$[\boldsymbol{A}]$ for matrix trace. $\boldsymbol{I}_N$ will denote the $N \times N$ identity matrix, and $[\boldsymbol{A}]_{m,n}$ will denote the (m,n)th entry of the matrix $\boldsymbol{A}$.

2 System Model

Our multi-antenna system employs N_t transmit- and N_r receive-antennas (see Fig. 1). The information bearing symbols are parsed into blocks of size $N_s \times 1$: $[\boldsymbol{s}(k)]_{n+1} := s(kN_s + n)$, where k denotes block index.

At the transmitter, each block $\boldsymbol{s}(k)$ is first encoded and/or multiplexed in space and time. The resulting N_t blocks $\{\boldsymbol{c}_\mu(k)\}_{\mu=1}^{N_t}$ have length N_c, and each is directed to one transmit antenna. At each, e.g., the μ-th transmit antenna, $\boldsymbol{c}_\mu(k)$ is processed by a matrix $\boldsymbol{A}_\mu$ of size $M \times N_c$, where $M \geq N_c$. Training blocks $\bar{\boldsymbol{b}}_\mu(k)$ of length M, which are known to both transmitter and receiver, are then superimposed to $\boldsymbol{A}_\mu \boldsymbol{c}_\mu(k)$ after the ST mapper to form $\bar{\boldsymbol{u}}_\mu(k) := \boldsymbol{A}_\mu \boldsymbol{c}_\mu(k) + \bar{\boldsymbol{b}}_\mu(k)$. Interblock interference (IBI) can be removed by processing blocks $\bar{\boldsymbol{u}}_\mu(k)$ with the $N \times M$ zero-padding matrix $\boldsymbol{T} := [\boldsymbol{I}_M \ \boldsymbol{0}_{M\times L}]^T$ to form blocks $\boldsymbol{u}_\mu(k) := \boldsymbol{T}\bar{\boldsymbol{u}}_\mu(k)$ of length $N = M + L$ with L denoting the maximum channel order across all antenna

*The work in this paper was supported by the NSF Grant No. 0122431, and ARL/CTA Grant No. DAAD19-01-2-011.

pairs. The blocks $\boldsymbol{u}_\mu(k)$ are then parallel-to-serial converted, pulse-shaped, carrier modulated, and transmitted from the μ-th transmit antenna. With IBI being eliminated, the ensuing channel estimator and symbol detectors operate on a block-by-block basis, we will hence forth omit the block index k.

Let $h^{(\nu,\mu)}(l)$, $l \in [0, L]$ denote the discrete-time baseband equivalent channel that includes transmit-receive filters as well as the frequency-selective propagation effects between the μ-th transmit antenna and the ν-th receive antenna. The samples at the ν-th receive-antenna filter output can be expressed as:

$$\boldsymbol{y}_\nu = \sum_{\mu=1}^{N_t} (\bar{\boldsymbol{H}}^{(\nu,\mu)} \boldsymbol{A}_\mu \boldsymbol{c}_\mu + \bar{\boldsymbol{H}}^{(\nu,\mu)} \bar{\boldsymbol{b}}_\mu) + \boldsymbol{\eta}_\nu, \qquad (1)$$

where $\bar{\boldsymbol{H}}^{(\nu,\mu)}$ is an $N \times M$ Toeplitz matrix with first column $[h^{(\nu,\mu)}(0), \cdots, h^{(\nu,\mu)}(L), 0, \cdots, 0]^T$.

In the next section, we will design optimal pairs $(\boldsymbol{A}_\mu, \bar{\boldsymbol{b}}_\mu)$, $\forall \mu \in [1, N_t]$. Using as criteria the conditional mutual information, and the MIMO channel's estimation error, we will seek optimal designs *for any* block size.

3 Optimal training design

Recall that a convolution between two vectors can be represented as the product of a Toeplitz matrix with a vector. Because convolution is a commutative operation, we deduce that $\bar{\boldsymbol{H}}^{(\nu,\mu)} \bar{\boldsymbol{b}}_\mu = \bar{\boldsymbol{B}}_\mu \boldsymbol{h}^{(\nu,\mu)}$, where $\bar{\boldsymbol{B}}_\mu$ is an $N \times (L+1)$ Toeplitz matrix having first column $[\bar{b}_\mu(1), \ldots, \bar{b}_\mu(M), 0, \ldots, 0]^T$ with $\bar{b}_\mu(m)$ being the m-th entry of $\bar{\boldsymbol{b}}_\mu$, and $[\boldsymbol{h}^{(\nu,\mu)}]_{l+1} := h^{(\nu,\mu)}(l)$. Eq. (1) becomes:

$$\boldsymbol{y}_\nu = \left[\bar{\boldsymbol{H}}^{(\nu,1)} \boldsymbol{A}_1 \; \cdots \; \bar{\boldsymbol{H}}^{(\nu,N_t)} \boldsymbol{A}_{N_t}\right] \boldsymbol{c} + \bar{\boldsymbol{B}} \boldsymbol{h}_\nu + \boldsymbol{\eta}_\nu, \quad (2)$$

where $\boldsymbol{h}_\nu := [(\boldsymbol{h}^{(\nu,1)})^T, \ldots, (\boldsymbol{h}^{(\nu,N_t)})^T]^T$, $\boldsymbol{c} := [\boldsymbol{c}_1^T, \ldots, \boldsymbol{c}_{N_t}^T]^T$, and $\bar{\boldsymbol{B}} := [\bar{\boldsymbol{B}}_1, \ldots, \bar{\boldsymbol{B}}_{N_t}]$. Concatenating the N_r received vectors into a single block $\boldsymbol{y} := [\boldsymbol{y}_1^T, \cdots, \boldsymbol{y}_{N_r}^T]^T$, we have

$$\boldsymbol{y} = \boldsymbol{\Phi} \boldsymbol{c} + (\boldsymbol{I}_{N_r} \otimes \bar{\boldsymbol{B}}) \boldsymbol{h} + \boldsymbol{\eta}, \qquad (3)$$

where $\boldsymbol{h} := [\boldsymbol{h}_1^T, \ldots, \boldsymbol{h}_{N_r}^T]^T$, and the $N_r N \times N_t N_c$ matrix $\boldsymbol{\Phi}$ is defined in terms of $\bar{\boldsymbol{H}}$'s and $\boldsymbol{A}$'s accordingly. The dependence of $\boldsymbol{\Phi}$ on $\boldsymbol{h}$, implies that estimating $\boldsymbol{h}$ and recovering $\boldsymbol{c}$ from $\boldsymbol{y}$ is a non-linear problem, whose joint solution is complicated or even impossible. Therefore, our first step is to judiciously design $(\boldsymbol{A}_\mu, \bar{\boldsymbol{b}}_\mu)$ in order to decouple channel estimation from symbol decoding.

3.1 Decoupling channel from symbol estimation

To estimate $\boldsymbol{h}$ from $\boldsymbol{y}$, we resort to the linear MMSE (LMMSE) estimator

$$\hat{\boldsymbol{h}} := \left(\sigma^2 \boldsymbol{R}_h^{-1} + \boldsymbol{I}_{N_r} \otimes (\bar{\boldsymbol{B}}^{\mathcal{H}} \bar{\boldsymbol{B}})\right)^{-1} (\boldsymbol{I}_{N_r} \otimes \bar{\boldsymbol{B}}^{\mathcal{H}}) \boldsymbol{y}, \quad (4)$$

where $\boldsymbol{R}_h := \mathrm{E}[\boldsymbol{h}\boldsymbol{h}^{\mathcal{H}}]$ is the channel covariance matrix, and σ^2 denotes the noise variance.

Since $\boldsymbol{c}$ is unknown, we cannot obtain $\hat{\boldsymbol{h}}$ directly from (3). To facilitate decoupling of channel estimation from symbol detection, we then need:

$$(\boldsymbol{I}_{N_r} \otimes \bar{\boldsymbol{B}}^{\mathcal{H}}) \boldsymbol{\Phi} \boldsymbol{c} = \boldsymbol{0}. \qquad (5)$$

For (5) to hold true for all $\boldsymbol{c}$, we need

Lemma 1 [Decoupling training from information] *Pairs $(\boldsymbol{A}_\mu, \bar{\boldsymbol{b}}_\mu)$ that guarantee decoupling of channel estimation from symbol detection, $\forall N$, have the form:*

$$\boldsymbol{A}_\mu = \begin{bmatrix} \boldsymbol{\Theta} \\ \boldsymbol{0}_{(M-N_c')\times N_c} \end{bmatrix}, \qquad \bar{\boldsymbol{b}}_\mu = \begin{bmatrix} \boldsymbol{0}_{N_c'+L} \\ \boldsymbol{b}_\mu \end{bmatrix}, \quad (6)$$

where the $(M - N_c' - L) \times 1$ vector $\mathbf{b}_\mu$ contains all the non-zero entries of $\bar{\boldsymbol{b}}_\mu$, and $\boldsymbol{\Theta}$ is an $N_c' \times N_c$ matrix that optionally precodes (if $\boldsymbol{\Theta} \neq \boldsymbol{I}_{N_c}$) $\boldsymbol{c}_\mu$ linearly.

It is worth emphasizing that we neither assumed insertion of training symbols *a fortiori*, nor we imposed time-division multiplexing (TDM) of training with information symbols at the outset. Lemma 1 revealed that in order to separate channel estimation from symbol decoding, we need to insert at least L zeros between the information symbols and the non-zero training symbols, per antenna.

Since the information blocks $\boldsymbol{c}_\mu$ have been encoded, in order to retain the structure of $\boldsymbol{c}_\mu$ and reduce decoding complexity, we simply choose $N_c' = N_c$, and $\boldsymbol{\Theta} = \boldsymbol{I}_{N_c}$. For this design, we then have

$$\bar{\boldsymbol{B}} := \begin{bmatrix} \boldsymbol{0}_{(N_c+L)\times N_t(L+1)} \\ \boldsymbol{B} \end{bmatrix}, \qquad (7)$$

where $\boldsymbol{B} := [\boldsymbol{B}_1 \cdots \boldsymbol{B}_{N_t}]$ is an $(N_b{+}L) \times N_t(L{+}1)$ matrix. Using (6) and (7), we can split $\boldsymbol{y}_\nu$ in (2) into two parts:

$$\boldsymbol{y}_\nu = \begin{bmatrix} \boldsymbol{y}_{\nu,c} \\ \boldsymbol{y}_{\nu,b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_c^{(\nu)} \boldsymbol{c} + \boldsymbol{\eta}_{\nu,c} \\ \boldsymbol{B} \boldsymbol{h}_\nu + \boldsymbol{\eta}_{\nu,b} \end{bmatrix}, \qquad (8)$$

where $\boldsymbol{H}_c^{(\nu)} := [\boldsymbol{H}_c^{(\nu,1)}, \ldots, \boldsymbol{H}_c^{(\nu,N_t)}]$, while $\boldsymbol{H}_c^{(\nu,\mu)}$ extracts the first $(N_c + L)$ rows and N_c columns of the channel matrix $\bar{\boldsymbol{H}}^{(\nu,\mu)}$. Stacking $\boldsymbol{y}_{\nu,b}$'s from all receive antennas into $\boldsymbol{y}_b$, we can now write the LMMSE channel estimator as:

$$\hat{\boldsymbol{h}} = \left(\sigma^2 \boldsymbol{R}_h^{-1} + \boldsymbol{I}_{N_r} \otimes (\boldsymbol{B}^{\mathcal{H}} \boldsymbol{B})\right)^{-1} (\boldsymbol{I}_{N_r} \otimes \boldsymbol{B}^{\mathcal{H}}) \boldsymbol{y}_b. \quad (9)$$

It can be readily verified that (9) is equivalent to (4). Notice that as $\hat{\boldsymbol{h}}$ in (9) depends only on the training blocks $\{\boldsymbol{b}_\mu\}_{\mu=1}^{N_t}$. Decoupling channel estimation from symbol detection offers flexibility in designing optimal training for MIMO channels without modifying ST designs having desirable rate-performance-complexity tradeoffs.

Defining the channel estimation error as $\breve{\boldsymbol{h}} := \boldsymbol{h} - \hat{\boldsymbol{h}}$, we can express its correlation as:

$$\boldsymbol{R}_{\breve{h}} := \mathrm{E}[\breve{\boldsymbol{h}} \breve{\boldsymbol{h}}^{\mathcal{H}}] = \left(\boldsymbol{R}_h^{-1} + \frac{1}{\sigma^2} \boldsymbol{I}_{N_r} \otimes (\bar{\boldsymbol{B}}^{\mathcal{H}} \bar{\boldsymbol{B}})\right)^{-1}. \quad (10)$$

The mean square error of $\hat{\boldsymbol{h}}$ is given by:

$$\sigma^2_{\breve{h}} := \mathrm{E}[\|\breve{\boldsymbol{h}}\|^2] = \mathrm{tr}\left[\left(\boldsymbol{R}_h^{-1} + \frac{1}{\sigma^2}\boldsymbol{I}_{N_r} \otimes (\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B})\right)^{-1}\right]. \quad (11)$$

Clearly, the design of training symbols across all transmit antennas affects $\sigma^2_{\breve{h}}$ through $\boldsymbol{B}$. To facilitate the ensuing analysis on the design of $\boldsymbol{B}$, we assume that:

A1) *The channel coefficients $h^{(\nu,\mu)}(l)$ are independent and the covariance matrices $\boldsymbol{R}_{h_\nu} := E[\boldsymbol{h}_\nu \boldsymbol{h}_\nu^{\mathcal{H}}]$ are the same $\forall \nu \in [1, N_r]$; i.e., $\boldsymbol{R}_h$ is diagonal and can be written as $\boldsymbol{R}_h = \boldsymbol{I}_{N_r} \otimes \boldsymbol{R}_{h_\nu}$ with trace $\mathrm{tr}[\boldsymbol{R}_h] = N_t N_r$.*

Note that A1) will not affect the optimality of our training design, simply because no CSI is assumed available at the transmitter. It can then be shown that $\sigma^2_{\breve{h}}$ in (11) is lower bounded by:

$$\sigma^2_{\breve{h}} \geq N_r \sum_{m=1}^{N_t(L+1)} \left([\boldsymbol{R}_{h_1}^{-1}]_{m,m} + \frac{1}{\sigma^2}[\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B}]_{m,m}\right)^{-1}, \quad (12)$$

where the equality holds if and only if $\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B}$ is a diagonal matrix. Therefore, the following condition is required to attain the minimum channel MSE:

C1) *For fixed N_b and N_c, the training symbols are inserted so that the matrix $\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B}$ is diagonal.*

For single-input single-output (SISO) channels, C1) coincides with that in [1].

3.2 Average capacity bounds

Our criterion for optimal training will be a lower bound of the average capacity, due to the difficulty associated with deriving an exact average capacity formula facilitating optimization. The optimal training parameters are those that maximize a lower bound of the average capacity. We will also invoke an upper bound of the average capacity to serve as a benchmark. Collecting the N_r received symbol blocks corresponding to all receive antennae, the information bearing part of (8) becomes:

$$\boldsymbol{y}_c = \boldsymbol{H}_c \boldsymbol{c} + \boldsymbol{\eta}_c. \quad (13)$$

Let $\mathcal{P}_c$ denote the power allocated to the information part. For a fixed power $\mathcal{P}_c := \mathrm{E}[\|\boldsymbol{c}\|^2]$, the mutual information between transmitted information symbols, and received symbols in (13) is given by $\mathcal{I}(\boldsymbol{y}_c; \boldsymbol{c}|\hat{\boldsymbol{h}})$. The channel capacity averaged over the random channel $\boldsymbol{h}$ is defined as:

$$C := \frac{1}{N}\mathrm{E}\left[\max_{p_c(\cdot),\, \mathcal{P}_c} \mathcal{I}(\boldsymbol{y}_c; \boldsymbol{c}|\hat{\boldsymbol{h}})\right], \quad (14)$$

where $p_c(\cdot)$ denotes the probability density function of $\boldsymbol{c}$.

When the channel estimate is perfect, i.e., $\hat{\boldsymbol{h}} \equiv \boldsymbol{h}$, the upper bound of the capacity can be obtained for a Gaussian distributed $\boldsymbol{c}$ with $\boldsymbol{R}_c := \mathrm{E}[\boldsymbol{c}\boldsymbol{c}^{\mathcal{H}}]$ (see e.g. [1, 6]), as:

$$\bar{C} := \frac{1}{N}\mathrm{E}\left[\max_{\boldsymbol{R}_c} \log\det\left(\boldsymbol{I}_{N_r(N_c+L)} + \frac{1}{\sigma^2}\boldsymbol{H}_c \boldsymbol{R}_c \boldsymbol{H}_c^{\mathcal{H}}\right)\right]. \quad (15)$$

Recall that $\boldsymbol{c}$ is obtained by encoding the information symbol block $\boldsymbol{s}$ in space and time. When $\boldsymbol{s}$ is Gaussian, $\boldsymbol{c}$ will also be approximately Gaussian for many ST mappers, such as block codes [9], and BLAST-type multiplexers. Therefore, in the following, we assume that:

A2) *The information bearing symbol block $\boldsymbol{c}$ is zero-mean Gaussian with covariance $\boldsymbol{R}_c = \bar{\mathcal{P}}_c \boldsymbol{I}_{N_t N_c}$, where $\bar{\mathcal{P}}_c := \mathcal{P}_c/(N_t N_c)$ is the normalized power.*

With A2), the capacity upper bound (15) becomes:

$$\bar{C} = \frac{1}{N}\mathrm{E}\left[\log\det\left(\boldsymbol{I}_{N_r(N_c+L)} + \frac{\bar{\mathcal{P}}_c}{\sigma^2}\boldsymbol{H}_c \boldsymbol{H}_c^{\mathcal{H}}\right)\right]. \quad (16)$$

When the estimate $\hat{\boldsymbol{h}}$ is not perfect, we have [c.f. (13)] $\boldsymbol{y}_c = \hat{\boldsymbol{H}}_c \boldsymbol{c} + \boldsymbol{v}$, where $\boldsymbol{v} := \breve{\boldsymbol{H}}_c \boldsymbol{c} + \boldsymbol{\eta}_c$ and $\breve{\boldsymbol{H}}_c := \boldsymbol{H}_c - \hat{\boldsymbol{H}}_c$. The correlation matrix of $\boldsymbol{v}$ is then given by

$$\boldsymbol{R}_v := \mathrm{E}[\boldsymbol{v}\boldsymbol{v}^{\mathcal{H}}] = \bar{\mathcal{P}}_c E[\breve{\boldsymbol{H}}_c \breve{\boldsymbol{H}}_c^{\mathcal{H}}] + \sigma^2 \boldsymbol{I}_{N_r(N_c+L)}. \quad (17)$$

Defining $\psi_l^{\breve{(1,\mu)}} := \mathrm{E}\left[\breve{h}^{(1,\mu)}(l)\breve{h}^{(1,\mu)*}(l)\right], \forall l \in [0, L]$, we can show that when $N_c \gg 2L$:

$$\mathrm{E}\left[\breve{\boldsymbol{H}}_c \breve{\boldsymbol{H}}_c^{\mathcal{H}}\right] \approx \frac{\sigma^2_{\breve{h}}}{N_r}\boldsymbol{I}_{N_r(N_c+L)}.$$

The resulting correlation matrix $\boldsymbol{R}_v$ in (17) is as follows:

$$\boldsymbol{R}_v \approx \left(\frac{\bar{\mathcal{P}}_c \sigma^2_{\breve{h}}}{N_r} + \sigma^2\right)\boldsymbol{I}_{N_r(N_c+L)}, \quad (18)$$

from which we deduce that as $\sigma^2_{\breve{h}}$ decreases, $\boldsymbol{R}_v$ decreases accordingly.

By employing the LMMSE channel estimator and using the channel Gaussian assumption, it has been shown in [6, Lemma 2] that the capacity in (14) is lower bounded by:

$$\underline{C} := \frac{1}{N}\mathrm{E}\left[\log\det\left(\boldsymbol{I}_{N_r(N_c+L)} + \bar{\mathcal{P}}_c \boldsymbol{R}_v^{-1}\hat{\boldsymbol{H}}_c \hat{\boldsymbol{H}}_c^{\mathcal{H}}\right)\right]. \quad (19)$$

Our objective is to select training parameters so that $\underline{C}$ is maximized. We have shown that [7]:

Lemma 2 *Suppose* C1), A1), *and* A2) *hold true, the information power $\bar{\mathcal{P}}_c$, and the training (information) block lengths N_b (N_c) are fixed. Then, maximizing $\underline{C}$ in* (19) *is equivalent to minimizing $\boldsymbol{R}_v$ in* (18)*, at high SNR.*

3.3 Optimal training parameters

When $\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B} = (\mathcal{P}_b/N_t)\boldsymbol{I}_{N_t(L+1)}$, C1) is satisfied, and the lower bound on the RHS of (11) is achieved:

$$\sigma^2_{\breve{h}} = N_r \sum_{\mu=1}^{N_t}\sum_{l=0}^{L}\left((\psi_l^{(1,\mu)})^{-1} + \frac{1}{\sigma^2}\frac{\mathcal{P}_b}{N_t}\right)^{-1}, \quad (20)$$

where $\psi_l^{(1,u)} := \mathrm{E}\left[h^{(1,\mu)}(l)h^{(1,\mu)*}(l)\right]$. Condition C1) can now be modified as:

C1') *For fixed N_b and N_c, the training symbols are inserted so that $\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B} = (\mathcal{P}_b/N_t)\boldsymbol{I}_{N_t(L+1)}$.*

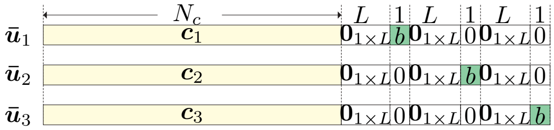


Figure 2: Training scheme example ($N_t = 3$).

By the definition of $\boldsymbol{B}$, the Toeplitz training matrices of every transmit antenna pair should satisfy:

$$\boldsymbol{B}_{\mu_1}^{\mathcal{H}}\boldsymbol{B}_{\mu_2} = \frac{\mathcal{P}_b}{N_t}\boldsymbol{I}_{L+1}\delta(\mu_1-\mu_2),\ \forall \mu_1,\mu_2 \in [1, N_t]. \quad (21)$$

Since, according to C1'), $\boldsymbol{B}$ is an $(N_b + L) \times N_t(L+1)$ matrix with full column rank, the minimum possible number of training symbols is given by $N_b = N_t(L+1) - L$, which suggests only one non-zero entry for each transmit antenna. In fact, for fixed $\mathcal{P}_c$ and $\mathcal{P}_b$, as N_b increases beyond $N_t(L+1) - L$, the average capacity bound $\underline{C}$ decreases. We summarize our results so far in the following:

Proposition 1 *Suppose* A1) *and* A2) *hold true. For fixed* $\mathcal{P}_c$ *and* $\mathcal{P}_b$*, the optimal placement of each block from the* μ*-th transmit antenna is* $[\boldsymbol{c}_\mu^T\ \mathbf{0}_L^T\ \boldsymbol{b}_\mu^T]^T$*, where* $\boldsymbol{b}_\mu$ *is selected to satisfy* (21) *with length* $N_b = N_t(L+1) - L$.

One simple example of the design is $\boldsymbol{b}_\mu = [\mathbf{0}^T_{(\mu-1)(L+1)}\ b\ \mathbf{0}^T_{(N_t-\mu)(L+1)}]^T$. With this structure, we have $\boldsymbol{B}_\mu = \sqrt{\bar{\mathcal{P}}_b}[\mathbf{0}_{(L+1)\times(\mu-1)(L+1)}\ \ \boldsymbol{I}_{L+1}\ \ \mathbf{0}_{(L+1)\times(N_t-\mu)(L+1)}]^T$. Although the proposed placement achieves the channel MMSE and maximizes the average capacity's lower bound $\underline{C}$, it is not unique for given N_t and L.

With the factor $\sigma^2_{\hat{H}_c} := \mathrm{tr}\left(\mathrm{E}\left[\hat{\boldsymbol{H}}_c\hat{\boldsymbol{H}}_c^{\mathcal{H}}\right]\right)$, we can normalize the estimated channel matrix as: $\bar{\hat{\mathbf{H}}}_c := \hat{\boldsymbol{H}}_c/\sigma_{\hat{H}_c}$. Substituting this and (18) into (19) yields the capacity lower bound with optimal placement:

$$\underline{C} = \frac{1}{N}\mathrm{E}\left[\log\det\left(\boldsymbol{I}_{N_r(N_c+L)} + \rho_{eff}\bar{\hat{\mathbf{H}}}_c\bar{\hat{\mathbf{H}}}_c^{\mathcal{H}}\right)\right], \quad (22)$$

where

$$\rho_{eff} := \frac{N_r\bar{\mathcal{P}}_c\sigma^2_{\hat{H}_c}}{\bar{\mathcal{P}}_c\sigma^2_{\check{h}} + N_r\sigma^2},$$

and it can be easily verified that $\sigma^2_{\hat{H}_c} = N_c(N_tN_r - \sigma^2_{\check{h}})$.

Consider now the total power $\mathcal{P} = \mathcal{P}_c + \mathcal{P}_b$, and define the power allocation factor $\alpha := \mathcal{P}_c/\mathcal{P} \in (0,1)$. Eq. (23) can then be expressed as:

$$\rho_{eff} = \frac{\alpha\mathcal{P}N_rN_c(N_tN_r - \sigma^2_{\check{h}})}{\alpha\mathcal{P}\sigma^2_{\check{h}} + N_tN_rN_c\sigma^2}. \quad (23)$$

Since $\sigma^2_{\check{h}}$ is dependent on $\psi_l^{(1,\mu)}$, it is difficult to find an optimal power allocation factor that does not depend on any CSI directly from (23). Therefore, we will consider the following three cases: low SNR, high SNR, and identically distributed channel taps.

Case A. Low SNR ($\mathcal{P}/N_t \ll \sigma^2$): In this case, (20) becomes $\sigma^2_{\check{h}} \approx N_tN_r - N_r(1-\alpha)\mathcal{P}/\sigma^2$. Plugging this result into (23), and differentiating ρ_{eff} with respect to α, we find that, at low SNR, the optimal power allocation factor is $\alpha \approx 1/2$.

Case B. High SNR ($\mathcal{P}/N_t \gg \sigma^2$): In this case, the optimal power allocation factor is $\alpha = 1/(1+\sqrt{\lambda})$, where $\lambda := N_t(L+1)/N_c$.

Case C. Identically distributed channel coefficients ($\psi_l^{(1,\mu)} = 1/(L+1)$): In this case, the optimal power allocation factor is given by

$$\alpha = \frac{\gamma - \sqrt{\gamma}\sqrt{\gamma-(1-\lambda)}}{1-\lambda}, \quad (24)$$

where $\gamma := 1 + N_t(L+1)\sigma^2/\mathcal{P}$.

In fact, the optimal placement and design of training symbols as illustrated in Fig. 2, are not unique. The structure of training blocks can be shuffled among the N_t transmit antennas without affecting either the channel MSE, or, the capacity lower bound.

4 Special Cases

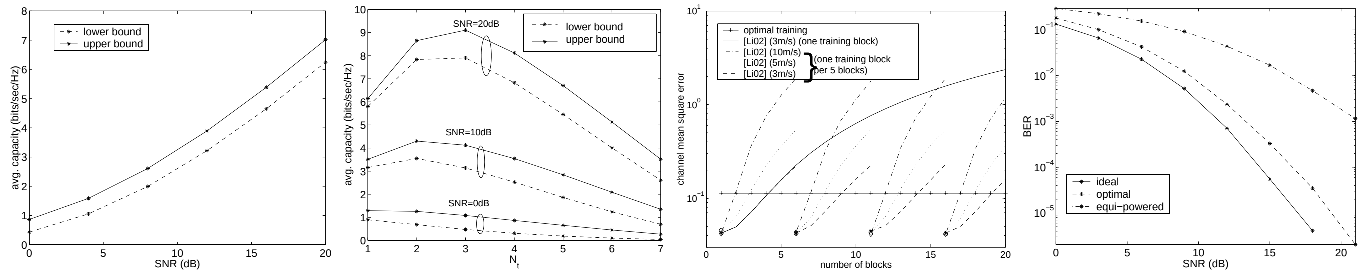
In the preceding section, we designed optimal training schemes for MIMO frequency-selective fading channels. The design leads to MIMO transmissions with TDM between training and information symbols. It is optimal in terms of minimizing channel MSE (without assuming any CSI besides the channel order L), and maximizing the lower bound on average capacity. The L guard symbols that are needed to decouple channel estimation from symbol decoding can be shared and play instrumental role in the ST code design of [9]; i.e., our optimal training designs merge well with ST mappers. Now, we will consider two special cases.

4.1 SISO frequency-selective fading channels

With one transmit and one receive antenna, the transmitted block $\boldsymbol{u} = \boldsymbol{T}_{zp}\boldsymbol{A}\boldsymbol{c} + \bar{\boldsymbol{b}}$ becomes $\boldsymbol{u} = [\boldsymbol{c}^T\ \mathbf{0}_L^T\ b\ \mathbf{0}_L^T]^T$, where the subscript μ is removed. This design has the same structure as the one in [1, Thm. 4]. In [8, Thm. 1], it was stated that the optimal number of training pilots is equal to the channel length $L+1$. The apparent discrepancy that requires $2L+1$ pilots comes from the fact that the redundancy of length L, which is needed to achieve IBI-free block transmission, is included in our training sequence, but is forgotten in [8]. At high SNR, our power allocation factor $\alpha = \sqrt{N_c}/(\sqrt{N_c} + \sqrt{L+1})$ coincides with the designs in [1, Thm. 5], and [8, Thm. 1].

4.2 MIMO flat-fading channels

When the MIMO channel is flat-fading, we have $L = 0$ which implies that one needs only a single pilot per block, per antenna. For such channels, the placement of training blocks has no effect on either the MSE or the average capacity, as long as the orthogonality among the N_t antennas is preserved. For this special case, a similar conclusion was reached in [4] where training signals with so-called



(a) (b) (c) (d)

Figure 3: (a) Average capacity bounds; (b) Average capacity bounds with increasing number of transmit antennas; (c) Channel MSE comparison between optimal training and the scheme in ; (d) BER comparison.

"orthonormal columns" is used. At high SNR, our optimal power allocation factor $\alpha = \sqrt{N_c}/(\sqrt{N_c} + \sqrt{N_t})$ is the same as that in [4, Corollary 1].

5 Simulations

In this section, we present simulations to validate our analyses and designs. The SNR is defined as the average received symbol power to noise ratio at each receive antenna.

Test case 1: In this example, we depict the relationships between the bounds on average capacity, and test the effect on average capacity of several training parameters.

(a) We select $(N_t, N_r, L, N) = (2, 2, 6, 62)$. We plot average capacity bounds versus SNR in Fig. 3(a). The capacity bounds increase monotonically as SNR increases.

(b) We depict the average capacity bounds versus the number of transmit-antennas in Fig. 3(b). Here we fix $(N_r, L, N) = (3, 6, 69)$, and the total power per block, $\mathcal{P}$. When $1 \leq N_t \leq 3$, the average capacity increases with the number of transmit antennas. However, when $N_t > 3$, the average capacity starts decreasing. This is because training symbols take over increasing portion of the whole transmitted block, leaving less to be used for information symbols.

Test case 2: We use as figure of merit the average channel MSE defined as: $\mathrm{E}[\|\hat{\boldsymbol{h}} - \boldsymbol{h}\|^2]$. We compare our schemes with the preamble training scheme proposed in [5]. The training blocks for [5] are selected according to [5, (12)]. We plot the channel MSE for the two cases when the channel is slowly time-varying. Each channel tap is generated by Jakes' model with a terminal speed of 3 (or 5 or 10) m/s, and a carrier frequency of 5.2 GHz. The variances of channel taps satisfy the exponential power profile. To equate the transmission rate, every training block is followed by four information blocks for the preamble scheme of [5]. Fig. 3(c) shows the channel MSE vs. the number of transmitted blocks at SNR$=10$dB. Although the channel is slowly time-varying, the MSE of our proposed schemes remains invariant across blocks, while that of [5] increases rapidly. Note that [5] yields smaller MSEs at the beginning of each re-training burst, because the total transmitted power is used for training.

Test case 3: To illustrate the compatibility of our optimal design with ST codes, we plot in Fig. 3(d) the BER performance. Two power allocation schemes are used: one with our optimal α, the other with $\bar{\mathcal{P}}_b = \bar{\mathcal{P}}_c$ (equi-powered). The ideal case corresponding to perfect channel estimates is also plotted. We select $(N_t, N_r, L, N) = (2, 1, 2, 128)$. The channel taps are independent complex Gaussian distributed with zero mean and variances with an exponential power profile. We use the ST code of [9]. Information symbols are estimated with zero-forcing estimators. We observe that: i) our optimal scheme outperforms the "equi-powered" scheme; ii) the penalty for inaccurate channel estimation is about 2dB.

References

[1] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal Placement of Training for Unknown Channels," *IEEE Trans. on Info. Theory*, pp. 2338–2353, Aug. 2002.

[2] J. K. Cavers, "An Analysis of Pilot Symbol Assisted Modulation for Rayleigh Fading Channels," *IEEE Trans. on Vehicular Technology*, pp. 686–693, Nov. 1991.

[3] C. Fragouli, N. Al-Dhahir, and W. Turin, "Finite-Alphabet Constant-Amplitude Training Sequence for Multiple-Antenna Broadband Transmissions," *Proc. of ICC*, vol. 1, pp. 6 -10, NY City, Apr., 2002.

[4] B. Hassibi and B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?," *IEEE Trans. on Info. Theory*, (submitted) Aug. 2000; URL: *http://mars.bell-labs.com/cm/ms/what/mars/index.html*.

[5] Y. Li, "Simplified Channel Estimation for OFDM Systems with Multiple Transmit Antennas," *IEEE Trans. on Wireless Comm.*, pp. 67–75, Jan. 2002.

[6] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal Training for Block Transmissions over Doubly-Selective Wireless Fading Channels," *IEEE Trans. on SP*, 2003 (to appear); see also *Proc. of ICASSP*, vol. 2, pp.1509–1512, Orlando, FL, May 13-17, 2002.

[7] X. Ma, L. Yang, and G. B. Giannakis, "Optimal Training for MIMO Frequency-Selective Fading Channels," *IEEE Trans. on Wireless Comm.*, (submitted), Oct., 2002.

[8] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "Optimal Training for Frequency-Selective Fading Channels," *Proc. of ICASSP*, Utah, vol. 4, pp. 2105–2108, 2001.

[9] S. Zhou and G. B. Giannakis, "Single-Carrier Space-Time Block Coded Transmissions over Frequency-Selective Fading Channels," *IEEE Trans. on Info. Theory*, Jan. 2003.