OPTIMAL TRAINING FOR MIMO FADING CHANNELS WITH TIME- AND FREQUENCY-SELECTIVITY

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

†Dept. of ECE, Univ. of Minnesota, Minneapolis, MN 55455

‡Dept. of ECE, Auburn University, Auburn, AL 36849

ABSTRACT

Demand for high data rate leads to frequency-selective propagation effects, whereas carrier frequency-offsets and Doppler effects induced by mobility introduce time-selectivity in wireless links. These fading channels, once acquired, offer joint multipath-Doppler diversity gains. In addition, 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. Optimal training sequences have so far been designed for flat-fading and frequency-selective multi-antenna systems. In this paper, we design a low complexity optimal training scheme for block transmissions over time- and frequency (a.k.a. doubly)-selective channels with multiple antennas. The optimality in designing our training schemes consists of maximizing a lower bound on the ergodic (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 which applies to both single- and multi-carrier transmissions.

1. INTRODUCTION

Demand for high data rate leads to time- and frequency-selective fading in mobile wireless channels. These fading channels, once acquired, offer joint multipath-Doppler diversity gains. Therefore, the quality of channel acquisition plays an important role on the overall system performance.

For the channel state information (CSI) acquisition at the receiver, two classes of methods are applicable: one is the class of so called blind methods, the other relies on training symbols known to the receiver. Relative to blind methods, training based schemes are more bandwidth consuming, but typically require shorter data records, and entail lower receiver complexity by decoupling symbol detection from channel estimation.

Training symbols can be either distributed throughout each burst, or, gathered as a preamble. In time-selective fading channels, the former is well motivated. Originally developed for time-selective channels [2], pilot symbol assisted modulation has been extended to MIMO flat and frequency-selective fading channels [6, 8], and single-antenna time- and/or frequency-selective channels [9, 7]. Clearly, optimal training needs to be designed for general MIMO time- and frequency-selective fading channels, which are frequently encountered in wireless communications with high rate and high mobility.

In this paper, we design optimal training parameters for MIMO *time- and frequency-selective* fading channels. Our training parameters are optimal in the sense of maximizing a capacity lower bound, and minimizing the channel estimation mean-square error (MSE). Since our MIMO doubly-selective channel subsumes many other channel types, the resultant optimal training scheme is the most general among existing ones, and can also be adapted to single-antenna, time- and/or frequency-selective fading channels.

Work was supported by ARL/CTA Grant No. DAAD19-01-2-0011.

2. SYSTEM MODEL

In this section, we first present the multi-antenna frequency- and time-selective channel model, and then propose a block transmission system model.

2.1. MIMO Doubly-Selective Channel Model

Our multi-antenna system contains N_t transmit antennas and N_r receive antennas. For the μth transmit antenna, symbols each of duration T_s are transmitted in $N \times 1$ blocks $\boldsymbol{u}_\mu(k)$, with k denoting the block index. The symbol sequence that is transmitted from the μth transmit antenna can thus be expressed as $u_\mu(kN+n)$, $\forall n \in [0, N-1]$.

Let $h^{(\nu,\mu)}(t;\tau)$ denote the time-varying impulse response of the channel including transmit-receive filters as well as the doubly-selective propagation effects between the μth transmit antenna and the νth receive antenna. Notice that $h^{(\nu,\mu)}(t;\tau)$ is two-dimensional: its dependence on t captures the time-variation of the channel; while its dependence on τ captures the frequency-selectivity of the channel. With $H^{(\nu,\mu)}(f;\tau)$ denoting the Fourier transform of $h^{(\nu,\mu)}(t;\tau)$ with respect to t, we observe that $H^{(\nu,\mu)}(f;\tau) \approx 0$, for $|f| > f_{max}$ or $\forall \tau > \tau_{max}$, where τ_{max} and f_{max} are the channel's delay spread and Doppler spread, respectively. Sampling $H^{(\nu,\mu)}(f;\tau)$ with period $1/(NT_s)$ in f and T_s in τ, we obtain $(Q+1)(L+1)$ samples with $L := \lceil \tau_{max}/T_s \rceil$ and $Q := 2\lceil f_{max}NT_s \rceil$.

From the channel estimation viewpoint, we treat the channel during each block duration of NT_s as deterministic, or, as a realization of a random process. Consequently, the $(Q+1)(L+1)$ time-frequency samples corresponding to each (μ,ν) pair stay invariant within each block duration. Using these samples, the discrete-time equivalent channel $h^{(\nu,\mu)}(kN+n;l)$, $\forall n \in [0, N-1]$, $l \in [0, L]$, can be expressed according to the Basis Expansion Model (BEM) as [7]:

$$h^{(\nu,\mu)}(kN+n;l) = \sum_{q=0}^{Q} h_q^{(\nu,\mu)}(k;l) e^{j\omega_q n}, \tag{1}$$

where $\omega_q := \frac{2\pi}{N}(q - \frac{Q}{2})$. As in [5, 7], we assume the following:

A1) *Parameters τ_{max}, f_{max} (and thus L, Q) are bounded, known, and satisfy* $2\tau_{max} f_{max} < 1$.

2.2. Block Transmission System Model

With the discrete-time equivalent channel model, the received sequence at the νth receive antenna can be expressed as $x_\nu(kN+n) = \sum_{\mu=1}^{N_t} \sum_{l=0}^{L} h^{(\nu,\mu)}(kN+n;l) u_\mu(kN+n-l) + \eta_\nu(kN+n)$, where $\eta_\nu(kN+n)$ is the additive complex Gaussian noise at the νth receive-antenna, with zero-mean and variance σ^2. Casting the received sequence into blocks of size $N > L$, the matrix-vector input-output relationship is given by:

$$\boldsymbol{x}_\nu(k) = \sum_{\mu=1}^{N_t} \Big[\boldsymbol{H}^{(\nu,\mu)}(k) \boldsymbol{u}_\mu(k) + \boldsymbol{H}_{\mathsf{ibi}}^{(\nu,\mu)}(k) \boldsymbol{u}_\mu(k-1) \Big] + \boldsymbol{\eta}_\nu(k),$$

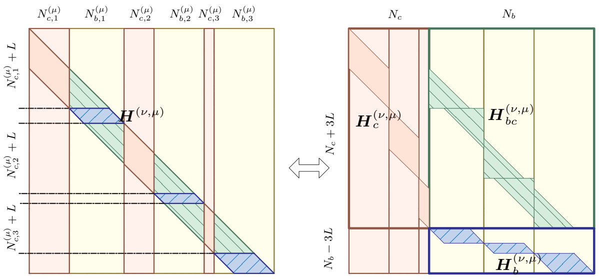

Fig. 1. *Left:* Structure of the channel matrix in (2); *Right:* Structure of the channel matrix in (4).

where $\boldsymbol{H}^{(\nu,\mu)}(k)$ and $\boldsymbol{H}_{\mathsf{ibi}}^{(\nu,\mu)}(k)$ are $N\times N$ lower and upper triangular matrices with entries $\left[\boldsymbol{H}^{(\nu,\mu)}(k)\right]_{n,m} = h^{(\nu,\mu)}(kN+n; n-m)$ and $\left[\boldsymbol{H}_{\mathsf{ibi}}^{(\nu,\mu)}(k)\right]_{n,m} = h^{(\nu,\mu)}(kN+n; N+n-m)$, $\forall n, m \in [0, N-1]$. The frequency-selective channel induced inter-block interference (IBI) term is captured in the second summand of the summation. Although the channel matrices $\boldsymbol{H}^{(\nu,\mu)}(k)$ and $\boldsymbol{H}_{\mathsf{ibi}}^{(\nu,\mu)}(k)$ both bear a banded structure, they are *not* Toeplitz, due to the channel's time-variation.

Since the non-zero entries of $\boldsymbol{H}_{\mathsf{ibi}}^{(\nu,\mu)}(k)$ are confined to its last L columns, the IBI term vanishes if each block $\boldsymbol{u}_\mu(k)$ contains L trailing zeros. A block transmission system model free of IBI can then be reached:

$$\boldsymbol{x}_\nu = \sum_{\mu=1}^{N_t} \boldsymbol{H}^{(\nu,\mu)}\boldsymbol{u}_\mu + \boldsymbol{\eta}_\nu, \tag{2}$$

where the block index k is omitted for notational simplicity, and the following condition is used:

C1) *Each block $\boldsymbol{u}_\mu$ contains L trailing zeros.*

In the next section, we will further specify the structure of transmitted blocks $\boldsymbol{u}_\mu$, the channel estimation scheme, and the corresponding mean-square error (MSE).

3. CHANNEL ESTIMATION AND MSE

To decouple channel estimation from symbol detection, each transmitted block $\boldsymbol{u}_\mu$ consists of segments of information and training symbols, with the general structure

$$\boldsymbol{u}_\mu = [\boldsymbol{c}_{\mu,1}^T, \boldsymbol{b}_{\mu,1}^T, \ldots, \boldsymbol{c}_{\mu,P_\mu}^T, \boldsymbol{b}_{\mu,P_\mu}^T]^T, \tag{3}$$

where $\boldsymbol{c}_{\mu,p}$ and $\boldsymbol{b}_{\mu,p}$ of lengths $N_{c,p}^{(\mu)}$ and $N_{b,p}^{(\mu)}$, $\forall p \in [1, P_\mu]$, denote information and training symbol sub-blocks, respectively. In general, these lengths satisfy

$$\sum_{p=1}^{P_\mu} N_{c,p}^{(\mu)} = N_c, \quad \sum_{p=1}^{P_\mu} N_{b,p}^{(\mu)} = N_b, \quad \text{and} \quad N_c + N_b = N,$$

to yield information symbol rate common across all transmit antennas. Moreover, since Condition C1) requires the last L entries of $\boldsymbol{b}_{\mu,P_\mu}$ to be zeros, we need $N_{b,P_\mu}^{(\mu)} > L, \forall \mu$.

Gathering information and training symbols per block $\boldsymbol{u}_\mu$ to form $\boldsymbol{c}_\mu$ and $\boldsymbol{b}_\mu$, respectively, we wish to estimate the time- and frequency-selective MIMO channel from the received samples corresponding to $\{\boldsymbol{b}_\mu\}_{\mu=1}^{N_t}$. Using the estimates $\hat{\boldsymbol{H}}^{(\nu,\mu)}$, $\forall \nu, \mu$, we can then recover the unknown information symbols $\boldsymbol{c}_\mu$.

Let us now take the noise-free received block corresponding to the (ν, μ) antenna pair $\boldsymbol{x}^{(\nu,\mu)} := \boldsymbol{H}^{(\nu,\mu)}\boldsymbol{u}_\mu$. From the structure of $\boldsymbol{u}_\mu$ and $\boldsymbol{H}^{(\nu,\mu)}$ (see Fig. 1), we deduce that some segments of $\boldsymbol{x}^{(\nu,\mu)}$ are free of unknown information symbols transmitted from the μth antenna and can be potentially used for channel estimation. However, since the received symbol block $\boldsymbol{x}_\nu$ is the superposition of $\boldsymbol{x}^{(\nu,\mu)}$'s over μ, it is possible that $\boldsymbol{x}_{b,p}^{(\nu,\mu)}$ is contaminated by unknown information symbols from other transmit antennas, and thus rendered unusable for channel estimation. To ensure the separation of information and training symbols and gain maximum usage of training pilots, the following condition has to be satisfied:

C2) *The training symbols should be located such that $P_\mu = P$, $N_{c,p}^{(\mu)} = N_{c,p}$ and $N_{b,p}^{(\mu)} = N_{b,p}$, for all $\mu \in [1, N_t]$.*

To establish our channel estimator, let us first re-arrange entries of $\boldsymbol{u}_\mu$ such that all training symbols are gathered together. Permutating $\boldsymbol{x}_\nu$ accordingly, we have at the νth antenna:

$$\begin{bmatrix}\boldsymbol{x}_{\nu,c}\\ \boldsymbol{x}_{\nu,b}\end{bmatrix} = \sum_{\mu=1}^{N_t}\begin{bmatrix}\boldsymbol{H}_c^{(\nu,\mu)} & \boldsymbol{H}_{bc}^{(\nu,\mu)}\\ \boldsymbol{0} & \boldsymbol{H}_b^{(\nu,\mu)}\end{bmatrix}\cdot\begin{bmatrix}\boldsymbol{c}_\mu\\ \boldsymbol{b}_\mu\end{bmatrix} + \begin{bmatrix}\boldsymbol{\eta}_{\nu,c}\\ \boldsymbol{\eta}_{\nu,b}\end{bmatrix}, \tag{4}$$

where the channel matrix is re-structured as illustrated in Fig. 1. Notice that the received sub-block $\boldsymbol{x}_{\nu,b}$ only relies on known training symbols, and can thus be used to establish our channel estimate. From Fig. 1, it is evident that $\boldsymbol{H}_b^{(\nu,\mu)}$ consists of P sub-matrices $\boldsymbol{H}_{b,p}^{(\nu,\mu)}$, each corresponding to a training sub-block $\boldsymbol{b}_{\mu,p}$. It can be shown that $\boldsymbol{H}_{b,p}^{(\nu,\mu)}$ has size $(N_{b,p}-L)\times N_{b,p}$. Therefore, for each training sub-block to contain sufficient training symbols for channel estimation, the following condition is needed:

C3) *The length of each training sub-block $\boldsymbol{b}_{\mu,p}$ is at least $L+1$; i.e., $N_{b,p} > L$, $\forall\, p$.*

Condition C3) shows that the pilot symbols should be inserted in sub-blocks of size at least $L+1$.

Applying (1) to each $\boldsymbol{H}_{b,p}^{(\nu,\mu)}$, $\forall p$, the training input-output relationship becomes [c.f. (4)]:

$$\boldsymbol{x}_{\nu,b} = \sum_{\mu=1}^{N_t}\sum_{q=0}^{Q}\begin{bmatrix}\boldsymbol{D}_{b,1,q}\boldsymbol{H}_{b,1,q}^{(\nu,\mu)}\boldsymbol{b}_{\mu,1}\\ \vdots\\ \boldsymbol{D}_{b,P_\mu,q}\boldsymbol{H}_{b,P,q}^{(\nu,\mu)}\boldsymbol{b}_{\mu,P}\end{bmatrix} + \boldsymbol{\eta}_{\nu,b}, \tag{5}$$

where $\boldsymbol{D}_{b,p,q}$ and $\boldsymbol{H}_{b,p,q}^{(\nu,\mu)}$ are corresponding sub-matrices of $\boldsymbol{D}_q = \text{diag}\{1, \exp(j\omega_q), \ldots, \exp(j\omega_q(N-1))\}$, and the lower triangular Toeplitz matrix $\boldsymbol{H}_q^{(\nu,\mu)}$ with first column $[h_q^{(\nu,\mu)}(0), \ldots, h_q^{(\nu,\mu)}(L), 0, \ldots, 0]^T$, respectively. Using the commutation property of convolution, we have $\boldsymbol{H}_{b,p,q}^{(\nu,\mu)}\boldsymbol{b}_{\mu,p} = \boldsymbol{B}_{\mu,p}\boldsymbol{h}_q^{(\nu,\mu)}$, with the $(N_{b,p}-L)\times(L+1)$ Toeplitz matrix:

$$\boldsymbol{B}_{\mu,p} := \begin{bmatrix} b_{\mu,p}(L) & \ldots & b_{\mu,p}(0)\\ \vdots & \ddots & \vdots\\ b_{\mu,p}(N_{b,p}-1) & \ldots & b_{\mu,p}(N_{b,p}-1-L)\end{bmatrix} \tag{6}$$

with $b_{\mu,p}(n)$ being the $(n+1)$st entry of sub-block $\boldsymbol{b}_{\mu,p}$, and $\boldsymbol{h}_q^{(\nu,\mu)}$ containing $\{h_q^{(\nu,\mu)}(l)\}_{l=0}^{L}$. Concatenate $\{\boldsymbol{h}_q^{(\nu,\mu)}\}_{q=1}^{Q}$ into a $(Q+1)(L+1)\times 1$ vector $\boldsymbol{h}^{(\nu,\mu)}$, and define matrices

$$\boldsymbol{\Xi}_\mu := \begin{bmatrix}\boldsymbol{D}_{b,1,0}\boldsymbol{B}_{\mu,1} & \ldots & \boldsymbol{D}_{b,1,Q}\boldsymbol{B}_{\mu,1}\\ \vdots & \ddots & \vdots\\ \boldsymbol{D}_{b,P,0}\boldsymbol{B}_{\mu,P} & \ldots & \boldsymbol{D}_{b,P,Q}\boldsymbol{B}_{\mu,P}\end{bmatrix}.$$

We can then re-express (5) as:

$$\boldsymbol{x}_{\nu,b} = \sum_{\mu=1}^{N_t}\boldsymbol{\Xi}_\mu\boldsymbol{h}^{(\nu,\mu)} + \boldsymbol{\eta}_{\nu,b} = \boldsymbol{\Xi}\boldsymbol{h}_\nu + \boldsymbol{\eta}_{\nu,b} \tag{7}$$

with obvious substitutions. Notice that the matrix $\boldsymbol{\Xi}$ is common for all $\nu \in [1, N_r]$. Collecting $\boldsymbol{x}_{\nu,b}$ and $\boldsymbol{h}_\nu$ for all ν, we have:

$$\boldsymbol{x}_b = (\boldsymbol{I}_{N_r}\otimes\boldsymbol{\Xi})\boldsymbol{h} + \boldsymbol{\eta}_b. \tag{8}$$

Similar to [6, 7], we will rely on the Wiener solution of (7) that yields the linear MMSE (LMMSE) channel estimator:

$$\hat{\boldsymbol{h}} = \left(\sigma^2 \boldsymbol{R}_h^{-1} + (\boldsymbol{I}_{N_r} \otimes \boldsymbol{\Xi}^{\mathcal{H}}\boldsymbol{\Xi})\right)^{-1} (\boldsymbol{I}_{N_r} \otimes \boldsymbol{\Xi}^{\mathcal{H}})\boldsymbol{x}_b, \tag{9}$$

where $\boldsymbol{R}_h := E[\boldsymbol{h}\boldsymbol{h}^{\mathcal{H}}]$ is the channel correlation matrix. To facilitate the ensuing analysis, we further assume that:

A2) *The channel coefficients are independent Gaussian distributed, and the channel covariance matrices $\boldsymbol{R}_{h_\nu} := E[\boldsymbol{h}_\nu \boldsymbol{h}_\nu^{\mathcal{H}}]$ are the same across $\nu \in [1, N_r]$; i.e., $\boldsymbol{R}_h = \boldsymbol{I}_{N_r} \otimes \boldsymbol{R}_{h_1}$ with trace $N_t N_r$.*

As a result, the covariance matrix of the channel estimation error $\breve{\boldsymbol{h}} := \boldsymbol{h} - \hat{\boldsymbol{h}}$, is given by:

$$\boldsymbol{R}_{\breve{h}} := E[\breve{\boldsymbol{h}}\breve{\boldsymbol{h}}^{\mathcal{H}}] = \boldsymbol{I}_{N_r} \otimes \left(\boldsymbol{R}_{h_1}^{-1} + \frac{1}{\sigma^2}\boldsymbol{\Xi}^{\mathcal{H}}\boldsymbol{\Xi}\right)^{-1}. \tag{10}$$

Consequently, the MSE of $\hat{\boldsymbol{h}}$ is given by:

$$\sigma_{\breve{h}}^2 := E[\|\breve{\boldsymbol{h}}\|^2] = N_r \cdot tr\left[\left(\boldsymbol{R}_{h_1}^{-1} + \frac{1}{\sigma^2}\boldsymbol{\Xi}^{\mathcal{H}}\boldsymbol{\Xi}\right)^{-1}\right], \tag{11}$$

and is lower bounded by:

$$\sigma_{\breve{h}}^2 \geq N_r \sum_m \frac{1}{[\boldsymbol{R}_{h_1}^{-1} + \frac{1}{\sigma^2}\boldsymbol{\Xi}^{\mathcal{H}}\boldsymbol{\Xi}]_{m,m}}, \tag{12}$$

where the equality holds if and only if $\boldsymbol{\Xi}^{\mathcal{H}}\boldsymbol{\Xi}$ is diagonal. Evidently, the design of training symbols across all transmit antennas affects the MMSE through $\boldsymbol{\Xi}$, and the following condition is required for our training strategy to attain the MMSE:

C4) *For fixed N_b and N_c, the training symbols should be inserted so that the matrix $\boldsymbol{\Xi}^{\mathcal{H}}\boldsymbol{\Xi}$ is diagonal.*

Condition C4) coincides with that in [1, 3, 8, 10].

So far, we have established our channel estimator in (9), and the conditions C1)-C4) for optimal training parameter design. But some training parameters remain to be decided, such as the placement and the optimal number of training symbols. These parameters affect the performance of the channel estimator, the effective transmission rate $\mathcal{R} := N_c/N$, the mutual information, as well as the bit error rate (BER). In the following, we will select these training parameters by optimizing an average capacity bound. Similar to the single-antenna case in [7], we will show that the optimization of the average capacity bound is equivalent to the minimization of the channel MMSE.

4. CAPACITY BOUNDS

Due to the difficulty associated with the evaluation of the average capacity itself, we choose its lower bound as the optimality criterion. The optimal training parameters are those which maximize this lower capacity bound, while the upper capacity bound is viewed as a benchmark for the maximum achievable rate.

Collecting the N_r received symbol blocks each corresponding to a receive antenna, the input-output relationship (4) becomes:

$$\begin{bmatrix}\boldsymbol{x}_c\\ \boldsymbol{x}_b\end{bmatrix} = \boldsymbol{H} \cdot \begin{bmatrix}\boldsymbol{c}\\ \boldsymbol{b}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\eta}_c\\ \boldsymbol{\eta}_b\end{bmatrix} = \begin{bmatrix}\boldsymbol{H}_c & \boldsymbol{H}_{bc}\\ \boldsymbol{0} & \boldsymbol{H}_b\end{bmatrix} \cdot \begin{bmatrix}\boldsymbol{c}\\ \boldsymbol{b}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\eta}_c\\ \boldsymbol{\eta}_b\end{bmatrix} \tag{13}$$

where

$$\boldsymbol{H}_a := \begin{bmatrix}\boldsymbol{H}_a^{(1,1)} & \dots & \boldsymbol{H}_a^{(1,N_t)}\\ \vdots & \ddots & \vdots\\ \boldsymbol{H}_a^{(N_r,1)} & \dots & \boldsymbol{H}_a^{(N_r,N_t)}\end{bmatrix},\ a = b,\ c,\ \text{or},\ bc;$$

and $\boldsymbol{c}$, $\boldsymbol{b}$ and $\boldsymbol{\eta}_c$, $\boldsymbol{\eta}_b$ are concatenated information symbols, training pilots and noise vectors, respectively.

Let $\mathcal{P}$ denote the total transmit-power per block, $\mathcal{P}_c$ the power allocated to the information signal part, and $\mathcal{P}_b$ the power assigned to the training part. With $\boldsymbol{b}$ conveying no information, the mutual information between transmitted information symbols, and received symbols in (13) is given by $\mathcal{I}(\boldsymbol{x}_c; \boldsymbol{c}|\hat{\boldsymbol{H}})$, where $\hat{\boldsymbol{H}}$ is an estimate of $\boldsymbol{H}$. The channel capacity averaged over the random channel $\boldsymbol{H}$ is defined as:

$$C := \max_{P_c(\cdot),\ \mathcal{P}_c} \frac{1}{N} E\left[\mathcal{I}(\boldsymbol{x}_c; \boldsymbol{c}|\hat{\boldsymbol{H}})\right] \tag{14}$$

for any fixed power $\mathcal{P}_c := E[\|\boldsymbol{c}\|^2]$, where $P_c(\cdot)$ denotes the probability density function of $\boldsymbol{c}$.

An upper bound of the capacity can be obtained when the channel estimate is perfect, i.e., $\hat{\boldsymbol{H}} \equiv \boldsymbol{H}$, and when $\boldsymbol{c}$ is Gaussian distributed with $\boldsymbol{R}_c := E[\boldsymbol{c}\boldsymbol{c}^{\mathcal{H}}]$:

$$\bar{C} := \max_{P_c(\cdot),\ \mathcal{P}_c} \frac{1}{N} E\left[\log\det\left(\boldsymbol{I}_{N_r(N_c+LP)} + \frac{1}{\sigma_w^2}\boldsymbol{H}_c \boldsymbol{R}_c \boldsymbol{H}_c^{\mathcal{H}}\right)\right].$$

Since $\boldsymbol{c}$ is the ST encoded information symbol block, it is approximately Gaussian for many ST mappers provided that the original information symbols are Gaussian. For simplicity, we hereafter assume that:

A3) *The information bearing symbol block $\boldsymbol{c}$ is zero-mean Gaussian with covariance $\boldsymbol{R}_c = \bar{\mathcal{P}}_c \boldsymbol{I}_{N_t N_c}$, and $\bar{\mathcal{P}}_c := \mathcal{P}_c/(N_t N_c)$.*

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

$$\boldsymbol{R}_v = \bar{\mathcal{P}}_c E[\breve{\boldsymbol{H}}_c \breve{\boldsymbol{H}}_c^{\mathcal{H}}] + E[\breve{\boldsymbol{H}}_{bc}\boldsymbol{b}\boldsymbol{b}^{\mathcal{H}}\breve{\boldsymbol{H}}_{bc}^{\mathcal{H}}] + \sigma^2 \boldsymbol{I}_{N_r(N_c+LP)}. \tag{15}$$

With $\boldsymbol{v}$ being uncorrelated with $\boldsymbol{c}$, it follows from [6, Lemma2] that the worst case noise is Gaussian with zero mean and $\boldsymbol{R}_v$. This implies that the lower bound of capacity is:

$$\underline{C} := \frac{1}{N} E\left[\log\det\left(\boldsymbol{I}_{N_r(N_c+LP)} + \bar{\mathcal{P}}_c \boldsymbol{R}_v^{-1}\hat{\boldsymbol{H}}_c \hat{\boldsymbol{H}}_c^{\mathcal{H}}\right)\right]. \tag{16}$$

Maximizing $\underline{C}$ entails minimizing $\boldsymbol{R}_v$ in the positive semi-definite sense [7]. Specifically, we establish the following result

Lemma 1 *Consider a fixed number of training symbols N_b adhering to C1)-C3). Among all $\{\boldsymbol{b}_{\mu,p}\}_{\mu=1}^{N_t}$ choices that satisfy C4) and lead to identical $\{\boldsymbol{R}_{\breve{h}_\nu}\}_{\nu=1}^{N_r}$, the design which satisfies that $N_{b,p} \geq 2L+1$ and has the first L and last L entries of $\{\boldsymbol{b}_{\mu,p}\}_{\mu=1}^{N_t}$, $\forall p \in [1, P]$ equal to zero, achieves the minimum $\boldsymbol{R}_v$.*

Accordingly, we modify condition C3) as follows:

C3') *The training sub-block is given by: $\boldsymbol{b}_{\mu,p} := [\boldsymbol{0}_L^T\ \bar{\boldsymbol{b}}_{\mu,p}^T\ \boldsymbol{0}_L^T]^T$, $\forall \mu \in [1, N_t], p \in [1, P]$, with $\bar{\boldsymbol{b}}_{\mu,p}$ of length $N_{b,p} - 2L \geq 1$.*

Intuitively, the L zeros between the information and training sub-blocks eliminate the inter-sub-block interference. It then follows that: $\boldsymbol{R}_v = \bar{\mathcal{P}}_c E[\breve{\boldsymbol{H}}_c \breve{\boldsymbol{H}}_c^{\mathcal{H}}] + \sigma^2 \boldsymbol{I}_{N_r(N_c+LP)}$. When $N_{c,p} \gg 2L$, the latter can be approximated as

$$\boldsymbol{R}_v \approx \left[\frac{\bar{\mathcal{P}}_c}{N_r}\sigma_{\breve{h}}^2 + \sigma^2\right] \boldsymbol{I}_{N_r(N_c+LP)}. \tag{17}$$

Clearly, $\boldsymbol{R}_v$ is minimized when $\sigma_{\breve{h}}^2$ is minimized and the capacity lower bound $\underline{C}$ in (16) is maximized accordingly.

5. OPTIMAL TRAINING PARAMETERS

Starting from the optimization criteria: capacity lower bound $\underline{C}$, and channel estimation MSE $\sigma_{\breve{h}}^2$, we have shown that the maximization of the former is equivalent to the minimization of the latter. Our objective then becomes selecting training parameters to maximize $\underline{C}$.

These parameters are: the placement of training symbols, the number of training symbols, and the power allocation among information and training symbols.

Regarding the placement of training symbols, we established the following:

Proposition 1 *Suppose* A1)-A3) *holds true. For fixed $\mathcal{P}_c$ and $\mathcal{P}_b$, the following placement is optimal: all information sub-blocks have identical length $N_{c,p} = N_c/P$, $\forall p$; the pilot sub-blocks from the μth transmit antenna have structure $[\mathbf{0}_{\mu(L+1)-1}, b, \mathbf{0}_{(N_t-\mu)(L+1)+L}]$ and lengths $N_{b,p} = N_t(L+1) + L$, $\forall p$; all pilot symbols are equipowered with $\mathcal{P}_b/(N_t P)$.*

It can be readily verified that the placement as in Proposition 1 satisfies conditions C1)-C4), and among all placement choices that satisfying C1)-C4), this placement maximized $\underline{C}$.

With the number of training symbols being fixed as $(N_t(L+1)+L)$ per sub-block, the optimal number of training symbols is uniquely determined by the number of sub-blocks P. Differentiating $\underline{C}$ with respect to P, we observe that $\underline{C}$ decreases monotonically as P increases. Moreover, notice that there are $N_t(Q+1)(L+1)$ unknowns for each receive antenna, we also need $N_b \geq PL + N_t(Q+1)(L+1)$ to guarantee the separation and channel estimation. The following result can then be established:

Proposition 2 *Suppose* A1)-A3) *holds true. For fixed $\mathcal{P}_c$ and $\mathcal{P}_b$, the placement satisfying* C1-C4*, and with block length N being an integer multiple of $(Q+1)$, then the optimal number of sub-blocks is $P = Q+1$.*

From Propositions 1 and 2, the total number of training symbols is $(N_t(L+1)+L)(Q+1)$. But notice that this includes also the L trailing zeros for IBI removal. The last parameter that remains to be determined is the power allocation factor $\alpha := \mathcal{P}_c/\mathcal{P} \in (0,1)$. To this end, we have established:

Proposition 3 *Suppose* A1-A3 *holds true. For placement satisfying* C1-C4 *with N_c information and N_b training symbols per block, the optimal power allocation factor is given by:*

$$\alpha = \frac{\sqrt{N_c}}{\sqrt{N_c} + \sqrt{N_t(Q+1)(L+1)}},$$

at high signal-to-noise ratio (SNR).

Summarizing, the optimal training parameters are as follows:

Parameters	Optimal training
Number of sub-blocks	$P = Q+1$
Placement of info. symbols	$N_{c,p} = N_c/P$, $\forall\, p \in [1, P]$
Number of training symbols	$N_t(L+1)+L$ per sub-block
Structure of training sub-blocks	$[\mathbf{0}_{\mu(L+1)-1}, b, \mathbf{0}_{(N_t-\mu)(L+1)+L}]$
Power allocation	$\alpha = \frac{\sqrt{N_c}}{\sqrt{N_c}+\sqrt{N_t(Q+1)(L+1)}}$

6. NUMERICAL EXAMPLES

In this section, we present simulations to test our designs. Adopting our training parameters summarized in the table, we depict the average capacity bounds versus SNR in Figure 2. The SNR is defined as the average received symbol power to noise ratio at each receive antenna. We select $L = 3$ for each channel, the transmitted block length $N = 126$, the carrier frequency 2GHz, mobile velocity 160Km/hr, and sampling period 26μs which result in $Q = 2$ bases per channel. We simulate the channel taps as independent and identically distributed across antennas, following an exponentially decaying power delay profile in the time domain and the Jakes' spectrum in frequency domain. As expected, the capacity bounds increase monotonically as SNR increases while the upper and lower bounds are tight. We also note that the $(N_t, N_r) = (1,2)$ case has larger average capacity than $(N_t, N_r) = (2,1)$ does, because we fix the transmission power.

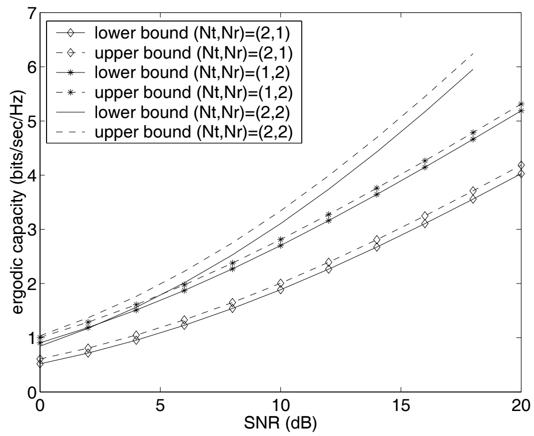

Fig. 2. Average capacity bounds with different number of antennas

However, the capacity bounds for these two cases are parallel in the figure (have the same slope). When $(N_t, N_r) = (2,2)$, the capacity bounds have sharper slopes. This result is consistent with Foschini's claim on coherent average capacity in [4].

7. REFERENCES

[1] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for frequency-selective block-fading channels," *IEEE Trans. on Information 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 Tech.*, pp. 686–693, Nov. 1991.

[3] S. A. Fechtel and H. Meyr, "Optimal parametric feedforward estimation of frequency-selective fading radio channels," *IEEE Trans. on Communications*, Feb./Mar./Apr. 1994.

[4] G. J. Foschini and M. J. Gans, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Communications*, pp. 311–335, Mar. 1998.

[5] G. B. Giannakis and C. Tepedelenlioğlu, "Basis Expansion Models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. of the IEEE*, pp. 1969–1986, Oct. 1998.

[6] B. M. Hochwald and B. Hassibi, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. on Information Theory*, pp. 951–963, Apr. 2003.

[7] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly-selective wireless fading channels," *IEEE Trans. on Signal Processing*, pp. 1351–1366, May 2003.

[8] X. Ma, L. Yang, and G. B. Giannakis, "Optimal training for MIMO frequency-selective fading channels," in *Proc. of Asilomar Conf. on Signals, Systems, and Computers*, 2002, pp. 1107–1111.

[9] S. Ohno and G. B. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," *IEEE Trans. on Communications*, pp. 2113–2123, Dec. 2002.

[10] Y. Zhang, M. P. Fitz, and S. B. Gelfand, "A performance analysis and design of equalization with pilot aided channel estimation," 1997, pp. 720–724.