# OPTIMAL TRAINING FOR BLOCK TRANSMISSIONS OVER DOUBLY-SELECTIVE FADING CHANNELS*

*Xiaoli Ma, Georgios B. Giannakis*

Dept. of ECE, University of Minnesota
Minneapolis, MN 55455, USA

*Shuichi Ohno*

Dept. of Mathematics and Computer Science
Shimane Univ., Shimane, 690-8504, Japan

## ABSTRACT

High data rates give rise to frequency-selective propagation, while carrier frequency-offsets and mobility-induced Doppler shifts introduce time-selectivity in wireless links. To mitigate the resulting time- and frequency-selective (or doubly-selective) channels, an optimal training strategy is designed in this paper for block transmissions over doubly-selective channels, relying on a basis expansion channel model. The optimality in designing our PSAM parameters consists of maximizing a tight lower bound on the average channel capacity, that is also shown to be equivalent to the minimization of the minimum mean-square channel estimation error. Numerical results corroborate our theoretical designs.

## 1. INTRODUCTION

High data rate wireless and mobile links suffer from time- and frequency-selective propagation effects. Mitigating these effects enables efficient transmission over such doubly-selective channels, and has justifiably received increasing attention over the last decade [3]. These fading channels are challenging to mitigate, but once acquired they offer joint multipath-Doppler diversity gains [6]. The quality of channel acquisition has a major impact on the system performance, especially when the fading is fast. Reliable estimation of doubly-selective channels is thus well motivated.

Two classes of methods are available for the receiver to acquire channel state information (CSI): one is based on training symbols that are *a priori* known to the receiver; while the other relies only on the received symbols to acquire CSI blindly. Albeit suboptimal and bandwidth consuming, training methods remain attractive in practice because they decouple symbol detection from channel estimation, which reduces complexity and relaxes the required identifiability conditions [7].

For time-invariant channels, a training sequence is usually sent at the beginning of each transmission burst. But when the channel is time-selective, this preamble-based training method may not work well. This motivates periodic insertion of training symbols during the transmission, which is known as pilot symbol aided modulation (PSAM). PSAM is not only useful for time-selective channels, but also for frequency-selective, and even doubly selective channels [3, 8, 10]. The number and placement of pilots affects not only the quality of CSI acquisition, but also the transmission rate. Within the general class of doubly-selective channels, PSAM has been optimized based on several criteria, but only for special channel models.

*The work in this paper was supported by the NSF Wireless Initiative Grant No. 9979443, NSF Grant No. 0122431, and by the ARL/CTA.

Optimization of PSAM for frequency-selective channels has relied on either average channel capacity bounds [1, 8, 11], or, the Cramér-Rao bound (CRB) of the adopted channel estimator [8]. PSAM for time-selective fading channels has been designed by minimizing the channel mean-square estimation error, and recently by optimizing an average capacity bound [9]. PSAM for time- and frequency-selective channels has been also considered (but not optimized) in [3, 5, 10, 12]. This paper's *objective* is to optimally design PSAM for doubly-selective channels by capitalizing on a parsimonious basis expansion channel model (BEM).

## 2. BEM FOR DOUBLY-SELECTIVE CHANNELS

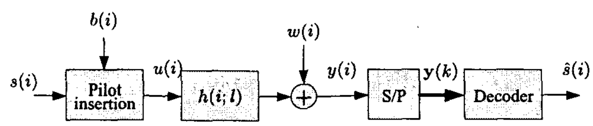


**Fig. 1.** Discrete-time baseband equivalent system model

Figure 1 depicts a general discrete-time baseband equivalent transmission format when communicating through the doubly selective channel with time-varying impulse response $h(i;l)$. Two types of sub-blocks can be identified in each transmitted block: one type contains the information symbols, while the other includes the training (or pilot) symbols. We use two arguments ($n$ and $k$) to describe the serial index $i = kN + n$ for $n \in [0, N-1]$, and denote the $(n+1)$st entry of the $k$th block as $[\mathbf{u}(k)]_n := u(kN+n)$. Each block $\mathbf{u}(k)$ includes $N_s$ information symbols and $N_b$ training symbols, which are known to both transmitter and receiver. After parallel to serial (P/S) multiplexing, the blocks $\mathbf{u}(k)$ are transmitted through a time- and frequency-selective channel. The $i$th received sample can be written as:

$$y(i) = \sum_{l=0}^{L} h(i;l)u(i-l) + w(i), \qquad (1)$$

where $w(i)$ is additive white Gaussian noise (AWGN) with mean zero, and variance $\sigma_w^2$, and the *discrete-time baseband equivalent* channel model is given by (see [7] for detailed derivations):

$$h(i;l) = \sum_{q=0}^{Q} h_q(\lfloor i/N \rfloor; l) e^{j\omega_q (i \bmod N)}, \quad l \in [0, L], \qquad (2)$$

where $\omega_q := 2\pi(q - Q/2)/N$, $L := \lfloor \tau_{\max}/T_s \rfloor$, and $Q := 2\lceil f_{\max} N T_s \rceil$; $\tau_{\max}$ and $f_{\max}$ denote the channel's delay and Doppler spread, respectively. Because both $\tau_{\max}$, and $f_{\max}$ can be measured experimentally in practice, we assume that:

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

When transmissions experience rich scattering, and no line-of-sight is present, one can appeal to the central limit theorem to validate that when $h(i;l)$ is treated as random:

A2) *The BEM coefficients $h_q(\lfloor i/N \rfloor; l)$ are zero-mean, complex Gaussian random variables.*

We will find it convenient to work with a block-form of the BEM which we construct, after serial to parallel (S/P) conversion, by collecting the samples $y(i)$ into $N \times 1$ blocks: $\mathbf{y}(k) = [y(kN), y(kN+1), \ldots, y(kN+N-1)]^T$. Selecting also $N \geq L$, we can write the matrix-vector counterpart of (1) as:

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{u}(k) + \mathbf{H}^{ibi}(k)\mathbf{u}(k-1) + \mathbf{w}(k), \qquad (3)$$

where $[\mathbf{w}(k)]_n := w(kN+n)$, while $\mathbf{H}(k)$ and $\mathbf{H}^{ibi}(k)$ are $N \times N$ upper and lower triangular matrices with entries $[\mathbf{H}(k)]_{n,m} = h(kN+n; n-m)$, and $[\mathbf{H}^{ibi}(k)]_{n,m} = h(kN+n; N+n-m)$ for $n, m = 1, \ldots, N$. The second term in the r.h.s. of (3) captures the interblock interference (IBI) that emerges due to the channel delay-spread. Recalling that $i = kN + n$ in (2), we can rewrite these channel matrices as:

$$\mathbf{H}(k) = \sum_{q=0}^{Q} \mathbf{D}(\omega_q)\mathbf{H}_q(k), \;\; \mathbf{H}^{ibi}(k) = \sum_{q=0}^{Q} \mathbf{D}(\omega_q)\mathbf{H}_q^{ibi}(k) \qquad (4)$$

where $\mathbf{D}(\omega_q) := \text{diag}[1, \ldots, \exp(j\omega_q(N-1))]$, $\mathbf{H}_q(k)$ is a lower triangular, and $\mathbf{H}_q^{ibi}(k)$ is an upper triangular Toeplitz matrix with entries $[\mathbf{H}_q(k)]_{n,m} = h_q(n-m)$, and $[\mathbf{H}_q^{ibi}(k)]_{n,m} = h_q(N+n-m)$, respectively.

In this paper, we wish to design the optimal training input for channel estimation, based on conditional mutual information and channel estimation error criteria. Our joint consideration of these criteria is intuitively appealing because of the apparent tradeoff: using more training symbols of higher power improves channel estimation, but also leads to reduced channel capacity.

## 3. LINKING CHANNEL ESTIMATION WITH CAPACITY

**Channel Estimation:** Since the channel coefficients $h_q(l)$ in (2) are time-invariant over $NT_s$ seconds, channel estimation has to be performed every $N$ symbols. To enable low-complexity block-by-block processing at the receiver, we need to remove the IBI. Hence, we construct $\mathbf{u}(k)$ to satisfy the condition:

C1) *Each block $\mathbf{u}(k)$ has the form $[\bar{\mathbf{u}}^T(k)\ \mathbf{0}_{1\times L}]^T$, where the $(N-L) \times 1$ vector $\bar{\mathbf{u}}(k)$ contains $N_s$ information symbols, and $N_b - L \geq 0$ training symbols.*

Since $\mathbf{H}^{ibi}(k)\mathbf{u}(k-1) = \mathbf{0}$, C1) guarantees the elimination of IBI. Without loss of generality, the placement of these symbols in $\mathbf{u}(k)$ can be expressed as

$$\mathbf{u}(k) = [\mathbf{s}_1^T(k),\ \mathbf{b}_1^T(k),\ \ldots,\ \mathbf{s}_P^T(k),\ \mathbf{b}_P^T(k)]^T, \;\; \forall\, k \qquad (5)$$

where we group consecutive information symbols and training symbols in sub-blocks: $\mathbf{s}_p(k)$ and $\mathbf{b}_p(k)$ of lengths $N_{s,p}$ and $N_{b,p}$, respectively. The structure of $\mathbf{u}(k)$ enables separation of each received block $\mathbf{y}(k)$ into two types of received sub-blocks: one, defined as $\mathbf{y}_b(k)$, that depends only on $\mathbf{H}(k)$ and $\{\mathbf{b}_p(k)\}_{p=1}^{P}$; and a second, defined as $\mathbf{y}_s(k)$, that depends on $\mathbf{H}(k)$, $\{\mathbf{s}_p(k)\}_{p=1}^{P}$, and $\{\mathbf{b}_p(k)\}_{p=1}^{P}$. Because the following analysis is based on a single block, we omit the block index $k$.

Corresponding to the separation of $\mathbf{y}$ to $\mathbf{y}_s$ and $\mathbf{y}_b$, the channel matrix $\mathbf{H}$ can be split into three matrices, namely $\mathbf{H}_s$, $\mathbf{H}_b$ and $\bar{\mathbf{H}}_b$ (see [7] for details). Each of them is constructed from sub-blocks of $\mathbf{H}$. After the separation of $\mathbf{y}$ and by taking C1) into account, we have two input-output relationships:

$$\mathbf{y}_s = \mathbf{H}_s\mathbf{s} + \bar{\mathbf{H}}_b\bar{\mathbf{b}} + \mathbf{w}_s \qquad (6)$$

$$\mathbf{y}_b = \mathbf{H}_b\mathbf{b} + \mathbf{w}_b, \qquad (7)$$

where $\mathbf{s} := [\mathbf{s}_1^T, \ldots, \mathbf{s}_P^T]^T$, $\mathbf{b} := [\mathbf{b}_1^T, \ldots, \mathbf{b}_P^T]^T$, $\bar{\mathbf{b}}$ contains the first $L$ and the last $L$ entries of $\mathbf{b}_p, \forall p$, while $\mathbf{w}_s$ and $\mathbf{w}_b$ denote the corresponding noise vectors. The term $\bar{\mathbf{H}}_b\bar{\mathbf{b}}$ captures the interference of the training sub-blocks to their adjacent information sub-blocks. After interchanging $\mathbf{H}_b$ with $\mathbf{b}$, the input-output relationship in (7) becomes (see [7] for details):

$$\mathbf{y}_b = \mathbf{\Phi}_b\mathbf{h} + \mathbf{w}_b, \qquad (8)$$

where $\mathbf{\Phi}_b$ depends on $\mathbf{b}$ and $\mathbf{D}_q$, and

$$\mathbf{h} := \begin{bmatrix} h_0(0) & \cdots & h_0(L) & \cdots & h_Q(L) \end{bmatrix}^T. \qquad (9)$$

Similar to [8], we will rely on the Wiener solution of (8) that yields the linear[1] MMSE (LMMSE) channel estimator:

$$\hat{\mathbf{h}} = \frac{1}{\sigma_w^2}\left(\mathbf{R}_h^{-1} + \frac{1}{\sigma_w^2}\mathbf{\Phi}_b^{\mathcal{H}}\mathbf{\Phi}_b\right)^{-1}\mathbf{\Phi}_b^{\mathcal{H}}\mathbf{y}_b, \qquad (10)$$

which requires $\mathbf{R}_h := E[\mathbf{h}\mathbf{h}^{\mathcal{H}}]$ to be known at the receiver.

Defining the channel error as $\tilde{\mathbf{h}} := \mathbf{h} - \hat{\mathbf{h}}$, we have:

$$\mathbf{R}_{\tilde{h}} := E[\tilde{\mathbf{h}}\tilde{\mathbf{h}}^{\mathcal{H}}] = \left(\mathbf{R}_h^{-1} + \frac{1}{\sigma_w^2}\mathbf{\Phi}_b^{\mathcal{H}}\mathbf{\Phi}_b\right)^{-1}, \; \sigma_{\tilde{h}}^2 := tr(\mathbf{R}_{\tilde{h}}). \;\; (11)$$

It is clear from (11) that the placement of training symbols affects $\mathbf{\Phi}_b$, and consequently $\sigma_{\tilde{h}}^2$.

**Capacity Bounds:** Let $\mathcal{P}_s$ denote the transmit-power allocated to the information signal part. Suppose first that the channel estimate $\hat{\mathbf{H}}$ is perfect; i.e., $\hat{\mathbf{H}} \equiv \mathbf{H}$. It can be verified that for fixed $\mathcal{P}_x$, the average channel capacity in this case is given by:

$$\overline{C} := \max_{P_s(\cdot)} \frac{1}{N} E\left[\log\det\left(\mathbf{I}_{N_s+LP} + \frac{1}{\sigma_w^2}\mathbf{H}_s\mathbf{R}_s\mathbf{H}_s^{\mathcal{H}}\right)\right], \quad (12)$$

where $P_s(\cdot)$ denotes the probability density function of $\mathbf{s}$. We underscore that $\overline{C}$ is an upper bound on the average channel capacity with estimated channels, because it expresses the ideal channel capacity without channel estimation error.

Consider now that the estimate of $\mathbf{H}$ is imperfect. Define $\tilde{\mathbf{H}}_s := \mathbf{H}_s - \hat{\mathbf{H}}_s$, $\tilde{\mathbf{H}}_b := \bar{\mathbf{H}}_b - \hat{\bar{\mathbf{H}}}_b$, and $\mathbf{v} := \tilde{\mathbf{H}}_s\mathbf{s} + \tilde{\mathbf{H}}_b\bar{\mathbf{b}} + \mathbf{w}_s$. In general, $\mathbf{v}$ is non-Gaussian distributed with correlation matrix $\mathbf{R}_v := E[\mathbf{v}\mathbf{v}^{\mathcal{H}}]$ given by:

$$\mathbf{R}_v = \bar{\mathcal{P}}_s E[\tilde{\mathbf{H}}_s\tilde{\mathbf{H}}_s^{\mathcal{H}}] + E[\tilde{\mathbf{H}}_b\bar{\mathbf{b}}\bar{\mathbf{b}}^{\mathcal{H}}\tilde{\mathbf{H}}_b^{\mathcal{H}}] + \sigma_w^2\mathbf{I}_{N_s+LP}. \; (13)$$

Since $\mathbf{v}$ is uncorrelated with $\mathbf{s}$, it follows from [8, Lemma 2] that the worst case noise is zero mean Gaussian with auto-correlation $\mathbf{R}_v$, and with $\mathbf{v}$ independent of $\mathbf{s}$. Although $\mathbf{s}$ is generally non-Gaussian, if $N_s$ is sufficiently large and $\mathbf{s}$ is channel coded (or linearly precoded), then $\mathbf{s}$ can be approximated as Gaussian; i.e.,

A3) *The information bearing symbol block $\mathbf{s}$ is zero-mean Gaussian with covariance $\mathbf{R}_s = \bar{\mathcal{P}}_s\mathbf{I}_{N_s}$, and $\bar{\mathcal{P}}_s := \mathcal{P}_s/N_s$.*

Taking $\mathbf{R}_s = \bar{\mathcal{P}}_s\mathbf{I}_{N_s}$ into account, we obtain the lower bound on the realistic average capacity $C$ as (see [7] for a proof):

$$C \geq \frac{1}{N}E[\log\det\{\mathbf{I}_{N_s+LP} + \bar{\mathcal{P}}_s\mathbf{R}_v^{-1}\hat{\mathbf{H}}_s\hat{\mathbf{H}}_s^{\mathcal{H}}\}] := \underline{C}. \quad (14)$$

[1] Under A2), $\mathbf{h}$ is Gaussian in the linear model (8); hence, the LMMSE coincides with MMSE optimal channel estimator.

The r.h.s. of (14) offers a lower bound on the average capacity of doubly-selective channels. The link between $\underline{C}$ and the channel MMSE $\sigma_{\tilde{h}}^2$ is established in the following proposition [7]:

**Proposition 1** *Suppose A1)–A3) hold true. If $N_{s,p} \gg 2L$, $\forall p$, then for fixed $N_{s,p}$ and $N_{b,p}$, the minimization of $\sigma_{\tilde{h}}^2$ in (11) is equivalent to the maximization of $\underline{C}$ in (14), at high SNR.*

## 4. OPTIMAL TRAINING PARAMETERS

Based on the link between the LMMSE channel estimation with the lower bound of the average channel capacity, we design our optimal training parameters as follows:

**Proposition 2** *Suppose A1)–A3) hold true. The following placement is optimal: all information sub-blocks have identical block lengths; i.e., $N_{s,p} = \bar{N}_s, \forall p$; the pilot sub-blocks have identical structure $[\mathbf{0}_L^T\ b\ \mathbf{0}_L^T]^T$, $\forall p$, and they are equi-powered with $b = \bar{\mathcal{P}}_b := \mathcal{P}_b/P$. The optimal number of sub-blocks is $P = Q + 1$, and with $\mathcal{P}_s := \alpha\mathcal{P}$, the optimal power allocation factor is:*

$$\alpha_{opt} = \frac{1}{1 + ((L+1)/\bar{N}_s)^{1/2}}. \quad (15)$$

Proposition 2 finalizes the *optimal structure* of our transmitted block $\mathbf{u}$ as

$$\mathbf{u} = [\mathbf{s}_0^T\ \mathbf{0}_L^T\ b\ \mathbf{0}_L^T\ \cdots\ \mathbf{s}_Q^T\ \mathbf{0}_L^T\ b\ \mathbf{0}_L^T]^T, \quad b = \sqrt{\bar{\mathcal{P}}_b}, \quad (16)$$

with $\mathcal{P}_b = (1-\alpha)\mathcal{P}$, and $\alpha$ given in (15).

## 5. SPECIAL CASES AND SAMPLING INTERPRETATION

**Frequency-Selective Channels:** Frequency-selective channels exhibit no (or negligible) variation during each transmitted block, and correspond to setting $Q = 0$ in (2). Hence, the optimum number of sub-blocks is $Q + 1 = 1$, and the transmitted block $\mathbf{u}$ in (16) reduces to

$$\mathbf{u} = [\mathbf{s}^T\ \mathbf{0}_L^T\ b\ \mathbf{0}_L^T]^T, \quad b := \sqrt{\bar{\mathcal{P}}_b}, \quad \mathcal{P}_b = (1-\alpha)\mathcal{P}, \quad (17)$$

where we removed the sub-script $_p$ for obvious reasons. Notice that $\mathbf{u}$ in (17) has the same structure as the design in [1, Theorem 3], which implies that [1] is subsumed by our design for doubly-selective channels. On the other hand, [8] formed $\mathbf{u}$ as an affine mapping of $\mathbf{s}$ and $\mathbf{b}$. The transmission in (17) can also be written in such an affine form [7].

Our main difference with [8, 1] is that a cyclic prefix (CP) is employed in [8, 1] to eliminate IBI, while we use zero-padding (ZP). It is interesting that the optimal number of redundant symbols is $2L + 1$ for both ZP- and CP-based training designs. Furthermore, although the power allocation parameter $\alpha$ in (15) and in [1] are identical, they mean different things. Due to the CP, $\alpha$ in [1] corresponds to the effective information power over the "total" power that excludes the CP. However, in our setup, $\alpha$ corresponds to the ratio of signal power over the total power per block, since we use ZP instead of CP to eliminate IBI. So, for a fixed total power per block, our ZP-scheme results in higher effective $\mathcal{P}_s$ and $\mathcal{P}_b$ than the CP-scheme. In the simulations section, we will further re-inforce this point.

**Time-Selective Channels:** In time-selective channels, the delay spread can be ignored, and the channel order $L = 0$ must be set in (2). In this case, the transmitted block $\mathbf{u}$ in (16) becomes

$$\mathbf{u} = [\mathbf{s}_0^T\ b\ \mathbf{s}_2^T\ b\ \cdots\ \mathbf{s}_Q^T\ b]^T, \quad b := \sqrt{\bar{\mathcal{P}}_b}, \quad \mathcal{P}_b = (1-\alpha)\mathcal{P}. \quad (18)$$

Eq. (18) coincides with the results in [2, 9]. Comparing (18) with [1, 8], we can observe the duality between periodic insertion of pilots tones in OFDM for frequency-selective channels, and the PSAM for time-selective channels [2]. There is, however, a notable difference between our scheme in (18), and the optimal design in [9]. In [9], the optimal distance between two consecutive pilots is $\lfloor 1/(2f_{\max}T_s) \rfloor$, where $\lfloor \cdot \rfloor$ denotes integer floor. In contrast, we find the optimal number of pilot symbols per superblock to be $(Q + 1)$ since we adopt the BEM as our channel model. In [7], we prove that these two conditions are equivalent. Since [9] obtained this optimal distance based on a general time-varying channel model, while we started from the BEM, the equivalence that we just mentioned, corroborates also the validity of our BEM.

**Time-Frequency Sampling Interpretations:** For time-selective channels, it is well known that the optimal PSAM samples uniformly the channel in the time-domain via periodic insertion of pilot symbols $b$ [2, 9]. Indeed, starting from the scalar input-output relationship for the training samples, $y_b(i) = h(i)b + w_b(i)$, one can estimate the channel as: $\hat{h}(i) = y_b(i)/b$. In a dual fashion, for frequency-selective channels, optimal PSAM with cyclic prefix samples uniformly the channel in the frequency-domain via periodic insertion of pilot tones $\tilde{b}$ [8, 1].

For doubly-selective channels, we can view the BEM coefficient $h_q(l)$ in (2) as the two-dimensional (2-D) channel sample at the ($q$th frequency bin, $l$th lag or time-slot). Intuitively thinking, the Kronecker deltas in (16) surrounded by zero-guards, implement time-domain sampling with pilot symbols; furthermore, the fact that these deltas are periodically inserted, implies that they are also equivalent to Kronecker deltas in the frequency-domain, and thus serve as pilot tones as well. To solidify this intuition, we will rearrange $\mathbf{y}_b$ in a 2-D fashion. Let us select the $Q+1$ entries from $\mathbf{y}_b$ in (8) with indices $\{(q+1)(l+1)\}_{q=0}^{Q}$ for a fixed lag $l$, defined as $\mathbf{y}^b(l)$, and then concatenate these vectors with $l \in [0, L]$, to form the $L + 1$ columns of the $(Q + 1) \times (L + 1)$ matrix $\mathbf{Y}_b := [\mathbf{y}^b(0)\ \mathbf{y}^b(1) \cdots \mathbf{y}^b(L)]$. Notice that the matrix $\mathbf{Y}_b$ contains all the training-based received data from $\mathbf{y}_b$ in (8) arranged in a 2-D format. If $\tilde{\mathbf{Y}}_b := \mathbf{F}_{Q+1}\mathbf{Y}_b/(Q+1)$ denotes the FFT of this 2-D received data array, where $\mathbf{F}_{Q+1}$ denotes the $(Q + 1)$-point FFT matrix with entries $[\exp(-j2\pi(m-1)(n-1)/(Q+1))]_{m,n}$, we can express the training input-output relationship after FFT processing as [7]:

$$\tilde{y}_q^b(l) = \tilde{b}_q(l)h_q(l) + \tilde{w}_q^b(l), \quad \tilde{b}_q(l) := \sqrt{\bar{\mathcal{P}}_b}e^{j\omega_q(\bar{N}_s+L+l)}. \quad (19)$$

Eq. (19) proves that indeed our optimal PSAM samples the BEM in time-frequency to enable estimation of the doubly-selective channel via: $\hat{h}_q(l) = \tilde{y}_q^b(l)/\tilde{b}_q(l)$. In fact, our optimal training sequence in (16), is precisely what one needs to obtain the channel model that is assumed *a fortiori* in [5].

## 6. NUMERICAL EXAMPLES

We now present test cases to validate our analysis and design.

**Test Case 1 (Optimal PSAM parameters):** The transmitted block size is $N = 63$, the number of information symbols $N_s = 42$, and the modulation is QPSK. The doubly-selective channel model is generated using the following parameters: carrier frequency $f_0 =$ 2GHz, sampling period $T_s = 53.6\mu$s, and mobile speed $v_{\max} =$ 160km/hr. With these parameters, we find that $Q = 2$. Our channel order is $L = 3$. All the channel coefficients $h_q(l)$ are generated as independent, standardized, complex Gaussian random with an exponential power profile (see [7] for details). Two parameters will be tested in this example. The first one is the number of the non-zero pilot symbols $N_{\bar{b},p}$. We let $N_{\bar{b},p} = N_{\bar{b}}, \forall p$, and adopt all the other parameters in Proposition 2 while changing $N_{\bar{b}}$. Figure 2 (left) depicts the lower bound on the average capacity (14) vs. $N_{\bar{b}}$. It can be seen that the capacity bound decreases monotonically as $N_{\bar{b}}$ increases for each SNR value considered (0dB, 10dB and 20dB). Furthermore, we notice that as the SNR increases, the effect of $N_{\bar{b}}$ increases. We depict the lower bound on the average capacity vs. $\alpha$ in Figure 2 (right). When $\alpha$ is too small (near 0), the average capacity is small since the information symbols do not have enough power to combat AWGN. When $\alpha$ is too large (near 1), the average capacity is also small since the training symbols do not have enough power to provide reliable channel estimation. The optimum $\alpha$ found from the peak confirms eq. (15).

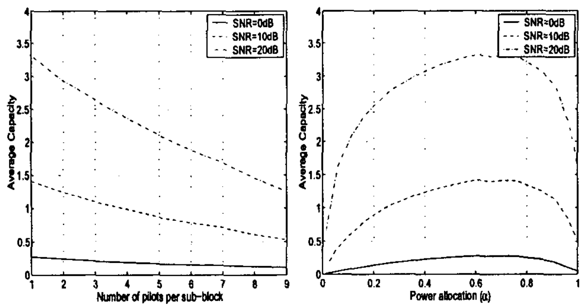


**Fig. 2.** Optimal PSAM parameters

**Test Case 2 (Comparison of ZP in** (17) **with CP in [1, 8]):** This test case is designed to compare our scheme in Section VI-A with [1, 8]. The channel is frequency-selective with i.i.d. taps. The channel order $L = 7$, and each tap is a zero mean Gaussian random variable with variance $1/(L+1)$. The number of information symbols per block is $\bar{N}_s = 48$, and the block length $N = \bar{N}_s + 2L + 1$. The total power per block is fixed to $\mathcal{P}$. Hence, the power ratio allocated between information symbols and training symbols for the CP-based scheme, is $\mathcal{P}(\bar{N}_s + L + 1)/N$. Figure 3 (left) depicts the average capacity bounds for both ZP- and CP-based alternatives. Here $SNR := \mathcal{P}/(\bar{N}_s + 1)$. We notice that the bounds (either upper or lower) for ZP are consistently greater than those of CP, which is partially due to the power loss incurred by the CP. We plot BER vs. SNR in Figure 3 (right). In the same figure, the ideal cases corresponding to perfect channel estimates are also plotted as benchmarks (the dashed lines). We computed MMSE channel estimates based on pilot symbols, and used zero-forcing (ZF) equalization for symbol detection in both cases. From Figure 3 (right), we observe that: i) ZP outperforms CP at high SNR, while CP has about 2dB advantage at BER= 0.1; ii) from the slopes of the curves, we notice that CP offers lower diversity order than ZP; and iii) for both cases, the penalty for inaccurate channel state information is about 1.5 dB.

In a nutshell, we showed that the optimal training for doubly-selective channels consists of equi-spaced and equi-powered pilot symbols surrounded by a number of zeros dictated by the channel's delay-spread, and inserted periodically with a period dictated by the channel's Doppler-spread.

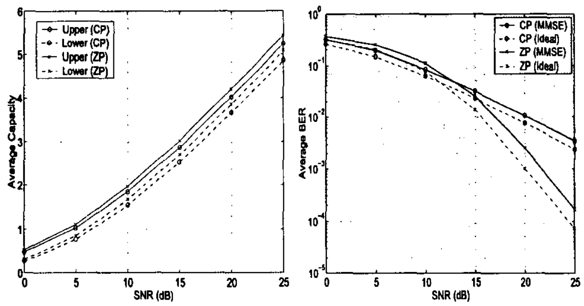


**Fig. 3.** CP vs. ZP for training frequency-selective channels

## 7. REFERENCES


[1] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal Placement of Training for Unknown Channels," *IEEE Trans. on Information Theory*, submitted Mar. 2001; URL: *http://www.ee.cornell.edu/~ltong/pubJ.html*

[2] J. K. Cavers, "Pilot Symbol Assisted Modulation and Differential Detection in Fading and Delay Spread," *IEEE Trans. on Comm.*, vol. 43, no. 7, pp. 2206–2212, July 1995.

[3] S. A. Fechtel and H. Meyr, "Optimal Parametric Feedforward Estimation of Frequency-Selective Fading Radio Channels," *IEEE Trans. on Comm.*, pp. 1639–1650, 1994.

[4] G. B. Giannakis and C. Tepedelenlioğlu, "Basis Expansion Models and Diversity Techniques for Blind Identification and Equalization of Time-Varying Channels," *Proceedings of the IEEE*, pp. 1969-1986, Nov. 1998.

[5] P. Höeher, S. Kaiser and P. Robertson, "Two-Dimensional Pilot-Symbol-Aided Channel Estimation by Wiener Filtering," *Proc. of ICASSP*, Münich, Germany, vol. 3, pp. 1845–1848, 1997.

[6] X. Ma and G. B. Giannakis, "Maximum-Diversity Transmissions over Doubly-Selective Wireless Channels," *IEEE Trans. on Inform. Theory*, submitted Aug. 2001; URL: *http://spincom.ece.umn.edu/xiaoli/double-diversity.pdf*

[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*, submitted Oct. 2001; URL: *http://spincom.ece.umn.edu/xiaoli/double-training.pdf*

[8] S. Ohno and G. B. Giannakis, "Capacity Maximizing Pilots for Wireless OFDM over Rapidly Fading Channels," *Proc. of Intl. Symp. on Signals, Systems and Electronics*, pp. 246–249, Tokyo, Japan, July 24-27, 2001.

[9] S. Ohno and G. B. Giannakis, "Average-Rate Optimal PSAM Transmissions over Time-Selective Fading Channels," *IEEE Trans. on Wireless Comm.*, submitted July 2001.

[10] M. K. Tsatsanis and Z. Xu, "Pilot Symbol Assisted Modulation in Frequency Selective Fading Wireless Channels," *IEEE Trans. on Signal Processing*, vol. 48, no. 8, pp. 2353–2365, Aug. 2000.

[11] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "Optimal Training for Frequency-Selective Fading Channels," *Proc. of ICASSP*, Salt Lake City, Utah, vol. 4, pp. 2105–2108, May 7–11, 2001.

[12] Y. Zhang, M.P. Fitz, and S. B. Gelfand, "A Performance Analysis and Design of Equalization with Pilot Aided Channel Estimation," *Proc. of 47th VTC*, vol. 2, pp. 720–724, Phoenix AZ, May 4–7, 1997.