Optimal Placement of Training for Unknown Channels

Srihari Adireddy, Lang Tong
School of Electrical Engineering
Cornell University
Ithaca, NY 14853
{srihari,ltong}@ee.cornell.edu

Harish Viswanathan
Lucent Technologies Bell Labs
Murray Hill, NJ
harishv@lucent.com

Abstract — **The problem of placing training symbols optimally for Single-Carrier Systems and OFDM is considered. The channel is assumed to be quasi-static with a finite impulse response of length $(L+1)$ samples. Under the assumptions that neither the transmitter nor the receiver knows the channel, and that the receiver forms a minimum mean square (MMSE) channel estimate based on training symbols only, we optimize the training by maximizing a tight lower bound of the mutual information. For Single-Carrier Systems, under the assumption that the training symbols are placed in clusters of length $\alpha \geq (2L+1)$, we show that the placements belonging to the family of QPP-α are optimal. For OFDM systems, we show that the lower bound is maximized by placing the known symbols periodically in frequency.**

I. Introduction

The problem of achieving the capacity of a linear, time-invariant Gaussian channel under the assumption that both the transmitter and the receiver know the channel is classical ([4] and the references in it). For wireless communications, especially mobile wireless, the channel is random and time-varying. Hence the assumption that either receiver or transmitter know the channel is unrealistic [2]. The rapid growth in mobile wireless applications has motivated the problem of finding and achieving the capacity of a fading channel under the assumption that neither the receiver nor the transmitter knows the channel (unknown channel).

The most popular and practical technique of learning the channel is by insertion of training symbols in the data stream. While insertion of known symbols leads to loss in information, it is mandatory in order to simplify the receiver implementation. The question then arises how one should optimize training to maximize the capacity of a training based system on a mobile wireless channel. This problem was considered for a multiple antenna system under Rayleigh block flat fading scenario by Hassibi and Hochwald [1]. They obtained tight lower bounds on the capacity of the training-based systems and optimized the fraction of training symbols, energy allocated to training and data to maximize this bound. Their paper provides a useful framework for analyzing the capacity achievable by training based schemes in general. An important insight of this analysis is that training is optimal at high SNR and suboptimal at low SNR. Similar techniques for lower bounding mutual information under imperfect knowledge of the channel have been proposed by Medard [5].

Demand for higher bit rate leads to frequency selective fading in mobile wireless channels. This motivates the question of designing training for frequency selective fading channels with block fading. A new degree of freedom that is specific to frequency selective channels is the placement of training. The performance for the flat fading scenario is independent of the placement of known symbols. Furthermore the problem of training symbol placement has to be addressed for both single-carrier and multi-carrier systems separately since the paradigm for training is different for the two transmission systems.

For single-carrier systems, the design of training, namely, the fraction of training, the choice of training symbols and energy trade-off between training and data, for frequency selective fading model was addressed in [3] under the assumption that all the training symbols are placed at the start of the packet. It was shown that at high SNR training-based schemes are capable of capturing most of the channel capacity, whereas at low SNR they are highly suboptimal. The placement of training though was assumed to be fixed.

The placement of training affects the capacity of the system through channel estimation and detection. We have previously considered the problem of joint optimization of symbol placement and equalizer for a symbol-by-symbol decision feedback receiver [7] under the assumption that the channel is known at the receiver. The performance criterion used was Average Mean Square Error (AMSE). It turns out that the optimal symbol placement is to separate the known symbols by at least the detection delay d of the decision feedback receiver. In order to obtain placement schemes that are optimal independent of the receiver, we have considered optimizing placement of training with respect to the i.i.d capacity of the channel under the assumption that the channel is linear, time-invariant and known to the receiver [8]. The channel was assumed to be known since the objective was to determine the effect of known symbol placement on detection. It was shown that in the single-carrier case, mutual information is maximized by breaking the known symbols into small blocks and placing them periodically. Optimal placement schemes were also obtained for the case when the frame length goes to infinity.

The problem of optimizing placement of training for minimizing the mean square error in channel estimate has been addressed for OFDM systems in [6]. The metric used for optimizing the placement is the mean square error (MSE) of the channel estimate. Optimal training placement schemes were obtained for the more general setting of block precoded transmissions with cyclic prefix in [12]. The metric for optimization was again MSE of the channel estimate. However, as alluded to earlier, channel estimation is just one facet of the problem. The placement of known symbols affects not only the chan-

*Part of this work was done when Srihari Adireddy was visiting Lucent Technologies. This work was supported in part by the National Science Foundation under Contract CCR-9804019 and the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564.

nel estimate but also the detection. In this paper, we take the holistic view and try to optimize the placement of known symbols by maximizing the mutual information.

We use the framework provided by [1] to optimize the design of training including placement, by maximizing a tight lower bound on the training-based capacity of OFDM and single-carrier systems. For single-carrier systems, under the assumption that the training symbol clusters are of length at least $\alpha \geq (2L+1)$, we show that the placement schemes in the class QPP-α [8] are optimal. The placement schemes in QPP-α are obtained by breaking the known symbols into as many clusters as possible and placing them such that the unknown symbols blocks are as "equal" as possible. For OFDM systems, under the assumption that all the training symbols have equal energy, the mutual information is maximized by placing the training symbols periodically in the OFDM symbol [9]. That is, we pick equally spaced tones for training. This is the placement scheme that was also obtained in [6, 12]. It is remarkable that this placement not only gives the best channel estimate but also maximizes the tight lower bound on mutual information.

This paper is organized as follows. In Section II, we introduce the system model. In Section III we first formulate the optimization problem for single-carrier systems and then determine optimal placement schemes. We give the results for OFDM systems in Section IV. In Section V we illustrate the ideas through simulations and finally conclude in Section VI.

II. System Model

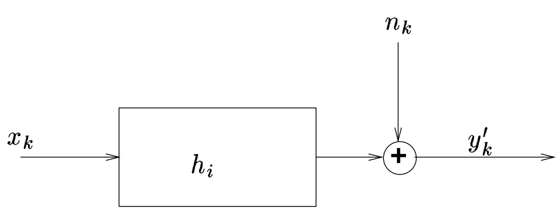

Fig. 1: System Model

The system model is shown in Fig 1. The channel $\mathbf{h} = [h_0, h_1, \cdots h_L]^t$ has a finite impulse response of length $(L+1)$ samples. We assume that taps of the channel $\mathbf{h}$ are i.i.d complex Gaussian with zero mean and variance equal to $\frac{1}{L+1}$. The fading coefficients remain constant for T symbol periods and change to an independent value. We assume that neither the receiver nor the transmitter knows the fading coefficients. The received signal is corrupted by additive white noise that is complex Gaussian with zero mean and variance σ_w^2. This model described above is an extension of the quasi-static flat fading to quasi-static frequency selective fading.

III. Optimal Placement for Single-Carrier Systems

A Single-Carrier System

Fig. 2 shows the processing performed at the transmitter of the single carrier system. We assume that the symbols are parsed into packets of length $(T-L)$ by the serial to parallel converters. A known symbol cluster $\mathbf{s}_k$ of length L is appended to the beginning of each block to form a super block. These known symbol clusters serve to remove the inter block interference (IBI) between consecutive blocks and facilitate block

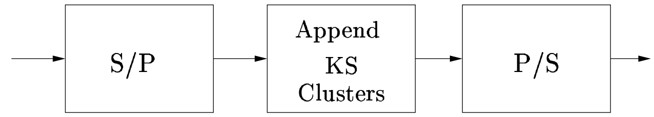

Fig. 2: Processing Performed at the Single-Carrier Transmitter

by block processing. A parallel to serial conversion is then performed on these super blocks and they are then transmitted through the channel. As shown in Fig. 3, we assume that

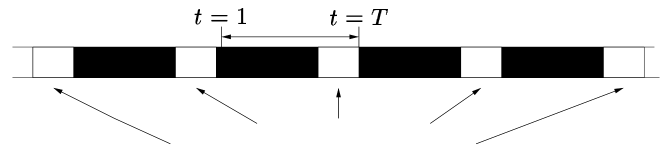

Fig. 3: The Period over which the Channel Stays Constant

the channel stays constant from $t = 1$ to $t = T$. Over the period for which the channel stays constant we have

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} h_L & \cdots & h_0 & & & \\ & h_L & \cdots & h_0 & & \\ & & \ddots & & \ddots & \\ & & & h_L & \cdots & h_0 \end{bmatrix}}_{\mathbf{H}} \begin{bmatrix} \mathbf{s}_k \\ s_1 \\ \vdots \\ s_{N+P} \\ \mathbf{s}_k \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_T \end{bmatrix}. \tag{1}$$

where $\mathbf{h} = [h_0, \cdots, h_L]^t$ is a realization of the channel. We note that the output vector $\mathbf{y}$ is a function of both the symbols in the current packet $\mathbf{s} = [\mathbf{s}_k, s_1, \cdots, s_{T-L}]^t$ and the known symbol cluster $\mathbf{s}_k$ at the start of the next packet.

Each packet $\mathbf{s}$ consists of N unknown and $(P+L)$ known symbols. The known symbols are placed in clusters of length equal to $\alpha \geq L$. Fig. 4 shows a placement scheme of the vector

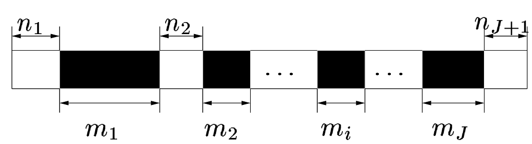

Fig. 4: Representation of Placement Schemes

$[\mathbf{s}\ \mathbf{s}_k]^t$. In general every placement can be specified by two tuples $(\mathbf{m}, \mathbf{n})$ where $\mathbf{m} = (m_1, \cdots, m_J)$ and $\mathbf{n} = (n_1, \cdots, n_{J+1})$. The tuple $\mathbf{m}$ gives the lengths of unknown symbol blocks and $\mathbf{n}$ gives the lengths of known symbol clusters. Since every packet starts with at least L known symbols, we know that n_1 is at least as big as L. We also note that the known symbol cluster $J+1$ includes the first L known symbols at the start of the next packet. Hence n_{J+1} is also at least as big as L. The minimum value of J is equal to one, which corresponds to placing all the training at the ends of the packet. We note that the number of elements in each tuple is a function of the placement scheme. We refer to the symbols between any two consecutive known symbol clusters as unknown symbol blocks. Let the set $\mathcal{P}$ be the set of all possible placement schemes $(\mathbf{m}, \mathbf{n})$.

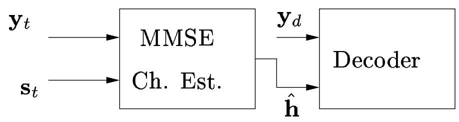

Fig. 5: Receiver Structure

As shown in Fig. 5, the receiver consists of a channel estimator block followed by a decoder. The channel estimator forms an estimate of the channel based on training only. Since the channel varies from block to block, we can only form a block-by-block estimate of the channel. If s_{it}^k denotes the k^{th} training symbol in the i^{th} cluster, then we define the vector of training symbols $\mathbf{s}_t = [s_{1t}^1 \cdots s_{1t}^{n_1} \cdots s_{(J+1)t}^1 \cdots s_{(J+1)t}^{n_{J+1}}]^t$. We note again that $\mathbf{s}_k = [s_{1t}^1 \cdots s_{1t}^L]^t = [s_{(J+1)t}^{n_{J+1}-L+1} \cdots s_{(J+1)t}^{n_{J+1}}]^t$. We define as $\mathbf{y}_t$ the part of the output vector $\mathbf{y}$ that is due to training alone. The remaining part of the output vector is grouped as $\mathbf{y}_d$. The channel estimator block forms the estimate of the channel $\hat{\mathbf{h}} = g(\mathbf{y}_t, \mathbf{s}_t)$. The decoder uses $\mathbf{y}_d, \hat{\mathbf{h}}$ and $\mathbf{s}_t$ optimally to perform the decoding.

We define as $\mathbf{s}_d$ the vector containing all the data symbols. The power constraint on the system is formulated as following.

$$\frac{1}{(N+P+L)}(\mathrm{E}\{\,\mathrm{tr}\,\mathbf{s}_d\mathbf{s}_d^H\} + \,\mathrm{tr}\,\mathbf{s}_t\mathbf{s}_t^H) = 1. \tag{2}$$

We do not constrain the data and training powers to be same. If $\rho_d = \frac{1}{N}\mathrm{E}\{\,\mathrm{tr}\,\mathbf{s}_d\mathbf{s}_d^H\}$ and $\rho_t = \frac{1}{P+L}\,\mathrm{tr}\,\mathbf{s}_t\mathbf{s}_t^H$, then the above equation can be written as

$$\frac{N\rho_d + (P+L)\rho_t}{N+P+L} = 1. \tag{3}$$

B Problem Statement

We now formulate the problem of optimal placement of training for single-carrier systems. The i.i.d capacity of the system [11] can be defined as

$$C(\mathcal{P}, \rho_d, \rho_t, \mathbf{s}_t) \triangleq \max_{f_{i.i.d}(\mathbf{s}_d)} I(\mathbf{y}_d, \hat{\mathbf{h}}; \mathbf{s}_d) \tag{4}$$

$$= \max_{f_{i.i.d}(\mathbf{s}_d)} I(\mathbf{y}_d; \mathbf{s}_d|\hat{\mathbf{h}}). \tag{5}$$

where the probability distribution $f_{i.i.d}(\mathbf{s}_d)$ and the training $\mathbf{s}_t$ are such that the input power constraint is satisfied. Our objective then is to obtain the optimal placement scheme $\mathcal{P}^*$, optimal energy trade-off (ρ_d^*, ρ_t^*) and optimal training symbols $\mathbf{s}_t^*$ as

$$(\mathcal{P}^*, \rho_d^*, \rho_t^*, \mathbf{s}_t^*) = \arg \max_{\mathcal{P}, \rho_d, \rho_t, \mathbf{s}_t} C(\mathcal{P}, \rho_d, \rho_t, \mathbf{s}_t). \tag{6}$$

C Lower Bound on Training-Based Capacity

In this section, since the problem of evaluating the i.i.d capacity is complicated, we obtain a tight lower bound on $C(\mathcal{P}, \rho_d, \rho_t, \mathbf{s}_t)$ and optimize training for this bound. The relationship between $\mathbf{y}_d$ and $\mathbf{s}_d$ is given by

$$\underbrace{\begin{bmatrix} \mathbf{y}_{1d} \\ \vdots \\ \mathbf{y}_{Jd} \end{bmatrix}}_{\mathbf{y}_d} = \underbrace{\begin{bmatrix} \mathbf{H}_{m_1} & \cdots & 0 \\ \vdots & \ddots & \\ 0 & & \mathbf{H}_{m_J} \end{bmatrix}}_{\mathbf{H}_d} \underbrace{\begin{bmatrix} \mathbf{s}_{1d} \\ \vdots \\ \mathbf{s}_{Jd} \end{bmatrix}}_{\mathbf{s}_d} + \underbrace{\begin{bmatrix} \mathbf{T}_1 \\ \vdots \\ \mathbf{T}_J \end{bmatrix}}_{\mathbf{T}} \mathbf{h} + \mathbf{w}_d. \tag{7}$$

The matrix $\mathbf{H}_{m_i}$ is a Toeplitz matrix of size $(m_i + L) \times m_i$ given by

$$\mathbf{H}_{m_i} = \begin{bmatrix} h_0 & 0 & \cdots & 0 \\ h_1 & h_0 & & \vdots \\ \vdots & h_1 & & 0 \\ h_L & \vdots & & h_0 \\ 0 & h_L & & h_1 \\ \vdots & 0 & & \vdots \\ 0 & \cdots & \cdots & h_L \end{bmatrix}_{(m_i+L)\times(m_i)} . \tag{8}$$

The fact that each training symbol cluster is at least as long as L leads to the matrix $\mathbf{H}_d$ being block diagonal with $\mathbf{H}_{m_i}$ having the structure shown above. The vector $\mathbf{s}_{id}$ is of length m_i and is composed of the data symbols in i^{th} unknown symbol block. The matrix $\mathbf{T}_i$ is a composed of the training symbols $(s_{it}^{n_i-L+1}, \cdots, s_{it}^{n_i})$ and $(s_{(i+1)t}^1, \cdots, s_{(i+1)t}^L)$. That is, $\mathbf{T}_i$ is a function of the L training symbols immediately before and after the i^{th} unknown symbol block. These matrices are introduced to account for the fact that the first L and the last L samples of $\mathbf{y}_{id}$ are affected by the training symbols.

We can express $\mathbf{y}_d$ in terms of the estimate $\hat{\mathbf{h}}$ and the error $\tilde{\mathbf{h}}$ as

$$\mathbf{y}_d = \hat{\mathbf{H}}_d\mathbf{s}_d + \mathbf{T}\hat{\mathbf{h}} + \tilde{\mathbf{H}}_d\mathbf{s}_d + \mathbf{T}\tilde{\mathbf{h}} + \mathbf{w}_d. \tag{9}$$

We subtract $\mathbf{T}\hat{\mathbf{h}}$ from $\mathbf{y}_d$ to obtain $\mathbf{y}_d'$. We thus have

$$\mathbf{y}_d' = \hat{\mathbf{H}}_d\mathbf{s}_d + \underbrace{\tilde{\mathbf{H}}_d\mathbf{s}_d + \mathbf{T}\tilde{\mathbf{h}} + \mathbf{w}_d}_{\boldsymbol{\nu}_d}. \tag{10}$$

It is easy to see that

$$I(\mathbf{y}_d; \mathbf{s}_d|\hat{\mathbf{h}}) = I(\mathbf{y}_d'; \mathbf{s}_d|\hat{\mathbf{h}}). \tag{11}$$

But it is difficult to obtain the latter analytically. We obtain a lower bound on the i.i.d channel capacity by varying the probability distribution of $\boldsymbol{\nu}_d$ among those that have the same first order and second order statistics. It is easy to see that $\boldsymbol{\nu}_d$ is zero mean. The auto-correlation of $\boldsymbol{\nu}_d$ is given by

$$\begin{aligned} \mathrm{E}\ \boldsymbol{\nu}_d\boldsymbol{\nu}_d^H &= \rho_d\mathrm{E}\ \tilde{\mathbf{H}}_d\tilde{\mathbf{H}}_d^H + \frac{1}{(L+1)(1+\frac{c}{\sigma^2})}\mathbf{T}\mathbf{T}^H + \sigma_w^2\mathbf{I} \\ &\triangleq \mathbf{R}_{\boldsymbol{\nu}} \end{aligned} \tag{12}$$

We also note that $\boldsymbol{\nu}_d$ is uncorrelated with the signal $\hat{\mathbf{H}}_d\mathbf{s}_d$ due to the property of the MMSE estimate.

It can be shown as in [1, 10], that the worst case noise is zero mean Gaussian with auto-correlation $\mathbf{R}_{\boldsymbol{\nu}}$ and is independent of $\mathbf{s}_d$. Therefore we have

$$\begin{aligned} C(\mathcal{P}, \rho_d, \rho_t, \mathbf{s}_t) &\geq C_{lb}(\mathcal{P}, \rho_d, \rho_t, \mathbf{s}_t) \\ C_{lb}(\mathcal{P}, \rho_d, \rho_t, \mathbf{s}_t) &\triangleq \mathrm{E}\{\log\det\left(\mathbf{I} + \rho_d\mathbf{R}_{\boldsymbol{\nu}}^{-1}\hat{\mathbf{H}}_d\hat{\mathbf{H}}_d^H\right)\} \end{aligned} \tag{13}$$

where the expectation is with respect to the random variable $\hat{\mathbf{h}}$. As in [3], we propose a lower-bound that is looser than the one given above but is simpler to handle. From (12), the matrix $\mathbf{R}_{\boldsymbol{\nu}}$ is a sum of three matrices. The first matrix is given by

$$\rho_d\mathrm{E}\{\tilde{\mathbf{H}}_d\tilde{\mathbf{H}}_d^H\} = \begin{bmatrix} \rho_d\mathrm{E}\{\tilde{\mathbf{H}}_{m_1}\tilde{\mathbf{H}}_{m_1}^H\} & & \\ & \ddots & \\ & & \rho_d\mathrm{E}\{\tilde{\mathbf{H}}_{m_J}\tilde{\mathbf{H}}_{m_J}^H\} \end{bmatrix} \tag{14}$$

We assume that the training is orthogonal which implies that the errors in the estimates of the taps are uncorrelated. Each of the matrix $\mathrm{E}\{\tilde{\mathbf{H}}_{m_i}\tilde{\mathbf{H}}_{m_i}^H\}$ is hence a diagonal matrix. We define as $\mathbf{R}_{\boldsymbol{\nu}_1}$, the matrix obtained by replacing $\mathrm{E}\{\tilde{\mathbf{H}}_d\tilde{\mathbf{H}}_d^H\}$ in (12) by a multiple of the identity matrix whose diagonal elements are to the largest entry in $\mathrm{E}\{\tilde{\mathbf{H}}_d\tilde{\mathbf{H}}_d^H\}$ [3]. From the convexity of determinants over non-negative hermitian matrices, it follows that

$$|\mathbf{I}+\rho_d\mathbf{R}_{\boldsymbol{\nu}_1}^{-1}\hat{\mathbf{H}}_d\hat{\mathbf{H}}_d^H| \leq |\mathbf{I}+\rho_d\mathbf{R}_{\boldsymbol{\nu}}^{-1}\hat{\mathbf{H}}_d\hat{\mathbf{H}}_d^H|. \quad (15)$$

This is used to propose the lower bound

$$C_{lb}(\mathcal{P},\rho_d,\rho_t,\mathbf{s}_t) = \mathrm{E}\{\log|\mathbf{I}+\rho_d\frac{\frac{c}{\sigma^2}}{1+\frac{c}{\sigma^2}}\mathbf{R}_{\boldsymbol{\nu}_1}^{-1}\bar{\mathbf{H}}_d\bar{\mathbf{H}}_d^H|\} \quad (16)$$

where $\bar{\mathbf{H}}_d$ is obtained by normalizing $\hat{\mathbf{H}}_d$. Specifically, the channel $\bar{\mathbf{h}}$ that generates $\bar{\mathbf{H}}_d$ is normalized to zero mean, i.i.d Gaussian with variance of each tap equal to $\frac{1}{L+1}$.

D Quasi-Periodic Placement Schemes

In this section we introduce a family of placement schemes called Quasi-Periodic Placement (QPP) schemes. This family is divided into different classes based on the minimum allowable cluster size. The class of schemes for which α is the minimum cluster size is denoted as QPP-α. Intuitively, QPP-α scheme is formed by first breaking the known symbols into as many clusters as possible each of length at least α and then placing these clusters such that the unknown symbol blocks are as "equal" as possible. We give the formal definition below.

Definition 1 *Given α and a frame with N unknown symbols and $P \geq \alpha$ known symbols, let $J_\alpha = \lfloor\frac{P}{\alpha}\rfloor + 1$. A placement scheme $\mathcal{P} = (\mathbf{n},\mathbf{m})$ belongs to QPP-α if and only if*

1. $\mathbf{n} \in \mathcal{N}^{J_\alpha}$ *where* $\mathcal{N}^{J_\alpha} = \{(n_1,\cdots,n_{J_\alpha+1}) : \sum_{i=2}^{J_\alpha} n_i = P \ \&\ n_1 = n_{J_\alpha+1} = L \ \&\ \min(\{n_2,\cdots,n_{J_\alpha}\}) \geq \alpha\}$
2. $\mathbf{m} \in \mathcal{M}^{J_\alpha}$ *where* $\mathcal{M}^{J_\alpha} = \{(m_1,\cdots,m_{J_\alpha}) : \sum_i m_i = N \ \&\ m_i \in \{\lfloor\frac{N}{J_\alpha}\rfloor, (\lfloor\frac{N}{J_\alpha}\rfloor+1)\}\}$

Any element of the set $\mathcal{N}^{J_\alpha}$ is denoted as $\mathbf{n}_{J_\alpha} = (\bar{n}_1,\cdots,\bar{n}_{J_\alpha})$ and similarly any element of the set $\mathcal{M}_{J_\alpha}$ is denoted as $\mathbf{m}_{J_\alpha} = (\bar{m}_1,\cdots,\bar{m}_{J_\alpha})$.

E Optimality of QPP-α Schemes for Unknown Channel

We obtain optimal training $(\mathcal{P}^*,\rho_d^*,\rho_t^*,\mathbf{s}_t^*)$ as

$$(\mathcal{P}^*,\rho_d^*,\rho_t^*,\mathbf{s}_t^*) = \arg\max_{\mathcal{P},\rho_d,\rho_t,\mathbf{s}_t} C_{lb}(\mathcal{P},\rho_d,\rho_t,\mathbf{s}_t) \quad (17)$$

The following theorem shows that under the assumption that $\alpha \geq 2L+1$, the placement schemes belonging to QPP-α are optimal. Further more, an optimal choice of training symbols is also given.

Theorem 1 *Given any energy trade-off, (ρ_d,ρ_t), under the assumption that $\alpha \geq (2L+1)$ and $P \geq \alpha$, the placement scheme $\mathcal{P}^*$ and training $\mathbf{s}_t^*$ is optimal if*

1. *$\mathcal{P}^*$ belongs to QPP-α .*
2. $$|s_{it}^k| = \begin{cases} \sqrt{\frac{(P+L)\rho_t}{J-1}} & \text{if } k=(L+1), \quad i=2,\cdots,J \\ 0 & \text{otherwise}\end{cases} \quad (18)$$

If $(L+1) \leq P < \alpha$, the known symbols are placed at the beginning and the end of the packet such that at least $(L+1)$ are at one of the ends. That is, a placement scheme $\mathcal{P}^*$ and training symbols $\mathbf{s}_t^*$ are optimal if

1. $\mathcal{P} = (\mathbf{m},\mathbf{n})$ *where* $\mathbf{m} = (N), \mathbf{n} = (2L+1+\beta, P-1-\beta)$ *and* $0 \leq \beta \leq P-(L+1)$.
2. $$|s_{it}^k| = \begin{cases} \sqrt{(P+L)\rho_t} & \text{if } k=(L+1), \quad i=1 \\ 0 & \text{otherwise}\end{cases} \quad (19)$$

In either case we have $C_{lb}(\mathcal{P}^*,\rho_d,\rho_t,\mathbf{s}_t^*)$ equal to

$$\sum_{i=1}^{J_\alpha} E\left\{\log\left(1+\frac{\rho_d}{\sigma_w^2}\frac{(P+L)\rho_t}{(P+L)\rho_t+(L+1)(\rho_d+\sigma_w^2)}\bar{\mathbf{H}}_{\bar{m}_i}^H\bar{\mathbf{H}}_{\bar{m}_i}\right)\right\} \quad (20)$$

We find that QPP-α placement schemes that were found to be optimal in the known channel scenario [8] are optimal for this scenario too. From (18) and (19), we find that for the optimal choice of training symbols, the L symbols at the beginning and the end of each known symbol cluster are zero. This is because if these symbols are non-zero, they contribute additional noise to the received data due to the error in the channel estimate. Also we find that in each cluster, there is only one non-zero training symbol. This design makes sure that the training is always orthogonal. For $L \leq \alpha \leq 2L$, it is difficult to analytically obtain the optimal placement schemes.

The minimum known symbol cluster size α is also a design parameter. The following theorem gives the optimal value of α.

Theorem 2 *For $\alpha \geq 2L+1$, $C_{lb}(\mathcal{P}^*,\rho_d,\rho_t,\mathbf{s}_t^*)$ is a monotonically decreasing function of α.*

The obtained placement schemes are optimal for any energy allocation. The following theorem gives the optimal energy allocation between training and data under the assumption that the optimal placement scheme and training symbols are used.

Theorem 3 *The optimal energy distribution is given by*

$$\rho_d^* = (\sqrt{\gamma}-\sqrt{\gamma-1})\frac{\sqrt{\gamma}}{g},$$
$$\rho_t^* = (\sqrt{\gamma}-\sqrt{\gamma-1})\frac{1}{\sqrt{hk}}. \quad (21)$$

where $h = \frac{P+L}{T}$, $g = \frac{N}{T}$, $k = \frac{(P+L)(N-L-1)}{(L+1)(T+N\sigma_w^2)}$ and $\gamma = \frac{h}{k}+1$.

The ratio of power in data to that in training is given by

$$\frac{g\rho_d^*}{h\rho_t^*} = \sqrt{1+\frac{N-L-1}{(L+1)(1+g\sigma_w^2)}}. \quad (22)$$

At low SNR we find that this ratio is equal to one. Hence half the energy is spent in training.

IV. Optimal Placement Scheme and Training for OFDM System

In this section, we state the results for OFDM systems [9]. For OFDM systems, known symbols introduced in frequency. Every OFDM symbols has N data and P training symbols. We assume that all the training tones have equal energy. A placement scheme is represented by the set $\mathcal{P}$ that contains the indices of the tones used for training.

Theorem 4 *For any energy trade off (ρ_d, ρ_t), under the assumption that $N = mP(m \geq 1)$, and $P \geq (L+1)$, all the following periodic placements are optimal*

$$\mathcal{P}^* = \{i, i+m+1, i+2(m+1), \cdots, i+(P-1)(m+1)\} \quad (23)$$

where i can take values from 1 to $(m+1)$. For any of these placements the lower bound is given by

$$C_{lb}(\mathcal{P}^*, \rho_d, \rho_t) = NE \log\left(1 + \frac{\rho_d}{\sigma_w^2} \frac{P\rho_t dd^*}{P\rho_t + (L+1)(\rho_d + \sigma_w^2)}\right) \quad (24)$$

where d is a complex Gaussian random variable with zero mean and unit variance.

It was shown in [6, 12] that the same set of placements minimizes the mean square error in the estimate of $\mathbf{h}$. Their performance metric is hence $\sum_{i=1}^{N+P} k_i$, the sum of MSE of both data and training tones. Our performance metric is quite different. In fact, the capacity lower-bound depends explicitly only on the MSE of data tones and not on those of training tones. It is surprising that the same set of placements is optimal for this metric too.

V. Simulation

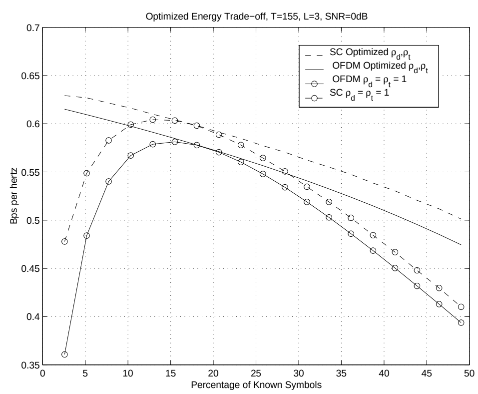

Fig. 6: Comparison of the variation of training based lower bound with the percentage of known symbols for OFDM and Single-Carrier (SC) systems with at $T = 155$ and $L = 3$ for SNR = 0dB.

In this section, we study the training-based capacity for single-carrier systems and OFDM systems through simulations. Figs. 6 and 7 compare the variation of the training-based lower bound with percentage of known symbols for OFDM and Single-Carrier(SC) systems with the coherence interval $T = 155$ and the channel length equal to 4. We find that the training-based capacity for single-carrier systems is better than that of OFDM systems consistently. For optimized (ρ_d, ρ_t), we find that the percentage difference is less than 5%. For equal energy case, at low SNR, we find that the single-carrier system performs considerably better than the OFDM system at small percentage of known symbols. This difference becomes smaller with the number of known symbols. At high SNR, the percentage difference between OFDM and single-carrier systems becomes much smaller. The optimum percentage of known symbols is approximately the same for both single-carrier and OFDM systems. From simulations, we find that for energy allocation optimized scenario the optimum number of known symbols is equal to $(L+1)$ for both the systems. A similar conclusion was reached in [3].

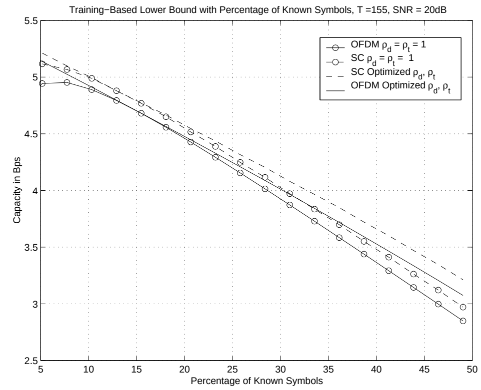

Fig. 7: Comparison of the variation of training based lower bound with the percentage of known symbols for OFDM and Single-Carrier (SC) systems with at $T = 155$ and $L = 3$ for SNR = 20dB.

We evaluate the asymptotic performance of training-based systems in Figs. 8 and 9. We plot the lower bound versus the coherence interval T for low SNR(0 dB) and high SNR(20 dB). The value of L was set to 3. The minimum cluster size α was made equal to $2L+1$. For each value of T, the optimum number of known symbols was used. The placement scheme used was a QPP-α scheme. As expected the known channel capacity for OFDM converges to that for SC systems at large T. We find that for optimized (ρ_d, ρ_t), the difference between OFDM and SC systems is quite small. For equal energy allocation, though, we find that at intermediate values of T, single-carrier systems can outperform OFDM systems by as much as 10%. We find that at high SNR, for both SC and OFDM systems, asymptotically training-based capacity approaches the known channel capacity. This is not true at low SNR.

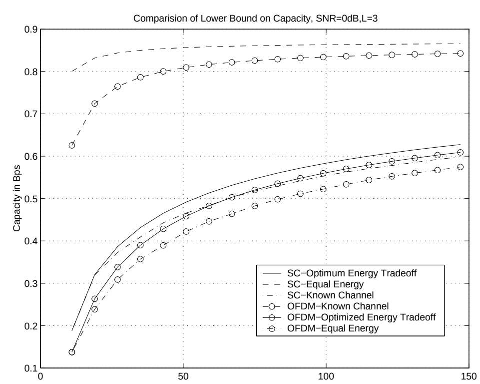

Fig. 8: Comparison of the variation of training based lower bound with the coherence interval T for OFDM and Single-Carrier(SC) systems for $L = 3$ at SNR = 0dB.

VI. Conclusions

The problem of designing optimal training symbol placement schemes for block frequency selective fading channels is presented. It is assumed that the receiver forms an MMSE estimate of the channel based on only training. The problem is addressed for both single-carrier systems and OFDM separately since the paradigm for channel estimation is different for each system. The metric used for optimization was a tight lower bound on the i.i.d capacity of the system.

For single-carrier system, we assume that the known symbols are placed in clusters of length $\alpha \geq L$. For $\alpha \geq (2L+1)$, we show that the placement schemes belonging to thal QPP-α family are optimal. Furthermore a choice of optimal training symbols is presented. Expressions for optimal energy allocation between data and energy are given. It is shown that for OFDM systems, under the assumption that the training tones are of equal energy, the optimal placement scheme is that for which the training tones are selected periodically.

We find from simulations that for both OFDM and single-carrier systems, when we optimize the energy allocation (ρ_d, ρ_t) the optimum number of known symbols is equal to $(L+1)$. For equal energy allocation scenario, the optimum percentage of known symbols decreases with SNR. From simulations, we find that at large values of T and at high SNR, training-based systems achieve most of the unknown channel capacity. At low SNR however, this is not true. This conclusion was also reached in [1, 3]. The comparison of the lower bound for OFDM and single-carrier systems shows that the single-carrier system performs better than the OFDM systems. This is to be expected because the OFDM system drops some received data for simpler receiver implementation. We find that for optimal energy allocation, the percentage difference between the two systems is quite small. For equal energy case, on the other hand, the single-carrier system might be considerably better than the OFDM system for some values of T and P.

References

[1] B.Hassibi and B.Hochwald. "How much Training is Needed in Multiple-Antenna Wireless Links". *Submitted to IEEE Trans. Information Theory*, August 2000.

[2] E.Biglieri, J.Proakis, and S.Shamai. "Fading Channels: information-theoretic and communication aspects". *IEEE Trans. Information Theory*, 44(6):2596–2618, October 1998.

[3] H.Vikalo, B.Hassibi, B.Hochwald, and T.Kailath. "Optimal Training for Frequency-Selective Fading Channels". In *Accepted to ICASSP*, Salt Lake City, UT, May 2001.

[4] G.D.Forney Jr. and G.Ungerboeck. "Modulation and Coding for linear Gaussian channels". *IEEE Trans. Information Theory*, 44(6):2596–2618, October 1998.

[5] M.Medard. "The effect upon channel capacity in wireless communication of perfect and imperfect knowledge of the channel". *IEEE Trans. Information Theory*, 46(3):933–946, May 2000.

[6] R.Negi and J.Cioffi. "Pilot Tone Selection for Channel Estimation in a Mobile OFDM System". *IEEE Tran. Consumer Electronics*, 44(3):1122–1128, August 1998.

[7] S.Adireddy and L.Tong. "Detection with Embedded Known Symbols : Optimal Symbol Placement and Equalization". In *Proc. ICASSP*, Istanbul, Turkey, June 2000.

[8] S.Adireddy and L.Tong. "Optimal Placement of Known Symbols". In *Proc. CISS*, Princeton, NJ, March 2000.

[9] S.Adireddy, L.Tong, and H.Viswanathan. "Optimal Placement of Training for Unknown Channels". In *Accepted to ICASSP*, Salt Lake City, UT, May 2001.

[10] S.Diggavi and T.Cover. "The Worst Additive noise under a Covariance Constraint". *Submitted to IEEE Tran. Information Theory.*

[11] Shlomo Shamai (Shitz) and Rajiv Laroia. " The Intersymbol Interference Channel: Lower bounds on Capacity and Channel Precoding Loss". *IEEE Trans. Inform. Theory*, 42(5):1388–1404, September 1996.

[12] S.Ohno and G.B.Giannakis. "Optimal training and redundant precoding for block transmissions with application to wireless OFDM". In *Accepted to ICASSP*, Salt Lake City, UT, May 2001.

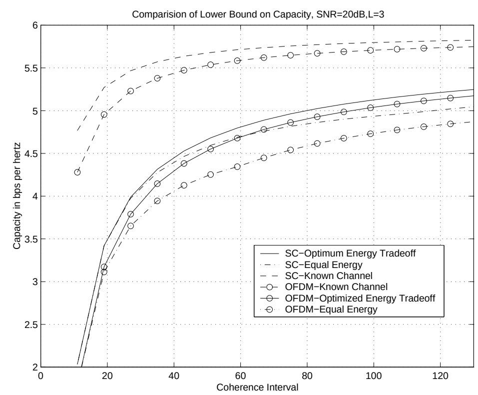

Fig. 9: Comparison of the variation of training based lower bound with the coherence interval T for OFDM and Single-Carrier(SC) systems for $L = 3$ at SNR = 20dB.