Optimal Design and Placement of Pilot Symbols for Channel Estimation

Min Dong, *Student Member, IEEE*, and Lang Tong, *Senior Member, IEEE*

***Abstract*—The problem of designing and placing pilot symbols for the estimation of frequency-selective random channels is considered. The channel is assumed to be a block-fading model with finite impulse response (FIR). For both single-input single-output (SISO) and multiple-input multiple-output (MIMO) channels, under the assumption of independent and identical distributed channel taps, the Cramér–Rao Bound (CRB) on the mean square error (MSE) of semi-blind channel estimators is derived and minimized with respect to pilot symbols and their placement. It is shown that the optimal strategy is to place pilot symbols satisfying certain orthogonality condition in such a way that data and pilot symbols with higher power are in the middle of the packet. The results also indicate that the optimal pilot placements are independent of channel probability distribution. For constant modulus symbols, we show that the quasi-periodic placement and its generalization in the multiuser case turn out to be optimal. We further consider estimating channels with correlated taps and show that the previous placement strategy is also optimal among orthogonal pilot sequences.**

***Index Terms*—Channel estimation, Cramér-Rao bound, optimal design, pilot symbols, placement schemes, semi-blind.**

I. Introduction

CHANNEL estimation plays a critical role in packet-switched wireless systems where it is often necessary to acquire the channel state for each packet. To facilitate channel estimation and synchronization, pilot symbols are usually embedded in a data stream. Consequently, it is important to fully utilize these symbols to obtain optimal estimation performance, and the placement of these pilot symbols can affect significantly the overall performance of a wireless system [1]–[4].

The optimization of pilot symbols and their placement has not been investigated until recently, although the design of optimal pilot sequence for training-based channel estimators is an old problem and has been investigated by many [5]–[8]. In [4] and [9], optimal pilot tone selection that minimizes the mean square error (MSE) of the minimum MSE (MMSE) channel estimator for orthogonal frequency division multiplexing (OFDM) systems are considered. In [10], Ling analyzed optimal performance of two pilot-assisted schemes in various aspects of code division multiple access (CDMA) systems. Adireddy and Tong considered the optimal placement problem for decision feedback equalization (DFE) [1]. From an information-theoretic perspective, they also optimized the known symbols placement for maximizing channel capacity or minimizing outage probability [2]–[11]. Sadler *et al.* [12] developed Cramér-Rao Bounds (CRBs) for estimating source and deterministic channel under the availability of side information by employing the constrained CRB formulation [13] and evaluated performance under different placements of known symbols through simulations. Carvalho and Slock [14] obtained expressions of CRBs for deterministic channels and examined the placement of pilot symbols via computer simulations. In their case, no optimal strategy was found as the CRB for the deterministic channel model is also a function of channel coefficients. For orthogonal space-time codes, the placement of superimposed pilot symbols for memoryless multiple-input multiple-output (MIMO) channels is considered in [15]. Aside from these previous results, however, the problem of pilot symbols placement for channel estimation in a wireless transmission system has yet to be fully studied, and optimal placement strategy is still unknown.

In this paper, we consider the optimal design and placement of pilot symbols for channel estimation. Since mobile users may choose different channel estimators, in searching for the optimal placement, it is desirable to use a criterion that is independent of any specific estimation technique used by individual receivers. A natural choice is the CRB on the MSE of channel estimators, and the objective of designing the pilot sequence and its optimal placement is to minimize the CRB.

The main contributions of this paper are as follows. For both single-input single-output (SISO) and MIMO finite impulse response (FIR) random channels, under the assumption of independent and identical distributed (i.i.d.) channel taps, we first obtain an expression of the CRB as a function of pilot symbols and their placement. It is then shown that the CRB is minimized by placing pilot symbols with smaller magnitudes closer to two ends of a packet and those with larger magnitudes closer to the center while satisfying certain orthogonality conditions. We show that, although the CRBs are functions of channel distributions, the optimal pilot placements are independent of probability distribution of the channel. This is especially important in broadcasting applications, where the pilot design should be optimal for channels of all users. We further consider estimation of channels with correlated taps and show that the previous placement strategy is also optimal among orthogonal pilot sequences.

For constant modulus pilot symbols with sufficient power, we show that the optimal strategy is to place pilot symbols, possibly

Manuscript received April 18, 2001; revised July 8, 2002. This work was supported in part by the Army Research Office under Grant ARO-DAAB19-00-1-0507, the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564, and Army Research Laboratory CTA on Communication and Networks. The associate editor coordinating the review of this paper and approving it for publication was Dr. Vikram Krishnamurthy.

The authors are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: mdong@ee.cornell.edu; ltong@ee.cornell.edu).

Digital Object Identifier 10.1109/TSP.2002.805504

in multiple clusters, in the middle of a packet. Although this result confirms the advantage of using the midamble placement as in the Global System for Mobile communication (GSM), it also suggests that some other placements are also optimal. One of such optimal placements is the quasi-periodic placement (QPP)-α scheme , which, under mild conditions, was shown to be optimal for DFE [1], as well as optimal in terms of maximizing channel capacity .[1]

This paper is organized as follows. In Section II, we introduce the basic SISO channel model and pilot symbol placement. In Section III, the CRB for the random channel vector as a function of pilot symbols and their placement is derived. In Section IV, we obtain optimal design and placement schemes that minimize the CRB, followed by discussions of the placement strategies and tradeoffs. In Section V, we extend our results to MIMO channels and obtain corresponding optimal placement schemes for the multiuser case. In Section VI, optimal placements of orthogonal pilot sequences for random channels with correlated taps are obtained. Numerical results are presented in Section VII.

Notation used in this paper are standard. Upper and lowercase bold letters denote matrices and vectors, respectively. $(\cdot)^*$ denotes the conjugation and $(\cdot)^H$ the Hermitian transpose. We use $\mathbf{A}_{m\times n}$ to denote a matrix $\mathbf{A}$ with size $m \times n$ and $[\mathbf{A}]_{ij}$ the ijth element of matrix $\mathbf{A}$. The Kronecker product of matrix $\mathbf{A}$ and $\mathbf{B}$ is denoted as $\mathbf{A} \otimes \mathbf{B}$. Matrix $\mathbf{I}$ stands for identity matrix.

II. Problem Statement

A. Model

We assume a frequency-selective block-fading model where the random channel remains constant for one data packet and changes to an independent value for the next packet. We further assume that channel estimation is performed within one transmitted packet. The estimation of an SISO FIR channel is first considered. Results for MIMO channels are presented in Section V.

Within one data packet, the channel is modeled by an FIR linear system with order L

$$y_k = \sum_{i=0}^{L} h_i s_{k-i} + n_k, \quad k = L+1, \ldots, N+P \tag{1}$$

where y_k is the received signal, $\mathbf{h} = [h_0, \ldots, h_L]^T$ is the channel vector, s_k is the input symbol, and n_k is the i.i.d. circular complex Gaussian noise with zero mean and variance σ_n^2.

We assume that each data packet consists of N data symbols denoted as $\mathbf{s}_d = [s_d[1], \ldots, s_d[N]]^T$ and P pilot symbols as $\mathbf{s}_p = [s_p[1], \ldots, s_p[P]]^T$. The vector channel model is used for the entire packet corresponding to N data symbols and P pilot symbols. Denoting $\mathbf{y} = [y_{N+P}, \ldots, y_{L+1}]^T$, $\mathbf{s} = [s_{N+P}, \ldots, s_1]^T$, we have

$$\mathbf{y} = \mathcal{T}(\mathbf{h})\mathbf{s} + \mathbf{n} = \mathcal{H}(\mathbf{s})\mathbf{h} + \mathbf{n} \tag{2}$$

[1]In , the channel capacity is maximized under the constraint that certain percentage of input symbols is used for training.

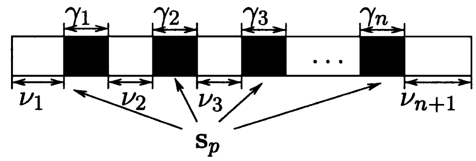


Fig. 1. Input data packet with multiple pilot clusters.

where $\mathcal{T}(\mathbf{h})$ is a Toeplitz matrix generated from $\mathbf{h}$ and $\mathcal{H}(\mathbf{s})$ a Hankel matrix from input $\mathbf{s}$

$$\mathcal{T}(\mathbf{h}) = \begin{pmatrix} h_0 & \cdots & h_L & & \\ & \ddots & & \ddots & \\ & & h_0 & \cdots & h_L \end{pmatrix}_{(N+P-L)\times(N+P)} \tag{3}$$

$$\mathcal{H}(\mathbf{s}) = \begin{pmatrix} s_{N+P} & \cdots & s_{N+P-L} \\ \vdots & \text{Hankel} & \vdots \\ s_{L+1} & \cdots & s_1 \end{pmatrix}_{(N+P-L)\times(L+1)}. \tag{4}$$

The channel $\mathbf{h}$ is to be estimated using the observation $\mathbf{y}$ of the entire packet, i.e., the estimation is semi-blind.

We also make the following assumptions:

A1) Data symbols are drawn from an i.i.d. sequence that has probability density function (pdf) $p_s(\cdot)$ with zero mean and variance σ_d^2. The power of pilot symbols is defined as $\sigma_p^2 \triangleq 1/P \sum_{i=1}^{P} |s_p[i]|^2$.

A2) Taps of the channel $\mathbf{h}$ are i.i.d. random variables with pdf $p_h(\cdot)$.

A3) The data $\mathbf{s}$, channel $\mathbf{h}$, and noise $\mathbf{n}$ are jointly independent.

Assumption A2 may be restrictive in practice when specific pulse shaping filters are used. In Section VI, this assumption is relaxed to deal with correlated channel coefficients.

B. Pilot Symbol Placement

In general, the placement of n clusters of pilot symbols can be described by $\mathcal{P} = (\boldsymbol{\nu}, \boldsymbol{\gamma})$, where $\boldsymbol{\nu} = [\nu_1, \ldots, \nu_{n+1}]$ is the data block length vector, and $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_n]$ the pilot cluster length vector, as illustrated in Fig. 1. Constraining the total number of data and pilot symbols, we have $\sum_{i=1}^{n+1} \nu_i = N$ and $\sum_{i=1}^{n} \gamma_i = P$. Moreover, for those placements starting with pilot symbols, $\nu_1 = 0$, and for those ending with pilot symbols, $\nu_{n+1} = 0$.

We also define the edge and midamble positions for each packet, as shown in Fig. 2. Edge positions are defined as the first and last L positions in a packet. The rest of the parts within interval $[L+1, N+P-L]$ are midamble positions.

For a training-based channel estimation, only those parts of the observations corresponding to pilot symbols are used. If there is a pilot cluster of length less than $L+1$, no pilot symbols corresponding to this cluster can be used for channel estimation. Therefore, it is intuitive that all pilot symbols should be grouped into a single cluster. This intuition, however, is not valid if all observations are used for channel estimation. Indeed, the use of multiple clusters results in a simpler design of pilot symbols as shown in Section IV and better detection performance (see [1]).

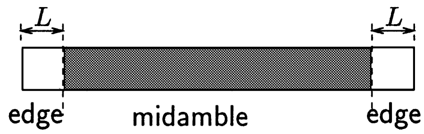


Fig. 2. Edge and midamble positions of one data packet.

The input symbol vector can be decomposed into the pilot and data parts

$$\mathbf{s} = \bar{\mathbf{s}}_p + \bar{\mathbf{s}}_d \quad (5)$$

where $\bar{\mathbf{s}}_p$ is obtained by setting the data part of $\mathbf{s}$ to zero. This introduces a similar decomposition of the input symbol matrix

$$\mathcal{H}(\mathbf{s}) = \mathcal{H}(\bar{\mathbf{s}}_p) + \mathcal{H}(\bar{\mathbf{s}}_d). \quad (6)$$

For convenience, we define their "autocorrelation" matrices as

$$\mathbf{R}_{\mathbf{s}} \stackrel{\Delta}{=} \mathcal{H}^H(\mathbf{s})\mathcal{H}(\mathbf{s}), \mathbf{R}_{\mathbf{s}_p} \stackrel{\Delta}{=} \mathcal{H}^H(\bar{\mathbf{s}}_p)\mathcal{H}(\bar{\mathbf{s}}_p). \quad (7)$$

Note that quantities $\mathbf{s}, \bar{\mathbf{s}}_p$ and their corresponding autocorrelation matrices are functions of placement $\mathcal{P}$. It follows that $\mathbf{R}_{\mathbf{s}}$ and $\mathbf{R}_{\mathbf{s}_p}$ are functions of $\mathcal{P}$ and $\mathbf{s}_p$.

III. Cramér–Rao Bound

The CRB for random channels is used as a performance measure for the design and placement of pilot symbols. The following theorem provides the expression of the CRB as a function of pilot symbols and their placement.

Theorem 1: Under the assumptions A1-A3 and the regularity conditions [16], [17], the MSE matrix of any channel estimator $\hat{\mathbf{h}}(\mathbf{y})$, which is defined as

$$\mathcal{M}(\hat{\mathbf{h}}) \stackrel{\Delta}{=} E\{[\hat{\mathbf{h}}(\mathbf{y}) - \mathbf{h}][\hat{\mathbf{h}}(\mathbf{y}) - \mathbf{h}]^H\} \quad (8)$$

satisfies the following inequality:

$$\mathcal{M}(\hat{\mathbf{h}}) \geq \mathbf{\Lambda}(\mathcal{P}, \mathbf{s}_p) \stackrel{\Delta}{=} \left(\frac{1}{\sigma_n^2} E\{\mathbf{R}_{\mathbf{s}}\} + \rho_h^2 \mathbf{I}\right)^{-1} \quad (9)$$

where $\mathbf{\Lambda}(\mathcal{P}, \mathbf{s}_p)$ is the complex CRB, and $\rho_h^2 = E\{|\partial \ln p_h(h)/\partial h^*|^2\}$ with the expectation taken with respect to $p_h(h)$.

Proof: See Appendix A.

The objective is to minimize the CRB of channel estimators, jointly with respect to pilot symbols and their placement under the pilot power constraint, i.e.,

$$(\mathcal{P}_*, \mathbf{s}_{p_*}) = \underset{\mathcal{P}, \mathbf{s}_p : \|\mathbf{s}_p\|^2 = P\sigma_p^2}{\arg\min} \operatorname{tr}[\mathbf{\Lambda}(\mathcal{P}, \mathbf{s}_p)]. \quad (10)$$

From (9), we note that the CRB for channel estimators depends on channel distribution through ρ_h^2. We show later in Section IV that, fortunately, the minimization of CRB with respect to $\mathbf{s}_p$ and $\mathcal{P}$ turns out to be independent of the channel distribution.

We also note that $\mathbf{\Lambda}(\mathcal{P}, \mathbf{s}_p)$ defined above completely determines the real CRB [18] under the circular complex Gaussian noise assumption, and

$$E\left\{\left(\frac{\partial \ln p_s(s_d)}{\partial {s_d}^*}\right)^2\right\} = 0, E\left\{\left(\frac{\partial \ln p_h(h)}{\partial h^*}\right)^2\right\} = 0. \quad (11)$$

The regularity conditions require that the joint distribution $p(\mathbf{y}, \mathbf{s}_d, \mathbf{h})$ be absolutely continuous with respect to $s_d[i]$. An example of such data sequences that satisfies the conditions is the sequence with Gaussian distribution.[2] For those drawn from discrete symbol constellations, the above theorem gives an approximation.

IV. Optimal Design and Placement for SISO Channels

A. Optimal Design and Placement

In this section, we consider the design and placement of n clusters of pilot symbols, as shown in Fig. 1. The placement of the clusters is specified by $\mathcal{P} = (\boldsymbol{\nu}, \boldsymbol{\gamma})$. For pilot symbols all placed in midamble positions, the following lemma concludes the shift invariant property of the CRB.

Lemma 1: For any $\mathbf{s}_p$, the midamble placement is shift invariant, i.e., for any $\mathcal{P} = \{(\boldsymbol{\nu}, \boldsymbol{\gamma}) : \nu_1 \geq L, \nu_{n+1} \geq L+1\}$

$$\mathbf{\Lambda}(\boldsymbol{\nu}, \boldsymbol{\gamma}, \mathbf{s}_p) = \mathbf{\Lambda}(\boldsymbol{\nu} + \mathbf{e}_1 - \mathbf{e}_{n+1}, \boldsymbol{\gamma}, \mathbf{s}_p) \quad (12)$$

where vector $\mathbf{e}_i$ denotes the unit row vector with 1 at the ith entry and 0 elsewhere.

Proof: See Appendix B.

Before we present the optimal pilot placement and design in Theorem 2, we first make some heuristic arguments and illustrate the idea in Fig. 3. With the invariance property given in Lemma 1, we know that placements in midamble positions are invariant with respect to shifts. Therefore, one should pay special attention to placements at the edge positions. Note that the channel model given in (3) assumes no knowledge about the channel input for $s_k, k = 0, -1, \ldots$, and those observations $\{y[k]\}_{k=1}^L$ relating to these unknown input symbols are discarded in channel estimation. However, $\{y[k]\}_{k=1}^L$ are related to input symbols $\{s[k]\}_{k=1}^L$, and discarding them prevents us from fully utilizing the first L input symbols. It is therefore logical that one should allocate minimum power to symbols at edge positions.

Fig. 3 illustrates the optimal placement given in Theorem 2. When there are many pilot symbols, i.e., $P \geq 2L+1$, the optimal design calls for setting zeros to symbols at two edges of a packet and putting the rest pilots in the midamble part in such a way that they satisfy certain orthogonality condition. On the other hand, when there are only a few pilot symbols, i.e., $P \leq 2L$, it is no longer possible to set all symbols at the edge positions zero. In such a case, it depends on how much power is allocated to the pilot symbols. If the total power of pilots is higher than the power of data symbols, then all the pilot power should be concentrated on one symbol placed in the midamble part of the packet. Otherwise, two pilot symbols, each with half of the total power, should be placed at the edge positions as

[2] We know that the capacity-achieving input distributions for known channels are Gaussian. In practice, the symbols may be shaped to approximate the Gaussian distribution.

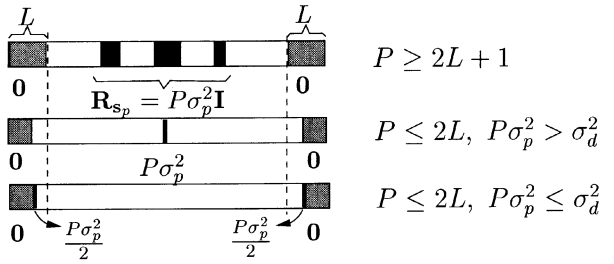


Fig. 3. Optimal pilot designs and placements for SISO channels.

close to the center as possible. The specification of the optimal placement and design of pilot symbols is given in Theorem 2.

Theorem 2: Given N data symbols with power σ_d^2 and P pilot symbols with power σ_p^2. Let $\lambda_i(\mathcal{P}, \mathbf{s}_p)$ be the ith diagonal entry of $\mathbf{\Lambda}(\mathcal{P}, \mathbf{s}_p)$. Under assumptions A1-A3, we have the following.

1) For $P \geq 2L+1$, the optimal placement $\mathcal{P}_*$ and design of pilot symbols $\mathbf{s}_{p_*}$ given by

$$\mathcal{P}_* \in \{(\boldsymbol{\nu}, \boldsymbol{\gamma}) : n > 2; \nu_1 = \nu_{n+1} = 0; \gamma_1 = \gamma_n = L\} \tag{13}$$

$$\mathbf{s}_{p_*} \in \{\mathbf{s}_p : s_p[i] = 0, \forall i \in [1, L] \cup [P-L+1, P] \quad \mathbf{R}_{\mathbf{s}_p} = P\sigma_p^2\mathbf{I}\}. \tag{14}$$

The minimum CRB is given by

$$\begin{aligned}\lambda_i(\mathcal{P}_*, \mathbf{s}_{p_*}) &= \min_{\mathcal{P}, \mathbf{s}_p : \|\mathbf{s}_p\|^2 = P\sigma_p^2} \lambda_i(\mathcal{P}, \mathbf{s}_p) \\ &= \frac{\sigma_n^2}{N\sigma_d^2 + P\sigma_p^2 + \rho_h^2\sigma_n^2} \\ & i = 1, \ldots, L+1.\end{aligned} \tag{15}$$

2) For $P \leq 2L$:

i) If $P\sigma_p^2 > \sigma_d^2$, the optimal placement $\mathcal{P}_*$ and design of pilot symbols $\mathbf{s}_{p_*}$ are given by

$$\begin{aligned}\mathcal{P}_* \in \{(\boldsymbol{\nu}, \boldsymbol{\gamma}) &: n = 3; \nu_1 = \nu_4 = 0 \\ &\gamma_2 = 1, \gamma_1, \gamma_3 \in \left\{ \left\lfloor \frac{P-1}{2} \right\rfloor, \left\lceil \frac{P-1}{2} \right\rceil \right\}\right\}\end{aligned} \tag{16}$$

$$\mathbf{s}_{p_*}[i] = \begin{cases} P\sigma_p^2, & \text{if } i = \gamma_1 + 1 \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

The minimum CRB is given by

$$\begin{aligned}\mathrm{tr}[\mathbf{\Lambda}(\mathcal{P}_*, \mathbf{s}_{p_*})] = &\sum_{i=\max\{L-P+2,1\}}^{L-\lceil (P-1)/2 \rceil} \frac{2\sigma_n^2}{(N-i)\sigma_d^2 + P\sigma_p^2 + \rho_h^2\sigma_n^2} \\ &+ \frac{(L-P+2)\sigma_n^2 \mathcal{U}(L-P)}{(N+P-L-1)\sigma_d^2 + P\sigma_p^2 + \rho_h^2\sigma_n^2} \\ &+ \frac{(\lceil \frac{P-1}{2} \rceil - \lfloor \frac{P-1}{2} \rfloor)\sigma_n^2}{(N-L+\lfloor \frac{P-1}{2} \rfloor)\sigma_d^2 + P\sigma_p^2 + \rho_h^2\sigma_n^2} \\ &+ \frac{\max\{P-L, 0\}\sigma_n^2}{N\sigma_d^2 + P\sigma_p^2 + \rho_h^2\sigma_n^2}\end{aligned} \tag{18}$$

where $\mathcal{U}(k)$ is the unit step function

$$\mathcal{U}(k) = \begin{cases} 1, & k \geq 0 \\ 0, & k < 0. \end{cases}$$

ii) If $P\sigma_p^2 \leq \sigma_d^2$, the optimal placement $\mathcal{P}$ and design of pilot symbols $\mathbf{s}_{p_*}$ are given by

$$\begin{aligned}\mathcal{P}_* \in \{(\boldsymbol{\nu}, \boldsymbol{\gamma}) &: n = 2; \nu_1 = \nu_3 = 0 \\ &\gamma_1, \gamma_2 \in \left\{ \left\lfloor \frac{P}{2} \right\rfloor, \left\lceil \frac{P}{2} \right\rceil \right\}\right\}\end{aligned} \tag{19}$$

$$\mathbf{s}_{p_*}[i] = \begin{cases} \frac{P\sigma_p^2}{2}, & \text{if } i = \gamma_1, \gamma_1 + 1 \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

The minimum CRB is given by

$$\begin{aligned}&\mathrm{tr}[\mathbf{\Lambda}(\mathcal{P}_*, \mathbf{s}_{p_*})] \\ &= \frac{\max\{P-L-1, 0\}\sigma_n^2}{N\sigma_d^2 + P\sigma_p^2 + \rho_h^2\sigma_n^2} \\ &\quad + \frac{\max\{L+1-P, 0\}\sigma_n^2}{(N+P-L)\sigma_d^2 + \rho_h^2\sigma_n^2} \\ &\quad + \sum_{i=\max\{L-P+1,0\}}^{L-\lceil P/2 \rceil} \frac{2\sigma_n^2}{(N-i)\sigma_d^2 + \frac{P\sigma_p^2}{2} + \rho_h^2\sigma_n^2} \\ &\quad + \frac{(\lceil \frac{P}{2} \rceil - \lfloor \frac{P}{2} \rfloor)\sigma_n^2}{(N-L+\lfloor \frac{P}{2} \rfloor)\sigma_d^2 + \frac{P\sigma_p^2}{2} + \rho_h^2\sigma_n^2}.\end{aligned} \tag{21}$$

Proof: See Appendix C.

Notice that while the value of the minimum CRB depends on the channel distribution through ρ_h^2, the optimal design and placements described in Theorem 2 are independent of ρ_h^2, and therefore, it is independent of the probability distribution of the channel. In other words, the placements are optimal for any channel distribution. For a sufficient number of pilot symbols, i.e., $P \geq 2L+1$, the denominator of the minimum CRB in (15) shows the total power of N data and P pilot symbols indicating that under the optimal placement, all power in the data packet is included in the estimation. Although Theorem 2 concludes that concentrating all the data and pilot power in the midamble positions leads to the minimum CRB, in the case when $P \geq 2L+1$, there is no specification on how many clusters of those midamble pilots should have or how they should be placed, as long as the orthogonality condition $\mathbf{R}_{\mathbf{s}_p} = P\sigma_p^2\mathbf{I}$ is satisfied. Since the optimal placement mandates the first and last L symbols in $\mathbf{s}_p$ being zeros, we only consider the midamble pilot symbols $\{s_p[i]\}_{i=L+1}^{P-L}$, which are denoted as $\tilde{\mathbf{s}}_p$. Notice that there always exists $\tilde{\mathbf{s}}_p$ satisfying the orthogonality condition in (14)—an obvious choice is the δ-sequence $\tilde{\mathbf{s}}_p = [0, \ldots, 0, \sqrt{P}\sigma_p, 0, \ldots, 0]^T$. However, such a sequence may not be a desirable choice in practice; it requires transmitters to have high peak-to-average power ratio. The design of orthogonal sequences for a single cluster is also not trivial, and for general L and P, there may not exist constant modulus pilot symbols.

For multiple clusters, the orthogonality requirement involves the joint pilot symbols and cluster design. Unlike the single cluster case, it is easier to find pilot symbols and their placement satisfying the orthogonality condition. An interesting case is the placement using only one single pilot symbol in each

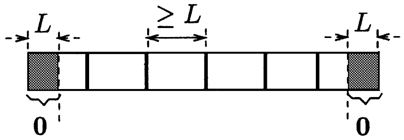


Fig. 4. Optimal multiple clusters placement scheme.

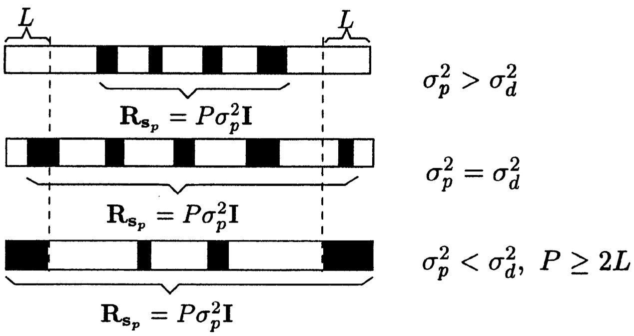


Fig. 5. Optimal placements of CM pilot symbols for SISO channels.

cluster, i.e., $\gamma_i = 1,\ i = 2,\dots,n-1$. In such a case, all the pilots in the midamble positions are at least L away from each other, as shown in Fig. 4. In this placement scheme, since each row of $\mathcal{H}(\bar{\mathbf{s}}_p)$ contains only one nonzero element, we have $\mathbf{R}_{\mathbf{s}_p} = P\sigma_p^2\mathbf{I}$, regardless of the values of these pilot symbols. Thus, the requirements for optimal placement and pilot design in part 1) of Theorem 2 are satisfied. This placement scheme is concluded in the following corollary.

Corollary 1: Assume $P \geq 2L+1$. Any $\mathbf{s}_p$ satisfying power constraint $\|\tilde{\mathbf{s}}_p\|^2 = P\sigma_p^2$ with $\mathcal{P} \in \{(\boldsymbol{\nu},\boldsymbol{\gamma}) : \nu_1 = \nu_{n+1} = 0; \nu_i \geq L, i = 3,\dots,n-1; \gamma_1 = \gamma_n = L; \gamma_i = 1, i = 2,\dots,n-1\}$ is optimal. Under the optimal $\mathbf{s}_{p_*}$ and $\mathcal{P}_*$, the minimum CRB is given in (15).

Although the optimal design benefits from the use of multiple clusters, existing estimation algorithms, on the other hand, favor single cluster placement. Multiple-cluster placement schemes, especially the scheme in Corollary 1, give an easy optimal design but make estimation harder. One expects such schemes to increase the difficulty and complexity in terms of channel estimation algorithms. Thus, a tradeoff between the choice of single cluster and multiple clusters exists.

B. Pilot Symbols With Constant Modulus Constraint

In many communication systems, pilot symbols with constant modulus (CM) property are used, i.e., $|s_p[i]|^2 = \sigma_p^2, i = 1,\dots,P$. We now consider the optimal placement and design of pilot symbols under such constraints following the same heuristic arguments. The optimal placements are illustrated in Fig. 5 and formally given in Theorem 3. For pilot symbols with sufficient power, placing pilot clusters all in the midamble positions leads to the lowest CRB. When pilot symbols have equal power to that of data symbols, optimal strategy is to design pilot symbols and placement jointly to satisfy the orthogonality condition. On the other hand, for sufficient amount of pilot symbols with low power, putting $2L$ pilot symbols at two edge parts leads to the lowest CRB.

Theorem 3: Given N data symbols with power σ_d^2 and P CM pilot symbols with $|s_p[i]|^2 = \sigma_p^2, i = 1,\dots,P$. Under assumption A1-A3, we have the following.

1) For $\sigma_p^2 > \sigma_d^2$, the placements and pilot sequences satisfying the following are optimal:

$$\mathcal{P}_* \in \{(\boldsymbol{\nu},\boldsymbol{\gamma}) : \nu_1, \nu_{n+1} \geq L\} \tag{22}$$
$$\mathbf{s}_{p_*} \in \{\mathbf{s}_p : \mathbf{R}_{\mathbf{s}_p} = P\sigma_p^2\mathbf{I}\}. \tag{23}$$

The minimum CRB is given by

$$\begin{aligned}\lambda_i(\mathcal{P}_*, \mathbf{s}_{p_*}) &= \min_{\mathcal{P},\mathbf{s}_p:|s_p[i]|^2=\sigma_p^2} \lambda_i(\mathcal{P},\mathbf{s}_p)\\ &= \frac{\sigma_n^2}{(N-L)\sigma_d^2 + P\sigma_p^2 + \rho_h^2\sigma_n^2}\\ &\quad i = 1,\dots,L+1.\end{aligned} \tag{24}$$

2) For $\sigma_p^2 = \sigma_d^2$, any placement with pilot symbols satisfying (23) is optimal. The minimum CRB takes the same formula as in (24).

3) For $\sigma_p^2 < \sigma_d^2$, if $P \geq 2L$, the placements and pilot sequences satisfying the following are optimal:

$$\mathcal{P}_* \in \{(\boldsymbol{\nu},\boldsymbol{\gamma}) : n > 2; \nu_1 = \nu_{n+1} = 0; \gamma_1 = \gamma_n = L\} \tag{25}$$
$$\mathbf{s}_{p_*} \in \{\mathbf{s}_p : \mathbf{R}_{\mathbf{s}_p} = P\sigma_p^2\mathbf{I}\}. \tag{26}$$

The minimum CRB is given by

$$\lambda_i(\mathcal{P}_*, \mathbf{s}_{p_*}) = \frac{\sigma_n^2}{N\sigma_d^2 + (P-L)\sigma_p^2 + \rho_h^2\sigma_n^2}. \tag{27}$$

Proof: See Appendix D.

For pilot symbols with $P < 2L$, the optimal placement is more complicated and varies with L and P. Due to the CM constraint, putting all pilot symbols into two clusters at two ends of a packet cannot satisfy the orthogonality requirement. Thus, this scheme is not guaranteed to be optimal. However, finding a placement to make $\mathbf{R}_{\mathbf{s}_p}$ a multiple of identity does not ensure the optimality over all possible pilot symbol placements and designs. Thus, an exhaustive search among all these possible placements may be necessary to achieve the minimum CRB.

Design of Orthogonal Sequences—Single Cluster vs. Multiple Clusters: For $\sigma_p^2 \geq \sigma_d^2$, we again encounter the problem of choosing between the single cluster or multiple clusters. As discussed earlier, the use of multiple clusters makes the orthogonality condition easy to satisfy. An interesting simple optimal placement where pilot symbols are scattered throughout the packet is shown in Fig. 6 and described by $\mathcal{P}_* \in \{(\boldsymbol{\nu},\boldsymbol{\gamma}) : n = P; \nu_i \geq L; \gamma_i = 1\}$. The actual values of CM pilot symbols are nonessential, provided that they satisfy the power constraint.

A generalization of the above scheme is the so-called QPP in a certain sense. In a QPP-α scheme, under the constraint that each pilot cluster length is no less than α, the pilot symbols are divided into as many clusters as possible. These clusters are then placed such that the data block lengths are as equal as possible. The QPP family is divided into different classes. The class

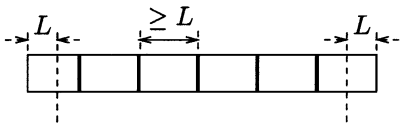


Fig. 6. Optimal multicluster placement scheme of CM pilot symbols.

of schemes for which α is the smallest allowable pilot symbol cluster length is denoted as QPP-α.

Definition 1: Given an α and a frame with N unknown symbols and P known symbols, let $J = \lfloor P/\alpha \rfloor + 1$. A placement scheme $\mathcal{P}(\boldsymbol{\nu}, \boldsymbol{\gamma})$ belongs to QPP-α if and only if

1) $\gamma_i \geq \alpha$, and $\sum_i^n \gamma_i = P$;
2) $\nu_i \in \{(\lfloor N/J \rfloor), (\lfloor N/J \rfloor + 1)\}$, and $\sum_i^{n+1} \nu_i = N$.

The QPP is a family of placement strategies that is shown to be optimal in the sense of maximizing mutual information . It is also optimal in the sense of minimizing average MSE associated with transmitted symbol when a decision feedback equalizer is used, again, assuming known channel [1]. In terms of minimizing the lower bound for channel estimation, it turns out that the orthogonality constraint on pilot symbols for QPP-1 is the easiest to satisfy. For $\sigma_p^2 \geq \sigma_d^2$, the scheme described in Fig. 6 indicates that if $N/(P+1) \geq L$, then QPP-1 is also optimal for channel estimation.

V. Optimal Design and Placement for MIMO Channels

A. Model

A multiuser channel can be modeled as a K-input M-output FIR linear system. The system inputs correspond to packets from K users, and the outputs come from M diversity channels that may result from temporal sampling or antenna array, etc. Denote L_k as the channel order for the kth user and $\mathbf{h}^{(k)}[i] \in \mathcal{C}^{M\times 1}, i = 0, \ldots, L_k$ as the channel impulse response vector for the single-input multiple-output channel between the kth user and the received M-dimensional data vector $\mathbf{y}_t$. The MIMO channel can then be described by

$$\mathbf{y}_t \triangleq [y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(M)}]^T = \sum_{k=1}^{K} \sum_{i=0}^{L_k} \mathbf{h}^{(k)}[i] s_{t-i}^{(k)} + \mathbf{n}_t \quad (28)$$

where the data packet from user k is denoted by $\mathbf{s}^{(k)}$, and $\boldsymbol{\mathcal{H}}(\mathbf{s}^{(k)})$, the correspoding input symbol matrix from user k, is defined the same as in (4). Stacking the corresponding vectors $\mathbf{s}^{(k)}$, $\mathbf{h}_k[i]$, respectively, we have the vector model

$$\mathbf{y} = \mathcal{F}(\mathbf{s})\mathbf{h} + \mathbf{n} \quad (29)$$

where $\mathcal{F}(\mathbf{s})$ is the overall input symbol matrix including both data and pilots from all the K users

$$\mathcal{F}(\mathbf{s}) \triangleq [\mathcal{H}(\mathbf{s}^{(1)}), \ldots, \mathcal{H}(\mathbf{s}^{(K)})] \otimes \mathbf{I}_M. \quad (30)$$

Let N_k, P_k be the number of data and pilot symbols of each packet from user k, respectively. The pilot symbols from the kth user is denoted by $\mathbf{s}_p^{(k)}$, and $\mathbf{s}_p = [\mathbf{s}_p^{(1)}, \ldots, \mathbf{s}_p^{(K)}]^T$ is the total pilot symbols from K users. In addition to A1-A3, we assume the following.

A4) The packet transmission system is slotted, i.e., for each time slot, each user transmits one packet through the channels.

A5) $P_k\sigma_{p_k}^2 > \sigma_{d_k}^2$, for $k = 1, \ldots, K$, where $\sigma_{p_k}^2$ and $\sigma_{d_k}^2$ are the pilot and data power for user k.

Assumption A4 ensures that channel estimation is performed within K transmitted packets: one from each user. Assumption A5 is introduced primarily because sufficient pilot power is generally guaranteed in communication systems.

B. Optimal Placement

In this section, we consider the optimal pilot design and placement for packet transmissions involving K users. Allowing pilot symbols to be placed independently for each user and assuming the number of pilot clusters for user k is n_k, the placement $\mathcal{P} = (\boldsymbol{\nu}, \boldsymbol{\gamma})$ is defined by $\boldsymbol{\nu} = [\boldsymbol{\nu}^{(1)}, \ldots, \boldsymbol{\nu}^{(K)}], \boldsymbol{\gamma} = [\boldsymbol{\gamma}^{(1)}, \ldots, \boldsymbol{\gamma}^{(K)}]$, where ($\boldsymbol{\nu}^{(k)}$, $\boldsymbol{\gamma}^{(k)}$) is the placement for user k. Given a placement $\mathcal{P}$ and the decomposition $\mathbf{s} = \bar{\mathbf{s}}_p + \bar{\mathbf{s}}_d$ as in (5), the "autocorrelation" matrices associated with input symbols and pilot symbols are defined by

$$\mathcal{R}_{\mathbf{s}} \triangleq \mathcal{F}^H(\mathbf{s})\mathcal{F}(\mathbf{s}), \mathcal{R}_{\mathbf{s}_p} \triangleq \mathcal{F}^H(\bar{\mathbf{s}}_p)\mathcal{F}(\bar{\mathbf{s}}_p) \quad (31)$$

where $\mathcal{R}_{\mathbf{s}}$ is, again, a function of $\mathcal{P}$ and $\mathbf{s}_p$.

Extending from Theorem 1 in the SISO model, the CRB for channel estimators under the MIMO model is given by

$$\mathcal{M}(\hat{\mathbf{h}}) \geq \boldsymbol{\Lambda}(\mathcal{P}, \mathbf{s}_p) \triangleq \left[\frac{1}{\sigma_n^2} E\{\mathcal{R}_{\mathbf{s}}\} + \text{diag}(\rho_{h_1}^2 \mathbf{I}_{L_1}, \ldots, \rho_{h_K}^2 \mathbf{I}_{L_K}) \otimes \mathbf{I}_M\right]^{-1}. \quad (32)$$

Again, the CRB has the shift invariant property in midamble positions.

Lemma 2: For any $\mathbf{s}_p^{(k)}$ with power $\sigma_{p_k}^2$, the midamble placement is shift invariant, i.e., for any $\mathcal{P} \in \{(\boldsymbol{\nu}, \boldsymbol{\gamma}) : \nu_1^{(k)} \geq L_k, \nu_{n_k+1}^{(k)} \geq L_k + 1, k = 1, \ldots, K\}$

$$\boldsymbol{\Lambda}(\boldsymbol{\nu}, \boldsymbol{\gamma}, \mathbf{s}_p) = \boldsymbol{\Lambda}(\boldsymbol{\nu}', \boldsymbol{\gamma}, \mathbf{s}_p) \quad (33)$$

where $\boldsymbol{\nu}'$ is related to $\boldsymbol{\nu}$ by

$$\boldsymbol{\nu}'^{(k)} = \boldsymbol{\nu}^{(k)} + \mathbf{e}_1^{(k)} - \mathbf{e}_{n_k+1}^{(k)}, k = 1, \ldots, K \quad (34)$$

where $\mathbf{e}_1^{(k)}$ and $\mathbf{e}_{n_k+1}^{(k)}$ are the unit vector for the kth user.

Proof: The proof is similar to the one for Lemma 1.

The optimal designs and placements in a two-user case is illustrated in Fig. 7, and that for the general MIMO channel is described in Theorem 4, where it indicates that under the MIMO model, the optimal placements are, again, independent of probability distributions of channels. The theorem concludes that within a packet from each user, the single-user optimal placement strategies should be used. Furthermore, the optimal placement involves the orthogonality design of pilot symbols among all users. For $P_k \leq 2L_k$, which indicates very small amount of pilot symbols, this orthogonality condition can be easily satisfied by the optimal placement described in Theorem 4. For the

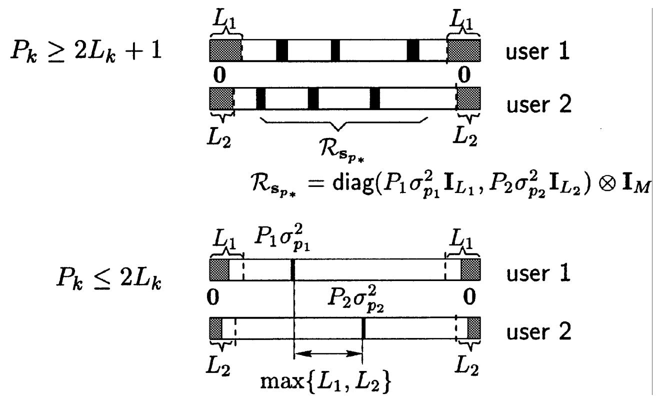


Fig. 7. Optimal placements of pilot symbols for the MIMO channel.

other case when $P_k \geq 2L_k + 1$, which is usually satisfied, the optimal design is nontrivial in general and is discussed in Section V-C

Theorem 4: Given N_k data symbols with power $\sigma_{d_k}^2$ and P_k pilot symbols with power $\sigma_{p_k}^2$ from the kth user, $k = 1, \ldots, K$. Let $\lambda_i^{(k)}(\mathcal{P}, \mathbf{s}_p)$ be the CRB for the ith channel coefficient of user k. Under assumptions A1-A4, we have the following.

1) For $P_k \geq 2L_k + 1$, $k = 1, \ldots, K$, the optimal placement $\mathcal{P}_*$ and design of pilot symbols $\mathbf{s}_{p_*}$ are given by

$$\begin{aligned}\mathcal{P}_* \in &\{(\boldsymbol{\nu}, \boldsymbol{\gamma}) : n_k > 2; \nu_1^{(k)} = \nu_{n_k+1}^{(k)} = 0 \\ &\gamma_1^{(k)} = \gamma_{n_k}^{(k)} = L_k; \forall k\}\end{aligned} \tag{35}$$

$$\begin{aligned}\mathbf{s}_{p_*} \in &\{\mathbf{s}_p : s_p^{(k)}[i] = 0, \forall i \in [1, L_k] \cup [P_k - L_k + 1, P_k] \\ &\mathcal{R}_{\mathbf{s}_p} = \text{diag}(P_1\sigma_{p_1}^2 \mathbf{I}_{L_1}, \ldots, P_K\sigma_{p_K}^2 \mathbf{I}_{L_K}) \otimes \mathbf{I}_M; \forall k\}.\end{aligned} \tag{36}$$

The minimum CRB is given by

$$\begin{aligned}\lambda_i^{(k)}(\mathcal{P}_*, \mathbf{s}_{p_*}) &= \min_{\substack{\mathcal{P}, \mathbf{s}_p : \|\mathbf{s}_p^{(k)}\|^2 = P_k\sigma_{p_k}^2, \\ 1 \leq k \leq K}} \lambda_i^{(k)}(\mathcal{P}, \mathbf{s}_p) \\ &= \frac{\sigma_n^2}{N_k\sigma_{d_k}^2 + P_k\sigma_{p_k}^2 + \rho_{h_k}^2\sigma_n^2} \\ &\quad i = 1, \ldots, (L_k + 1)M, k = 1, \ldots, K.\end{aligned} \tag{37}$$

2) For $P_k \leq 2L_k$, $k = 1, \ldots, K$, under assumption A5, the optimal placement $\mathcal{P}_*$ and design of pilot symbols $\mathbf{s}_{p_*}$ are given by

$$\begin{aligned}\mathcal{P}_* \in \Big\{&(\boldsymbol{\nu}, \boldsymbol{\gamma}) : n_k = 3; \nu_1^{(k)} = \nu_4^{(k)} = 0 \\ &|(\nu_2^{(k)} + \gamma_2^{(k)}) - (\nu_2^{(j)} + \gamma_2^{(j)})| \geq \max_k\{L_k\}, \forall j \neq k \\ &\gamma_2^{(k)} = 1, \gamma_1^{(k)}, \gamma_3^{(k)} \in \left\{\left\lfloor \frac{P_k - 1}{2} \right\rfloor, \left\lceil \frac{P_k - 1}{2} \right\rceil\right\}; \forall k\Big\}\end{aligned} \tag{38}$$

$$\mathbf{s}_{p_*}^{(k)}[i] = \begin{cases} P_k\sigma_{pk}^2, & \text{if } i = \gamma_1^{(k)} + 1 \\ 0, & \text{otherwise}\end{cases} \quad k = 1, \ldots, K. \tag{39}$$

The minimum CRB is given by

$$\begin{aligned}&\text{tr}[\boldsymbol{\Lambda}(\mathcal{P}_*, \mathbf{s}_{p_*})] \\ &= \min_{\substack{\mathcal{P}, \mathbf{s}_p : \|\mathbf{s}_p^{(k)}\|^2 = P_k\sigma_{p_k}^2, \\ 1 \leq k \leq K}} \text{tr}[\boldsymbol{\Lambda}(\mathcal{P}, \mathbf{s}_p)] \\ &= M\sum_{k=1}^{K}\left\{\sum_{i=\max\{L_k - P_k + 2, 1\}}^{L_k - \lceil (P_k - 1)/2 \rceil} \frac{2\sigma_n^2}{(N_k - i)\sigma_{d_k}^2 + P_k\sigma_{p_k}^2 + \rho_{h_k}^2\sigma_n^2}\right. \\ &\quad + \frac{(L_k - P_k + 2)\sigma_n^2\mathcal{U}(L_k - P_k)}{(N_k + P_k - L_k - 1)\sigma_{d_k}^2 + P_k\sigma_{p_k}^2 + \rho_{h_k}^2\sigma_n^2} \\ &\quad + \frac{(\lceil \frac{P_k - 1}{2} \rceil - \lfloor \frac{P_k - 1}{2} \rfloor)\sigma_n^2}{(N_k - L_k + \lfloor \frac{P_k - 1}{2} \rfloor)\sigma_{d_k}^2 + P_k\sigma_{p_k}^2 + \rho_{h_k}^2\sigma_n^2} \\ &\quad \left. + \frac{\max\{P_k - L_k, 0\}\sigma_n^2}{N_k\sigma_{d_k}^2 + P_k\sigma_{p_k}^2 + \rho_{h_k}^2\sigma_n^2}\right\}\end{aligned} \tag{40}$$

where $\mathcal{U}(k)$ is the unit step function.

Proof: See Appendix E.

C. Multiuser Placement Strategies

In this section, we only consider the case when $P_k \geq 2L_k + 1$. As discussed in Section IV, it is difficult to design orthogonal sequences for those pilot symbols in the midamble positions if they are grouped in a single pilot cluster. Multiple clusters should be considered. Then, the next question follows: Should we align pilot clusters from each user at the same position? Theorem 4 indicates that as long as the pilots between users are orthogonal, the placement is still optimal. However, by doing this, we should consider all K pilot sequences jointly, which increases the difficulty of the sequence design. An easy way to simplify the problem is to place the pilot clusters staggered among users. As an example shown in Fig. 8, two users are present in the system. Clusters from users 1 and 2 are offset to each other so that $\mathcal{R}_{\mathbf{s}_p}$ in (31) is block diagonal. The orthogonality condition between users are automatically satisfied. Note that the pilot sequence design can now be done independently. Moreover, smaller cluster size also simplifies the pilot design.

Furthermore, the easiest way to satisfy the orthogonality condition, perhaps, is the scheme described in Fig. 4 extended for the multiple-user case as illustrated in Fig. 9. The lowest CRBs can be obtained if pilot symbols in the midamble positions are scattered in such a way that they are at least $\max_k\{L_k\}$ apart. By this way, the actual values of pilot symbols are nonessential as long as they satisfy the power constraint.

D. Pilot Symbols With Constant Modulus Constraint

Consider all K users using CM pilot symbols, i.e., $|s_p^{(k)}[i]|^2 = \sigma_{p_k}^2$, for all k. The optimal placements are illustrated in a two-user case in Fig. 10 and described in the following theorem.

Theorem 5: Assume CM pilot symbols. Under assumption A1-A4, we have the following.

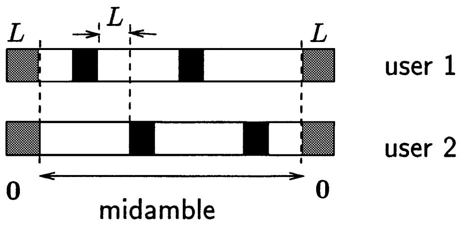


Fig. 8. Optimal placement for two-user case.

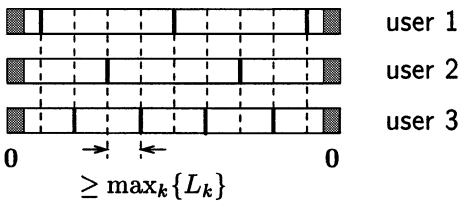


Fig. 9. Optimal placement for three-user case.

1) For $\sigma^2_{p_k} > \sigma^2_{d_k}, k = 1, \ldots, K$, the optimal placement $\mathcal{P}_*$ and design of pilot symbols $\mathbf{s}_{p_*}$ are given by

$$\mathcal{P}_* \in \{(\boldsymbol{\nu}, \boldsymbol{\gamma}) : \nu_1^{(k)}, \nu_{n_k+1}^{(k)} \geq L_k, \forall k\} \tag{41}$$

$$\mathbf{s}_{p_*} \in \{\mathbf{s}_p : \mathcal{R}_{\mathbf{s}_{p_*}} = \mathrm{diag}(P_1\sigma^2_{p_1}\mathbf{I}_{L_1}, \ldots, P_K\sigma^2_{p_K}\mathbf{I}_{L_K}) \otimes \mathbf{I}_M\}. \tag{42}$$

The minimum CRB is given by

$$\lambda_i^{(k)}(\mathcal{P}_*, \mathbf{s}_{p_*}) = \min_{\substack{\mathcal{P}, \mathbf{s}_p : |s_p^{(k)}[i]|^2 = \sigma^2_{p_k} \\ 1 \leq k \leq K}} \lambda_i^{(k)}(\mathcal{P}, \mathbf{s}_p) = \frac{\sigma_n^2}{(N_k - L_k)\sigma^2_{d_k} + P_k\sigma^2_{p_k} + \rho^2_{h_k}\sigma^2_n}$$
$$i = 1, \ldots, (L_k+1)M, k = 1, \ldots, K. \tag{43}$$

2) For $\sigma^2_{p_k} = \sigma^2_{d_k}, k = 1, \ldots, K$, any placement with pilot symbols satisfying (42) is optimal. The minimum CRB takes the same formula in (43).

Proof: The proof is similar to the one in Theorem 3.

For $\sigma^2_{p_k} < \sigma^2_{d_k}$, the optimal placement should satisfy the single-user optimal placement requirement and, at the same time, satisfies cross-user orthogonality condition. However, such pilot sequences satisfying both conditions might not exist. Thus, finding an optimal placement scheme may follow an exhaustive search among all possible placements. The resulting optimal placement then depends on each specific situation.

For sufficient pilot power, the easiest scheme perhaps is QPP-1 scheme extended for multiple users as illustrated in Fig. 11, which can be summarized by the following corollary.

Corollary 2: For any $\mathbf{s}_p$ satisfying $\sigma^2_{p_k} \geq \sigma^2_{d_k}$, $k = 1, \ldots, K$, the placement $\mathcal{P}_* = (\boldsymbol{\nu}, \mathbf{1})$ is optimal if $\boldsymbol{\nu}$ satisfies

1) $\nu_1^{(k)}, \nu_{n_k+1}^{(k)} \geq L_k, \forall k$;
2) $\min_{m,j,k,l} |\sum_{i=1}^m (\nu_i^{(k)} + 1) - \sum_{i=1}^j (\nu_i^{(l)} + 1)| \geq \max_k\{L_k\}$.

Although the extended QPP-1 scheme is the easiest to satisfy the orthogonality condition, it requires each user's packet length to be sufficiently long to allow pilot symbols to scatter out. On the other hand, the longer pilot cluster length, the shorter packet size is required; this is especially tractable for short packet communication scenarios. Therefore, there again exists a tradeoff between the choice of short and long pilot clusters.

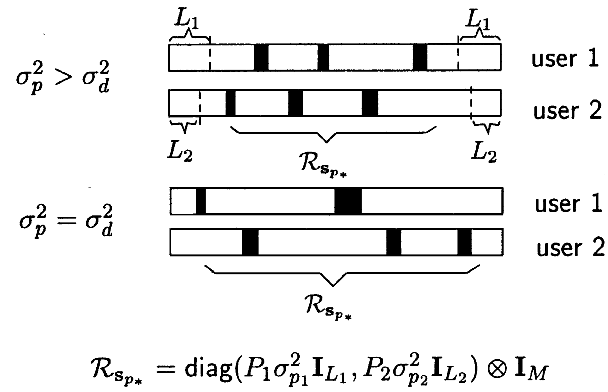


Fig. 10. Optimal placements of CM pilot symbols for the MIMO channel.

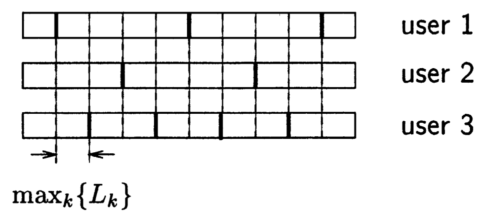


Fig. 11. Optimal QPP-1 placement for three users.

VI. Placement for Channels With Correlated Taps

In the previous sections, we discussed the design and placement of pilot symbols for random channels with taps being i.i.d. In this section, we look into a more general case where channel taps are correlated. The SISO channel model is considered. Specifically, the channel is the combination of the pulse shaping filter and propagation channel. Although the propagation channel appears random changes from packet to packet, due to the pulse shaping filter, channel taps are correlated to each other in general. Thus, channel $\mathbf{h}$ appears random but is restricted within a certain subspace. Therefore, assuming A1 and A3, we relax A2 to the following assumption.

A2') The channel $\mathbf{h}$ can be represented by

$$\mathbf{h} = \mathbf{G}\mathbf{v} \tag{44}$$

where $\mathbf{G} = (\mathbf{g}_1, \ldots, \mathbf{g}_r) \in \mathcal{C}^{(L+1)\times r}$ has r orthonormal columns, and vector $\mathbf{v}$ consists of r i.i.d. zero mean random variables with pdf $p_v(\cdot)$ and variance σ_v^2.

When $r < L+1$, channel taps are correlated with covariance $E[\mathbf{h}\mathbf{h}^H] = \sigma_v^2\mathbf{G}\mathbf{G}^H$. In the special case when $r = L+1$, assumption A2' reduces to A2.

A. CRB

The complex CRB for transformations of deterministic parameters was derived in [19]. For linear transformation of random parameters, the complex CRB becomes

$$\boldsymbol{\Lambda}_{\mathbf{f}} = E\left[\frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}}\right] \boldsymbol{\Lambda}_{\boldsymbol{\theta}} E\left[\frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}}\right]^H \tag{45}$$

where $\boldsymbol{\theta}$ is a random parameter vector, and $\mathbf{f} = \mathbf{A}\boldsymbol{\theta}$; $\boldsymbol{\Lambda}_{\boldsymbol{\theta}}$ is the CRB for $\boldsymbol{\theta}$. Given the channel model in (44), the complex CRB for channel estimators is then derived as

$$\boldsymbol{\Lambda}(\mathcal{P}, \mathbf{s}_p) = \mathbf{G}\left(\frac{1}{\sigma_n^2}\mathbf{G}^H E\{\mathbf{R}\mathbf{s}\}\mathbf{G} + \rho_v^2\mathbf{I}\right)^{-1}\mathbf{G}^H. \quad (46)$$

B. Placement

It is not hard to see that for general random channels, Lemma 1 still holds. Our objective is to find $\mathcal{P}_*$ and $\mathbf{s}_{p_*}$ among all orthogonal pilot sequences such that

$$(\mathcal{P}_*, \mathbf{s}_{p_*}) = \mathop{\arg\min}_{\substack{\mathcal{P},\mathbf{s}_p: \|\mathbf{s}_p\|^2 = P\sigma_p^2 \\ \mathbf{R}_{\mathbf{s}_p}\,\text{diagonal}}} \text{tr}[\boldsymbol{\Lambda}(\mathcal{P}, \mathbf{s}_p)]. \quad (47)$$

Theorem 6: Assume $P \geq 2L + 1$. Among all orthogonal pilot sequences, i.e., $\{\mathbf{s}_p : \mathbf{R}_{\mathbf{s}_p} \text{diagonal}\}$, the optimal placement is given by

$$\begin{aligned} \mathcal{P}_* \in &\{(\boldsymbol{\nu}, \boldsymbol{\gamma}) : n > 2; \nu_1 = \nu_{n+1} = 0 \\ &\gamma_1 = \gamma_n = L\} \quad (48) \\ \mathbf{s}_{p_*} \in &\{\mathbf{s}_p : s_p[i] = 0, \forall i \in [1, L] \cup [P - L + 1, P] \\ &\mathbf{R}_{\mathbf{s}_p} = P\sigma_p^2\mathbf{I}\}. \quad (49) \end{aligned}$$

The minimum CRB is given by

$$\text{tr}[\boldsymbol{\Lambda}(\mathcal{P}_*, \mathbf{s}_{p_*})] = \frac{r\sigma_n^2}{N\sigma_d^2 + P\sigma_p^2 + \rho_v^2\sigma_n^2}. \quad (50)$$

Proof: See Appendix F.

For pilot symbols with CM property, following Theorem 6 and using the similar proof, we see that when $\sigma_p^2 > \sigma_d^2$, among all orthogonal pilot sequences, placing all pilot symbols in the midamble positions is optimal:

$$\mathcal{P}_* \in \{(\boldsymbol{\nu}, \boldsymbol{\gamma}) : \nu_1, \nu_{n+1} \geq L\} \quad (51)$$

and the minimum CRB is given by

$$\text{tr}[\boldsymbol{\Lambda}(\mathcal{P}_*, \mathbf{s}_{p_*})] = \frac{r\sigma_n^2}{(N - L)\sigma_d^2 + P\sigma_p^2 + \rho_v^2\sigma_n^2}. \quad (52)$$

When $\sigma_p^2 = \sigma_d^2$, $\boldsymbol{\Lambda}(\mathcal{P}, \mathbf{s}_p)$ is invariant under $\mathcal{P}$ among orthogonal sequences, i.e., different placements result in equal performance.

All the above show that the optimal placement strategy for channels with i.i.d. taps is also optimal, among all orthogonal pilot sequences, for channels with correlated taps. Note that our results of optimal placements are confined in searching among all possible orthogonal pilot sequences. It does not imply that this placement minimizes $\text{tr}[\boldsymbol{\Lambda}(\mathcal{P}, \mathbf{s}_p)]$ for all choices of pilot sequences. Indeed, in general, the placement that gives the minimum $\text{tr}[\boldsymbol{\Lambda}(\mathcal{P}, \mathbf{s}_p)]$ depends heavily on $\mathbf{G}$ and each specific realization of pilot sequences.

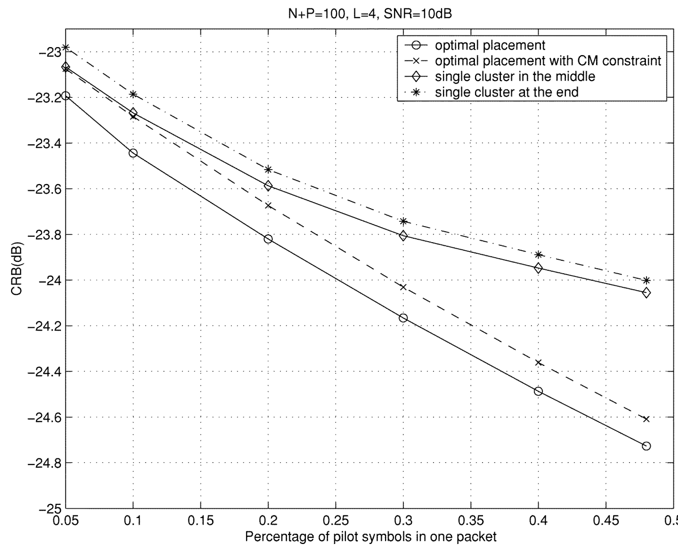
VII. Numerical Results

A. Placement Schemes in Single User Case

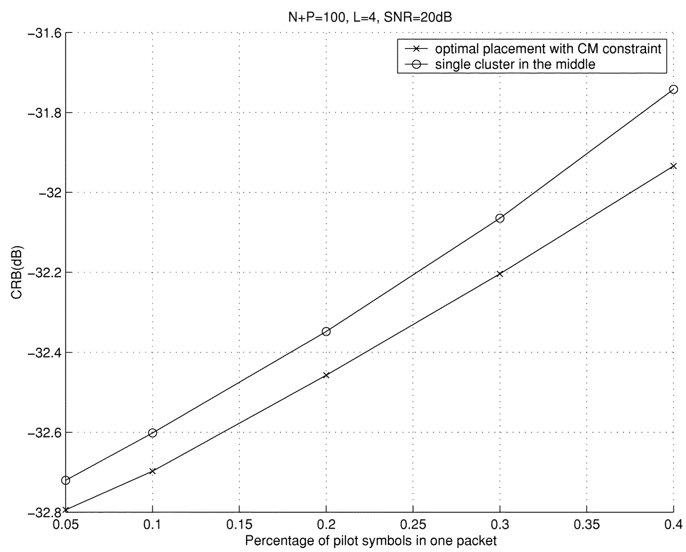
We first compared the CRBs of channel estimators under optimal and nonoptimal pilot design and placement schemes in the SISO model. Channel order was $L = 4$. We assumed that channel taps are i.i.d. complex Gaussian with zero mean and variance $\sigma_h^2 = 1/(L+1) = 0.2$. The data packet length was 100. Data and pilot powers were $\sigma_d^2 = 1$ and $\sigma_p^2 = 2$, respectively. Four placement schemes were considered:

1) the optimal placement allowing power allocation;
2) the optimal placement for pilot symbols with CM constraint;
3) a single cluster with CM pilot symbols used in 2) placed in the middle of the packet and the pilot sequence violated the orthogonality requirement;
4) the same single cluster placed at one end of packet.

In the first optimal scheme, we used the placement described in Corollary 1. For the second one with CM constraint, QPP-3 placement was used with each pilot cluster being $[\sqrt{2}, \sqrt{2}, -\sqrt{2}]$. Fig. 12(a) shows the traces of CRBs of these four schemes under increasing percentage of pilot symbols per packet at SNR $= 10$ dB. We observe that the gain of the



(a)



(b)

Fig. 12. (a) CRBs of different placements versus percentage of pilot symbols at SNR $= 10$ dB. (b) CRBs versus different placements under low pilot power at SNR $= 20$ dB.

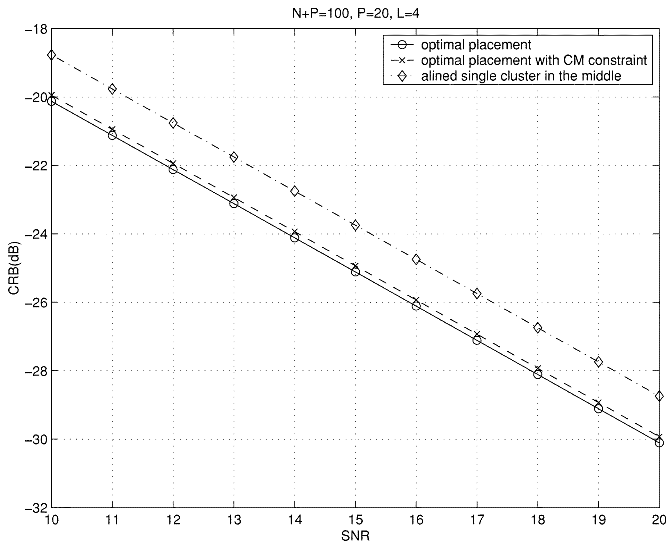


Fig. 13. CRBs of different placement schemes in a two-user case.

optimal scheme increases with increasing percentage of pilot symbols.

Finally, for pilot symbols with CM constraint, we give an example when pilot symbols with low power $\sigma_p^2 = 0.5\sigma_d^2$. Two schemes were compared: 1) optimal placement for pilot symbols with CM constraint and 2) single cluster with the same pilot sequence placed in the middle of the packet. Fig. 12(b) plots the CRBs versus the percentage of pilot symbols at $SNR = 20$ dB. We can see that in this case, putting pilot symbols at two ends of the frame resulted in lower CRBs. Notice that because the total power from data and pilot symbols decreases with the percentage of pilot symbols increasing, the corresponding CRBs increases.

B. Placement Schemes in Multiple-User Case

We next consider the placements in the multiuser case, where $K = M = 2$. The channels, data, and pilot powers used were still the same as in the single-user case. Two users were considered with the same packet length of 100. Each packet consisted of 20 pilot symbols. Three schemes were compared:

1) optimal placement;
2) optimal placement under the CM constraint;
3) conventional single cluster with the CM pilot symbols used in 2) aligned in the middle of the packet from each user.

Again, for the first scheme, pilot symbols in the midamble positions were placed similarly as in Fig. 9. For the second scheme, we used the QPP-3 scheme with pilot clusters shifted between users, which is similar as in Fig. 11. Fig. 13 shows the trace of the CRBs under these threes scenarios. We observe that about a 1.5-dB gain is obtained by placing pilot symbols optimally. This shows that the importance of optimal placement in the multiuser case is more significant than that in the single-user case. Note also that there is little performance loss by imposing the CM constraint on pilot symbols.

VIII. Conclusion

In this paper, we presented the optimization of the placement and design of pilot symbols for semi-blind channel estimation. We have shown that the CRB is shift-invariant among midamble positions, and the basic principle of optimal placements is to concentrate higher power symbols in the midamble positions of a packet while placing symbols with lower power at two ends. Our results also indicate that the optimal placements are independent of any channel distribution. While the merit of placing pilot symbols in the middle of a packet is justified by our theory, we found many other placements that are also optimal. Among those, the use of multiple clusters makes the design of optimal pilot sequence simpler. However, placing pilot symbols in multiple clusters may increase the complexity of channel estimation.

We noticed that under the SISO model, the difference of CRBs between optimal and nonoptimal placements does not appear to be significant. Therefore, more consideration should be given to the placement design for optimal detection performance. It is reassuring to find that the QPP-α scheme for the detection and maximizing channel capacity is, in our results, also optimal for channel estimation. Under the MIMO model, as the number of users in the system increases, much can be gained from the optimal placements.

Finally, we note that a pilot placement may have effects on the estimator performance that are different from that on the CRB. The optimality of placements for a specific class of estimations may also be of interest.

Appendix A
Proof of Theorem 1

Let $\boldsymbol{\theta} = [\mathbf{s}_d^H, \mathbf{h}^H]^H$. Under the regularity conditions [19], , the MSE matrix of any estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$ is lower bounded by

$$\mathcal{M}(\hat{\boldsymbol{\theta}}) \stackrel{\Delta}{=} E\{[\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta}][\hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta}]^H\} \geq \mathbf{J}^{-1}$$

with the complex Fisher information matrix (FIM) $\mathbf{J}$ defined as

$$\mathbf{J} = E\left\{\left[\frac{\partial \ln p(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]\left[\frac{\partial \ln p(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]^H\right\} \tag{53}$$

where $p(\mathbf{y},\boldsymbol{\theta})$ is the joint distribution of $\mathbf{y}$ and $\boldsymbol{\theta}$, and the expectation is taken over $\boldsymbol{\theta}$ and $\mathbf{y}$.

Under the regularity condition, we have

$$\begin{aligned} &E\left\{\left[\frac{\partial \ln p(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]\left[\frac{\partial \ln p(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]^H |\boldsymbol{\theta}\right\} \\ &=E\left\{\left[\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^*}\right]\left[\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^*}\right]^H |\boldsymbol{\theta}\right\} \\ &\quad + \left[\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]\left[\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]^H \\ &=\frac{1}{\sigma_n^2}\begin{pmatrix} \mathbf{H}_d^H\mathbf{H}_d & \mathbf{H}_d^H\mathcal{H}(\mathbf{s}) \\ \mathcal{H}(\mathbf{s})^H\mathbf{H}_d & \mathcal{H}(\mathbf{s})^H\mathcal{H}(\mathbf{s}) \end{pmatrix} \\ &\quad + \left[\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]\left[\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]^H \end{aligned} \tag{54}$$

where $\mathbf{H}_d$ is obtained from $\mathcal{T}(\mathbf{h})$ by deleting columns corresponding to pilot symbols.

By assumptions of $\mathbf{h}$ and $\mathbf{s}_d$ in A1-A3, we have

$$E\{\mathbf{H}_d^H\mathcal{H}(\mathbf{s})\} = E\{\mathbf{H}_d\}^H E\{\mathcal{H}(\mathbf{s})\} = \mathbf{0}. \quad (55)$$

Now, we can obtain the expression of the FIM

$$\begin{aligned}\mathbf{J} =& E\left\{E\left\{\left[\frac{\partial \ln p(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]\left[\frac{\partial \ln p(\mathbf{y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]^H |\boldsymbol{\theta}\right\}\right\}\\ =& \begin{pmatrix} \frac{1}{\sigma_n^2}E\{\mathbf{H}_d^H\mathbf{H}_d\} & \mathbf{0}\\ \mathbf{0} & \frac{1}{\sigma_n^2}E\{\mathbf{Rs}\}\end{pmatrix}\\ &+ E\left\{\left[\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]\left[\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]^H\right\}. \end{aligned} \quad (56)$$

Under the regularity conditions and assumptions A1-A3, since $p(\boldsymbol{\theta}) = p_{\mathbf{s}}(\mathbf{s}_d)p_{\mathbf{h}}(\mathbf{h})$, where $p_{\mathbf{s}}(\mathbf{s}_d)$ and $p_{\mathbf{h}}(\mathbf{h})$ are joint pdf of $\mathbf{s}_d$ and $\mathbf{h}$, respectively, the second term in (56) becomes

$$\begin{aligned}&E\left\{\left[\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]\left[\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*}\right]^H\right\}\\ &= \begin{pmatrix} E\{\frac{\partial \ln p_{\mathbf{s}}(\mathbf{s}_d)}{\partial \mathbf{s}_d^*}\frac{\partial^H \ln p_{\mathbf{s}}(\mathbf{s}_d)}{\partial \mathbf{s}_d^*}\} & \mathbf{0}\\ \mathbf{0} & E\{\frac{\partial \ln p_{\mathbf{h}}(\mathbf{h})}{\partial \mathbf{h}^*}\frac{\partial^H \ln p_{\mathbf{h}}(\mathbf{h})}{\partial \mathbf{h}^*}\}\end{pmatrix}\\ &= \begin{pmatrix}\rho_s^2\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \rho_h^2\mathbf{I}\end{pmatrix} \end{aligned} \quad (57)$$

where ρ_s^2, ρ_h^2 are defined as

$$\rho_s^2 \triangleq E\left\{\left|\frac{\partial \ln p_s(s_d)}{\partial s_d^*}\right|^2\right\}, \rho_h^2 \triangleq E\left\{\left|\frac{\partial \ln p_h(h)}{\partial h^*}\right|^2\right\} \quad (58)$$

where the expectation is taken with respect to $p_s(s_d)$ and $p_h(h)$, respectively.

Therefore, the complex FIM is

$$\mathbf{J} = \begin{pmatrix} \frac{1}{\sigma_n^2}E\{\mathbf{H}_d^H\mathbf{H}_d\} + \rho_s^2\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \frac{1}{\sigma_n^2}E\{\mathbf{Rs}\} + \rho_h^2\mathbf{I}\end{pmatrix}. \quad (59)$$

Consequently

$$\begin{aligned}&\mathcal{M}(\hat{\boldsymbol{\theta}})\\ &\geq \begin{pmatrix} (\frac{1}{\sigma_n^2}E\{\mathbf{H}_d^H\mathbf{H}_d\} + \rho_s^2\mathbf{I})^{-1} & \mathbf{0}\\ \mathbf{0} & (\frac{1}{\sigma_n^2}E\{\mathbf{Rs}\} + \rho_h^2\mathbf{I})^{-1}\end{pmatrix}.\end{aligned}$$

Notice that the FIM for $\boldsymbol{\theta}$ is block diagonal, and CRBs for the channel and data symbols are decoupled. The complex CRB for the MSE of channel estimators is then given by

$$\mathcal{M}(\hat{\mathbf{h}}) \geq \left(\frac{1}{\sigma_n^2}E\{\mathbf{Rs}\} + \rho_h^2\mathbf{I}\right)^{-1} \triangleq \boldsymbol{\Lambda}(\mathcal{P}, \mathbf{s}_p). \quad (60)$$

□

APPENDIX B
PROOF OF LEMMA 1

When $\nu_1, \nu_{n+1} \geq L$, all P pilot symbols are placed in the midamble positions. For the n-cluster case, $\mathbf{R}_{\mathbf{s}_p}$ defined in (7) is a function of $\boldsymbol{\nu}$ and $\boldsymbol{\gamma}$, which is denoted as $\mathbf{R}_{\mathbf{s}_p}(\boldsymbol{\nu},\boldsymbol{\gamma})$. Since $\{s_d[i]\}_{i=1}^P$ is i.i.d. with zero mean, we have

$$\begin{aligned}E\{\mathbf{R}_\mathbf{s}\} =& E\{\mathcal{H}(\bar{\mathbf{s}}_d)^H\mathcal{H}(\bar{\mathbf{s}}_d)\} + \mathbf{R}_{\mathbf{s}_p}(\boldsymbol{\nu},\boldsymbol{\gamma})\\ =& (N-L)\sigma_d^2\mathbf{I} + \mathbf{R}_{\mathbf{s}_p}(\boldsymbol{\nu},\boldsymbol{\gamma}).\end{aligned} \quad (61)$$

Substituting the above into (9), the CRB becomes

$$\boldsymbol{\Lambda}(\boldsymbol{\nu},\boldsymbol{\gamma},\mathbf{s}_p) = \sigma_n^2[(N-L)\sigma_d^2\mathbf{I} + \mathbf{R}_{\mathbf{s}_p}(\boldsymbol{\nu},\boldsymbol{\gamma}) + \rho_h^2\sigma_n^2\mathbf{I}]^{-1}. \quad (62)$$

Notice that $(\boldsymbol{\nu} + \mathbf{e}_1 - \mathbf{e}_{n+1})$ is corresponding to shifting the n clusters to the right by 1 without changing their relative distances. From the structure of $\mathcal{H}(\bar{\mathbf{s}}_p)$, it is not hard to see that when $\nu_1 \geq L$, $\nu_{n+1} \geq L+1$

$$\mathbf{R}_{\mathbf{s}_p}(\boldsymbol{\nu},\boldsymbol{\gamma}) = \mathbf{R}_{\mathbf{s}_p}(\boldsymbol{\nu} + \mathbf{e}_1 - \mathbf{e}_{n+1},\boldsymbol{\gamma}). \quad (63)$$

This means, for fixed $\boldsymbol{\gamma}$ and $[\nu_2,\ldots,\nu_n]$, $\mathbf{R}_{\mathbf{s}_p}(\boldsymbol{\nu},\boldsymbol{\gamma})$ is invariant for different ν_1 and ν_{n+1}, and we have

$$\boldsymbol{\Lambda}(\boldsymbol{\nu},\boldsymbol{\gamma},\mathbf{s}_p) = \boldsymbol{\Lambda}(\boldsymbol{\nu} + \mathbf{e}_1 - \mathbf{e}_{n+1},\boldsymbol{\gamma},\mathbf{s}_p). \quad (64)$$

Therefore, $\boldsymbol{\Lambda}(\boldsymbol{\nu},\boldsymbol{\gamma},\mathbf{s}_p)$ is invariant under shifting of the n clusters among the midamble positions. □

APPENDIX C
PROOF OF THEOREM 2

From (9), we know that

$$\boldsymbol{\Lambda}^{-1}(\mathcal{P},\mathbf{s}_p)]_{ii} = \frac{1}{\sigma_n^2}[E\{\mathbf{Rs}\}]_{ii} + \rho_h^2, i = 1,\ldots,L+1 \quad (65)$$

where $[E\{\mathbf{Rs}\}]_{ii}$ is given by

$$\begin{aligned}[E\{\mathbf{Rs}\}]_{ii} =& [E\{\mathcal{H}^H(\bar{\mathbf{s}}_d)\mathcal{H}(\bar{\mathbf{s}}_d)\} + \mathbf{R}_{\mathbf{s}_p}]_{ii}\\ =& (N+P-L)\sigma_d^2 + \sum_{k=1}^{P}(\rho_k^2 - \sigma_d^2)I_i(k)\end{aligned}$$

where $\rho_k^2 = |s_p[k]|^2$ and $I_i(k)$ an indicator function, defined by $I_i(k) = 1$ if pilot symbol $s_p[k]$ appears in the ith column of $\mathcal{H}(\bar{\mathbf{s}}_p)$ and $I_i(k) = 0$ otherwise. Note that $\sum_{k=1}^P \rho_k^2 I_i(k) \leq P\sigma_p^2$. Denoting $q_i = \sum_{i=1}^P I_i(k)$ as the total number of pilot symbols in the ith column of $\mathcal{H}(\bar{\mathbf{s}}_p)$, we have

$$[E\{\mathbf{Rs}\}]_{ii} = (N+P-L)\sigma_d^2 + \sum_{k=1}^{P}\rho_k^2 I_i(k) - q_i\sigma_d^2. \quad (66)$$

Define

$$\begin{aligned}f(\mathcal{P},\mathbf{s}_p) \triangleq& \sum_{i=1}^{L+1}\frac{1}{[\boldsymbol{\Lambda}^{-1}(\mathcal{P},\mathbf{s}_p)]_{ii}}\\ =& \sum_{i=1}^{L+1}\frac{\sigma_n^2}{(N+P-L)\sigma_d^2 + \sum_{k=1}^{P}\rho_k^2 I_i(k) - q_i\sigma_d^2 + \rho_h^2\sigma_n^2}.\end{aligned} \quad (67)$$

Case 1—$P \geq 2L+1$: Let β_1, β_2 be the total number of pilot symbols in the two edge parts belonging to the beginning and end of a packet, respectively. Note that $\beta_1, \beta_2 \leq L$. We now bound q_i as follows:

$$\begin{aligned} q_i \geq & P - \beta_1 - \beta_2 + \max\{i - (L + 1 - \beta_1), 0\} \\ & + \max\{\beta_2 - i + 1, 0\} \\ \geq & P - L \\ & i = 1, \ldots, L + 1. \end{aligned} \tag{68}$$

The equalities hold for all i if and only if $\beta_1 = \beta_2 = L$. Therefore, the minimum number of pilot symbols in each column is $P - L$.

From (66), we have

$$\begin{aligned} [E\{\mathbf{Rs}\}]_{ii} \leq & (N + P - L)\sigma_d^2 + P\sigma_p^2 - (P - L)\sigma_d^2 \\ & i = 1, \ldots, L + 1 \end{aligned} \tag{69}$$

with equality if and only if $\beta_1 = \beta_2 = L$ and $\sum_{k=1}^{P} \rho_k^2 I_i(k) = P\sigma_p^2$, i.e., the total power $P\sigma_p^2$ is allocated on the $(P - 2L)$ pilot symbols that are in the midamble positions. From Lemma 1, we note that any midamble placements are shift invariant. Thus, the placement described in Theorem 2.1 maximizes $[E\{\mathbf{R_s}\}]_{ii}$ and minimizes $f(\mathcal{P}, \mathbf{s}_p)$ in (67)

$$\min_{\mathcal{P}, \mathbf{s}_p : \|\mathbf{s}_p\|^2 = P\sigma_p^2} f(\mathcal{P}, \mathbf{s}_p) = \frac{(L + 1)\sigma_n^2}{N\sigma_d^2 + P\sigma_p^2 + \rho_h^2\sigma_n^2}. \tag{70}$$

By the Cauchy–Schwartz inequality,[3] , $\lambda_i(\mathcal{P}, \mathbf{s}_p)$ is lower bounded by

$$\lambda_i(\mathcal{P}, \mathbf{s}_p) \geq \frac{1}{[\mathbf{\Lambda}^{-1}(\mathcal{P}, \mathbf{s}_p)]_{ii}} \tag{71}$$

and

$$\mathrm{tr}\{\mathbf{\Lambda}(\mathcal{P}, \mathbf{s}_p)\} \geq f(\mathcal{P}, \mathbf{s}_p) \tag{72}$$

where the equalities hold if and only if $\mathbf{R}_{\mathbf{s}_p} = P\sigma_p^2\mathbf{I}$. Thus, we have

$$\begin{aligned} \min_{\mathcal{P}, \mathbf{s}_p : \|\mathbf{s}_p\|^2 = P\sigma_p^2} \lambda_i(\mathcal{P}, \mathbf{s}_p) \geq & \min_{\mathcal{P}, \mathbf{s}_p : \|\mathbf{s}_p\|^2 = P\sigma_p^2} \frac{1}{[\mathbf{\Lambda}^{-1}(\mathcal{P}, \mathbf{s}_p)]_{ii}} \\ = & \frac{\sigma_n^2}{N\sigma_d^2 + P\sigma_p^2 + \rho_h^2\sigma_n^2} \end{aligned} \tag{73}$$

and

$$\begin{aligned} \min_{\mathcal{P}, \mathbf{s}_p : \|\mathbf{s}_p\|^2 = P\sigma_p^2} \mathrm{tr}\{\Lambda(\mathcal{P}, \mathbf{s}_p)\} \geq & \min_{\mathcal{P}, \mathbf{s}_p : \|\mathbf{s}_p\|^2 = P\sigma_p^2} f(\mathcal{P}, \mathbf{s}_p) \\ = & \frac{(L + 1)\sigma_n^2}{N\sigma_d^2 + P\sigma_p^2 + \rho_h^2\sigma_n^2} \end{aligned} \tag{74}$$

where the equalities hold under the optimal placement described in Theorem 2.1.

Case 2—$P \leq 2L$: i) $P\sigma_p^2 > \sigma_d^2$: We prove that the placement described in part 2) i) of Theorem 2 minimizes $f(\mathcal{P}, \mathbf{s}_p)$.

a) We first show that for any fixed placement, allocating total pilot energy on those symbols in the midamble positions, decreases $f(\mathcal{P}, \mathbf{s}_p)$.

From (67), we have

$$\begin{aligned} & f(\mathcal{P}, \mathbf{s}_p) \\ & \geq \sum_{i=1}^{L+1} \frac{\sigma_n^2}{(N + P - L)\sigma_d^2 + P\sigma_p^2 - q_i\sigma_d^2 + \rho_h^2\sigma_n^2} \\ & = \sum_{i=1}^{L+1} \frac{\sigma_n^2}{(N + P - L)\sigma_d^2 + (P\sigma_p^2 - \sigma_d^2) - (q_i - 1)\sigma_d^2 + \rho_h^2\sigma_n^2} \end{aligned} \tag{75}$$

with equality if and only if $q_i \geq 1$ and $\sum_{k=1}^{P} \rho_k^2 I_i(k) = P\sigma_p^2$, i.e., there exist pilot symbols in the midamble positions, and their total power is $P\sigma_p^2$, whereas the power of those at the edge parts are all zeros. Since $(P\sigma_p^2 - \sigma_d^2) > 0$, for any fixed q_i (fixed placement), (75) gives the minimum $f(\mathcal{P}, \mathbf{s}_p)$.

b) If (75) is satisfied, the only variable in $f(\mathcal{P}, \mathbf{s}_p)$ is q_i. Notice that placing pilot symbols at two ends decreases q_i for some i, thus decreasing $f(\mathcal{P}, \mathbf{s}_p)$ in (75). Therefore, to minimize (75), all $(P - 1)$ pilot symbols should be placed at the two ends. In other words, assign $P\sigma_p^2$ to a single pilot in the midamble position, and split the rest $(P - 1)$ pilot symbols into two clusters at two ends of a packet. Furthermore, among all possible ways of splitting these $(P - 1)$ symbols, dividing them evenly $(\lfloor (P - 1)/2 \rfloor, \lceil (P - 1)/2 \rceil)$ at two ends minimizes $f(\mathcal{P}, \mathbf{s}_p)$[4] . Thus the placement described in the Theorem minimizes $f(\mathcal{P}, \mathbf{s}_p)$.

We now calculate $\min f(\mathcal{P}, \mathbf{s}_p)$. Under the optimal placement, q_{i_*} for the ith column is given by (76), shown at the bottom of the page. Substituting q_{i_*} into (75), we obtain the minimum $f(\mathcal{P}, \mathbf{s}_p)$ as in (77), shown at the bottom of the next page. Combining the common terms, we obtain (18). Since, by this placement, $\mathbf{R}_{\mathbf{s}_p} = P\sigma_p^2\mathbf{I}$, following the same argument as in Case 1, we have

$$\min \mathrm{tr}\{\mathbf{\Lambda}(\mathcal{P}, \mathbf{s}_p)\} = \min f(\mathcal{P}, \mathbf{s}_p). \tag{78}$$

ii) $P\sigma_p^2 \leq \sigma_d^2$: In part i), under the optimal placement

$$\begin{aligned} & f(\mathcal{P}, \mathbf{s}_p) \\ & = \sum_{i=1}^{L+1} \frac{\sigma_n^2}{(N + P - L)\sigma_d^2 + (P\sigma_p^2 - \sigma_d^2) - (q_{i_*} - 1)\sigma_d^2 + \rho_h^2\sigma_n^2} \end{aligned} \tag{79}$$

where q_{i_*} is defined in (76). In the case when $P\sigma_p^2 \leq \sigma_d^2$, $f(\mathcal{P}, \mathbf{s}_p)$ can be further reduced by removing $(P\sigma_p^2 - \sigma_d^2)$ from

[3]The Cauchy-Schwartz inequality is that for any positive definite matrix $\mathbf{A}$, $(\mathbf{A}^{-1})_{ii} \geq 1/\mathbf{A}_{ii}$, with equality iff $\mathbf{A}$ is a diagonal matrix.

[4]This is because the following inequality: $(1/a) + (1/(a + b)) \geq (1/(a + b/2)) + (1/(a + b/2))$, where $a, b > 0$.

$$q_{i_*} = \begin{cases} \max\{\lceil \frac{P-1}{2} \rceil - i + 1, 0\} + 1, & 1 \leq i \leq L - \lfloor \frac{P-1}{2} \rfloor \\ P - L, & L - \lfloor \frac{P-1}{2} \rfloor + 1 \leq i \leq \lceil \frac{P-1}{2} \rceil + 1 \\ i - (L + 1 - \lfloor \frac{P-1}{2} \rfloor) + 1, & \max\{L - \lfloor \frac{P-1}{2} \rfloor + 1, \lceil \frac{P-1}{2} \rceil + 2\} \leq i \leq L + 1\ . \end{cases} \tag{76}$$

the denominator and reducing the quantity q_{i_*} for some i. It is not hard to see that $f(\mathcal{P}, \mathbf{s}_p)$ under the optimal placement in part i), which is shown in (79), monotonically decreases by moving the pilot symbol with power $P\sigma_p^2$ from the midamble part to the position next to the pilot, which is in the edge part and is closest to the center. Therefore, we have

$$\begin{aligned} &f(\mathcal{P}, \mathbf{s}_p) \\ &> \sum_{i=1}^{\lfloor P/2 \rfloor} \frac{\sigma_n^2}{(N+P-L)\sigma_d^2 + (P\sigma_p^2 - \sigma_d^2) - (q_{i_*}-1)\sigma_d^2 + \rho_h^2\sigma_n^2} \\ &\quad + \sum_{i=\lfloor P/2 \rfloor+1}^{L+1} \frac{\sigma_n^2}{(N+P-L)\sigma_d^2 - (q_{i_*}-1)\sigma_d^2 + \rho_h^2\sigma_n^2}. \end{aligned} \tag{80}$$

Thus, all pilot symbols should be placed at two ends while allocating the total pilot power on the two pilot symbols closest to the center. Finally, by the same argument in part i), allocating power evenly ($P\sigma_p^2/2$) to each pilot closest to the center minimizes $f(\mathcal{P}, \mathbf{s}_p)$. Thus, the placement described in part 2) ii) of Theorem 2.2 is optimal. Under this placement, we calculate $\sum_{k=1}^{P}(\rho_k^2 - \sigma_d^2)I_i(k)$ as in (81), shown at the bottom of the page, and the minimum $f(\mathcal{P}, \mathbf{s}_p)$ is as in (82), also shown at the bottom of the page. Rearranging the index, we obtain (21). Since $\mathbf{R}_{\mathbf{s}_p}$ is diagonal under this placement, we have

$$\min \operatorname{tr}\{\mathbf{\Lambda}(\mathcal{P}, \mathbf{s}_p)\} = \min f(\mathcal{P}, \mathbf{s}_p). \qquad \square$$

APPENDIX D
PROOF OF THEOREM 3

Case 1—$\sigma_p^2 > \sigma_d^2$: In this case, since $\sigma_p^2 > \sigma_d^2$, (65) becomes

$$\begin{aligned} &[\mathbf{\Lambda}^{-1}(\mathcal{P}, \mathbf{s}_p)]_{ii} \\ &\quad = \frac{1}{\sigma_n^2}((N+P-L)\sigma_d^2 + q_i(\sigma_p^2 - \sigma_d^2)) + \rho_h^2 \qquad (83) \\ &\quad \le \frac{1}{\sigma_n^2}((N+P-L)\sigma_d^2 + P(\sigma_p^2 - \sigma_d^2)) + \rho_h^2 \qquad (84) \end{aligned}$$

with equality if and only if $\sum_{k=1}^{P} I_i(k) = P$ for all i. In other words, all the pilot symbols should be placed in the midamble positions

$$\mathcal{P}_* \in \{(\boldsymbol{\nu}, \boldsymbol{\gamma}) : \nu_1, \nu_{n+1} \ge L\}. \tag{85}$$

Following (71), we have

$$\begin{aligned} \min_{\mathcal{P}, \mathbf{s}_p : |s_p[i]|^2 = \sigma_p^2} \lambda_i(\mathcal{P}, \mathbf{s}_p) &\ge \min_{\mathcal{P}, \mathbf{s}_p : \|\mathbf{s}_p\|^2 = P\sigma_p^2} \frac{1}{[\mathbf{\Lambda}^{-1}(\mathcal{P}, \mathbf{s}_p)]_{ii}} \\ &= \frac{\sigma_n^2}{(N-L)\sigma_d^2 + P\sigma_p^2 + \rho_h^2\sigma_n^2} \end{aligned} \tag{86}$$

with equality when the conditions in (22) and (23) are satisfied.

Case 2—$\sigma_p^2 = \sigma_d^2$: When pilot and data powers are equal, we see that (83) becomes

$$[\mathbf{\Lambda}^{-1}(\mathcal{P}, \mathbf{s}_p)]_{ii} = \frac{\sigma_n^2}{(N+P-L)\sigma_d^2 + \rho_h^2\sigma_n^2}. \tag{87}$$

$$\begin{aligned} \min f(\mathcal{P}, \mathbf{s}_p) = &\frac{\max\{P-L, 0\}\sigma_n^2}{N\sigma_d^2 + P\sigma_p^2 + \rho_h^2\sigma_n^2} \\ &+ \sum_{i=1}^{L-\lfloor (P-1)/2 \rfloor} \frac{\sigma_n^2}{(N+P-L)\sigma_d^2 + P\sigma_p^2 - (\max\{\lceil \frac{P-1}{2} \rceil - i + 1, 0\} + 1)\sigma_d^2 + \rho_h^2\sigma_n^2} \\ &+ \sum_{i=\max\{L-\lfloor (P-1)/2 \rfloor+1, \lceil (P-1)/2 \rceil + 2\}}^{L+1} \frac{\sigma_n^2}{(N+P-L)\sigma_d^2 + P\sigma_p^2 - (i - [L - \lfloor \frac{P-1}{2} \rfloor])\sigma_d^2 + \rho_h^2\sigma_n^2}. \end{aligned} \tag{77}$$

$$\sum_{k=1}^{P}(\rho_k^2 - \sigma_d^2)I_i(k) = \begin{cases} \frac{P\sigma_p^2}{2} - (\lceil \frac{P}{2} \rceil - i + 1)\sigma_d^2, & 1 \le i \le \min\{L+1-\lfloor \frac{P}{2} \rfloor, \lceil \frac{P}{2} \rceil\} \\ P\sigma_p^2 - (P-L)\sigma_d^2, & \text{if } L+1 \le P : L+2-\lfloor \frac{P}{2} \rfloor \le i \le \lceil \frac{P}{2} \rceil \\ 0, & \text{if } L \ge P : \lceil \frac{P}{2} \rceil + 1 \le i \le L+1-\lfloor \frac{P}{2} \rfloor \\ \frac{P\sigma_p^2}{2} - [i - (L+1-\lfloor \frac{P}{2} \rfloor)]\sigma_d^2, & \max\{L+2-\lfloor \frac{P}{2} \rfloor, \lceil \frac{P}{2} \rceil + 1\} \le i \le L+1 \end{cases} \tag{81}$$

$$\begin{aligned} \min f(\mathcal{P}, \mathbf{s}_p) = &\frac{\max\{P-L-1, 0\}\sigma_n^2}{N\sigma_d^2 + P\sigma_p^2 + \rho_h^2\sigma_n^2} + \frac{\max\{L+1-P, 0\}\sigma_n^2}{(N+P-L)\sigma_d^2 + \rho_h^2\sigma_n^2} \\ &+ \sum_{i=1}^{\min\{L+1-\lfloor P/2 \rfloor, \lceil P/2 \rceil\}} \frac{\sigma_n^2}{(N+P-L)\sigma_d^2 + \frac{P\sigma_p^2}{2} - (\lceil \frac{P}{2} \rceil - i + 1)\sigma_d^2 + \rho_h^2\sigma_n^2} \\ &+ \sum_{i=\max\{L+2-\lfloor P/2 \rfloor, \lceil P/2 \rceil + 1\}}^{L+1} \frac{\sigma_n^2}{(N+P-L)\sigma_d^2 + \frac{P\sigma_p^2}{2} - (i - (L+1-\lfloor \frac{P}{2} \rfloor))\sigma_d^2 + \rho_h^2\sigma_n^2}. \end{aligned} \tag{82}$$

Thus, different placements do not affect $[\mathbf{\Lambda}^{-1}(\mathcal{P},\mathbf{s}_p)]_{ii}$, and we have

$$\min_{\mathcal{P},\mathbf{s}_p:|s_p[i]|^2=\sigma_p^2} \lambda_i(\mathcal{P},\mathbf{s}_p) = \frac{\sigma_n^2}{(N-L)\sigma_d^2+P\sigma_p^2+\rho_h^2\sigma_n^2}$$

with equality if and only if

$$\mathbf{R}_{\mathbf{s}_p} = P\sigma_p^2\mathbf{I}.$$

Case 3—$\sigma_p^2 < \sigma_d^2$ and $P \geq 2L$: From (68), q_i is lower bounded by $q_i \geq P-L$. Then, in this case, $[\mathbf{\Lambda}^{-1}(\mathcal{P},\mathbf{s}_p)]_{ii}$ in (83) satisfies

$$[\mathbf{\Lambda}^{-1}(\mathcal{P},\mathbf{s}_p)]_{ii} \leq \frac{1}{\sigma_n^2}(N\sigma_d^2+(P-L)\sigma_p^2+\rho_h^2\sigma_n^2). \quad (88)$$

Note that the equality holds when $q_i = P-L$, i.e., all the edge positions are filled with pilot symbols. Thus

$$\min_{\mathcal{P},\mathbf{s}_p:|s_p[i]|^2=\sigma_p^2} \lambda_i(\mathcal{P},\mathbf{s}_p) = \frac{\sigma_n^2}{N\sigma_d^2+(P-L)\sigma_p^2+\rho_h^2\sigma_n^2}$$

with equality if and only if

$$\mathbf{R}_{\mathbf{s}_p} = P\sigma_p^2\mathbf{I}.$$

□

APPENDIX E
PROOF OF THEOREM 4

Given in (32), the CRB under the MIMO model is

$$\begin{aligned}\mathbf{\Lambda}(\mathcal{P},\mathbf{s}_p) &\\ &= \sigma_n^2[(E\{\mathcal{R}\mathbf{s}\}+\sigma_n^2\text{diag}(\rho_{h_1}^2\mathbf{I}_{L_1},\ldots,\rho_{h_K}^2\mathbf{I}_{L_K})\otimes\mathbf{I}_M]^{-1}.\end{aligned}$$

Again, as $\mathbf{s}$ can be decomposed into $\bar{\mathbf{s}}_p$ and $\bar{\mathbf{s}}_d$, we have

$$\begin{aligned}E\{\mathcal{R}_\mathbf{s}\} =& E\{\mathcal{F}^H(\bar{\mathbf{s}}_d)\mathcal{F}(\bar{\mathbf{s}}_d)\}+\mathcal{F}^H(\bar{\mathbf{s}}_p)\mathcal{F}(\bar{\mathbf{s}}_p)\\ =& \begin{pmatrix}\mathbf{R}_{\mathbf{s}_d}^{(1)}+\mathbf{R}_{\mathbf{s}_p}^{(1)} & & * \\ & \ddots & \\ * & & \mathbf{R}_{\mathbf{s}_d}^{(K)}+\mathbf{R}_{\mathbf{s}_p}^{(K)}\end{pmatrix}\otimes\mathbf{I}_M\end{aligned} \quad (89)$$

where $\mathbf{R}_{\mathbf{s}_d}^{(k)}$ and $\mathbf{R}_{\mathbf{s}_p}^{(k)}$ are the autocorrelation matrices for the kth user defined under the SISO model.

From the above equation, we can see that the expression of the kth diagonal block (corresponding to the kth user) is the same as that in the SISO case. Thus, we have

$$[\mathbf{\Lambda}^{-1}(\mathcal{P},\mathbf{s}_p)]_{ii}^{(k)} = \frac{1}{\sigma_n^2}[E\{\mathbf{R}^{(k)}\mathbf{s}\}]_{ii}+\rho_h^2. \quad (90)$$

Define (91), shown at the bottom of the page.

Case 1—$P_k \geq 2L_k+1, \forall k$: Notice that (90) only involves $E\{\mathbf{R}_\mathbf{s}^{(k)}\}$ from the kth user; thus, to maximize $[\mathbf{\Lambda}^{-1}(\mathcal{P},\mathbf{s}_p)]_{ii}^{(k)}$, the placement for the packet from user k should be the same as described in the SISO case. Therefore, we have

$$\begin{aligned}\min_{\substack{\mathcal{P},\mathbf{s}_p:\|\mathbf{s}_p^{(k)}\|^2=P_k\sigma_{p_k}^2\\1\leq k\leq K}} &\frac{1}{[\mathbf{\Lambda}^{-1}(\mathcal{P},\mathbf{s}_p)]_{ii}^{(k)}}\\ &= \frac{\sigma_n^2}{N_k\sigma_{d_k}^2+P_k\sigma_{p_k}^2+\rho_{h_k}^2\sigma_n^2}.\end{aligned} \quad (92)$$

Using the Cauchy–Schwartz inequality again, we have

$$\begin{aligned}\min_{\substack{\mathcal{P},\mathbf{s}_p:\|\mathbf{s}_p^{(k)}\|^2=P_k\sigma_{p_k}^2\\1\leq k\leq K}} &\lambda_i^{(k)}(\mathcal{P},\mathbf{s}_p)\\ &\geq \min_{\substack{\mathcal{P},\mathbf{s}_p:\|\mathbf{s}_p^{(k)}\|^2=P_k\sigma_{p_k}^2\\1\leq k\leq K}} \frac{1}{[\mathbf{\Lambda}^{-1}(\mathcal{P},\mathbf{s}_p)]_{ii}^{(k)}}\end{aligned} \quad (93)$$

with equality if and only if

$$\mathcal{R}_{\mathbf{s}_p} = \text{diag}(P_1\sigma_{p_1}^2,\ldots,P_K\sigma_{p_K}^2)\otimes\mathbf{I}_M.$$

Notice the that optimal design and placement involves pilot symbols among all users. This cross-user effect on placement can be seen in $\mathcal{R}_{\mathbf{s}_p}$, where the off-diagonal ijth block is the cross "correlation" matrix between user i and j.

Case 2—$P_k \leq 2L_k, \forall k$: We do not give a detailed proof in this case since it can be similarly derived from previous results. Similarly, as in Case 1, we see that by Theorem 2, the placement described in this case minimizes $f(\mathcal{P},\mathbf{s}_p)$ in (91). Therefore, to satisfy

$$\min_{\substack{\mathcal{P},\mathbf{s}_p:\|\mathbf{s}_p^{(k)}\|^2=P_k\sigma_{p_k}^2\\1\leq k\leq K}} \text{tr}[\mathbf{\Lambda}(\mathcal{P},\mathbf{s}_p)] = \min_{\substack{\mathcal{P},\mathbf{s}_p:\|\mathbf{s}_p^{(k)}\|^2=P_k\sigma_{pk}^2\\1\leq k\leq K}} f(\mathcal{P},\mathbf{s}_p)$$

we require

$$\mathcal{R}_{\mathbf{s}_p} = \text{diag}(P_1\sigma_{p_1}^2,\ldots,P_K\sigma_{p_K}^2)\otimes\mathbf{I}_M.$$

□

APPENDIX F
PROOF OF THEOREM 6

From the CRB $\mathbf{\Lambda}(\mathcal{P},\mathbf{s}_p)$ for channel estimators given in (46), we have

$$\begin{aligned}\text{tr}\mathbf{\Lambda}(\mathcal{P},\mathbf{s}_p) =& \text{tr}\left[\mathbf{G}\left(\frac{1}{\sigma_n^2}\mathbf{G}^H E\{\mathbf{R}\mathbf{s}\}\mathbf{G}+\rho_v^2\mathbf{I}\right)^{-1}\mathbf{G}^H\right]\\ =& \text{tr}\left[\left(\frac{1}{\sigma_n^2}\mathbf{G}^H E\{\mathbf{R}\mathbf{s}\}\mathbf{G}+\rho_v^2\mathbf{I}\right)^{-1}\right]\\ =& \sum_{i=1}^{r}\frac{\sigma_n^2}{\lambda_i(\mathbf{G}^H E\{\mathbf{R}_\mathbf{s}\}\mathbf{G})+\rho_v^2\sigma_n^2}.\end{aligned} \quad (94)$$

where $E\{\mathbf{R}_\mathbf{s}\}_{ii}$ is upper bounded by

$$E\{\mathbf{R}_\mathbf{s}\}_{ii} \leq N\sigma_d^2+P\sigma_p^2 \quad (95)$$

$$f(\mathcal{P},\mathbf{s}_p) \triangleq \sum_{k=1}^{K}\sum_{i=1}^{L_k+1}\frac{\sigma_n^2}{(N_k+P_k-L_k)\sigma_{d_k}^2+\sum_{j=1}^{P_k}(\rho_j^2-\sigma_{d_k}^2)I_i^{(k)}(j)+\rho_{h_k}^2\sigma_n^2}. \quad (91)$$

with the equality iff the placement described in (48) and (49) is satisfied, where the total pilot power is concentrated on those pilots in the midamble positions.

Note that by definition, for any orthogonal pilot sequence $\mathbf{s}_p$, $\mathbf{R}_{\mathbf{s}_p}$ is diagonal. Consequently, $E\{\mathbf{R}_\mathbf{s}\}$ is a diagonal matrix. Therefore, we have

$$\begin{aligned}\sum_{i=1}^{r}\lambda_i(\mathbf{G}^H E\{\mathbf{R}_\mathbf{s}\}\mathbf{G}) =& \mathrm{tr}[\mathbf{G}^H E\{\mathbf{R}_\mathbf{s}\}\mathbf{G}] = \sum_{i=1}^{r}\mathbf{q}_i^H E\{\mathbf{R}_\mathbf{s}\}\mathbf{q}_i \\ \le& r\lambda_{\max}(E\{\mathbf{R}_\mathbf{s}\}) = r\max_i E\{\mathbf{R}_\mathbf{s}\}_{ii} \\ \le& r(N\sigma_d^2 + P\sigma_p^2) \quad (96)\end{aligned}$$

where equalities hold when $E\{\mathbf{R}_\mathbf{s}\} = (N\sigma_d^2 + P\sigma_p^2)\mathbf{I}$.

Thus, minimizing $\mathrm{tr}\boldsymbol{\Lambda}(\mathcal{P}, \mathbf{s}_p)$ is equivalent to

$$\min_{\lambda_i}\sum_{i=1}^{r}\frac{\sigma_n^2}{\lambda_i(\mathbf{G}^H E\{\mathbf{R}_\mathbf{s}\}\mathbf{G}) + \rho_v^2\sigma_n^2} \quad \text{subject to} \quad \sum_{i=1}^{r}\lambda_i(\mathbf{G}^H E\{\mathbf{R}_\mathbf{s}\}\mathbf{G}) \le r(N\sigma_d^2 + P\sigma_p^2). \quad (97)$$

Hence, $\mathrm{tr}\boldsymbol{\Lambda}(\mathcal{P}, \mathbf{s}_p)$ is minimized by making all $\lambda_i(\mathbf{G}^H E\{\mathbf{R}_\mathbf{s}\}\mathbf{G})$ equal

$$\lambda_i(\mathbf{G}^H E\{\mathbf{R}_\mathbf{s}\}\mathbf{G}) = N\sigma_d^2 + P\sigma_p^2, i = 1, \ldots, r$$

and, at the same time, satisfying

$$E\{\mathbf{R}_\mathbf{s}\} = (N\sigma_d^2 + P\sigma_p^2)\mathbf{I}.$$

Therefore, the placement described in the Theorem gives the minimum CRB shown in (50). □

REFERENCES

[1] S. Adireddy and L. Tong, “Detection with embedded known symbols: Optimal symbol placement and equalization,” in *Proc. ICASSP Conf.*, vol. 5, Istanbul, Turkey, June 2000, pp. 2541–2544.

[2] S. Adireddy, L. Tong, and H. Viswanathan, “Optimsl placement of training for frequency selective block-fading channels,” *IEEE Trans. Inform Theory*, vol. 48, pp. 2338–2353, Aug. 2002.

[3] B. Hassibi and B. Hochwald, “How much training is needed in multiple-antenna wireless links,” *IEEE Trans. Inform. Theory*, Aug. 2000, submitted for publication.

[4] S. Ohno and G. B. Giannakis, “Optimal training and redundant precoding for block transmissions with application to wireless OFDM,” *IEEE Trans. Commun.*, Nov. 2000, submitted for publication.

[5] R. Heimiller, “Phase shift pulse codes with good periodic correlation properties,” *IRE Trans. Inform. Theory*, vol. IT-7, pp. 254–257, Oct. 1961.

[6] B. Popovic, “Generalized chirp-like polyphase sequences with optimum correlation properties,” *IEEE Trans. Inform. Theory*, vol. 38, pp. 1406–1409, July 1992.

[7] J. C. L. Ng, K. B. Letaief, and R. D. Murch, “Complex optimal sequences with constant magnitude for fast channel estimation initialization,” *IEEE Trans. Commun.*, vol. 46, pp. 305–308, Mar. 1998.

[8] C. Tellambura, M. G. Parker, Y. J. Guo, S. J. Shepherd, and S. K. Barton, “Optimal sequences for channel estimation using discrete Fourier transform techniques,” *IEEE Trans. Commun.*, vol. 47, pp. 230–238, Feb. 1999.

[9] R. Negi and J. Cioffi, “Pilot tone selection for channel estimation in a mobile OFDM system,” *IEEE Tran. Consum. Electron.*, vol. 44, pp. 1122–1128, Aug. 1998.

[10] F. Ling, “Optimal reception, performance bound, and cutoff rate analysis of references-assisted coherent CDMA communications with applications,” *IEEE Trans. Commun.*, vol. 47, pp. 1583–1592, Oct. 1999.

[11] S. Adireddy and L. Tong, “Optimal placement of known symbols for nonergodic broadcast channels,” *IEEE Trans. Inform. Theory*, submitted for publication.

[12] B. M. Sadler, R. J. Kozick, and T. Moore, “Constrained CRB’s for channel and signal estimation of mimo systems,” in *Proc. 35st Annu. Conf. Inform. Sci. Syst.*, Baltimore, MD, Mar. 2001.

[13] J. D. Gorman and A. O. Hero, “Lower bounds for parametric estimation with constraints,” *IEEE Trans. Inform. Theory*, vol. 36, pp. 1285–1301, Nov. 1990.

[14] E. de Carvahlo and D. T. M. Slock, “Semi-blind methods for FIR multichannel estimation,” in *Signal Processing Advances in Wireless and Mobile Communications*, Giannakis, Hua, Stoica, and Tong, Eds. Englewood Cliffs, NJ: Prentice-Hall, 2000, pp. 211–254.

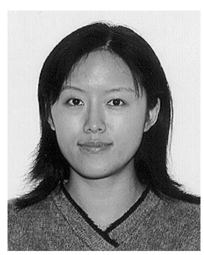
[15] C. Budianu and L. Tong, “Channel estimation for space-time orthogonal block code,” *IEEE Trans. Signal Processing*, vol. 50, pp. 2515–2528, Oct. 2002.

[16] H. L. Van Trees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1968, vol. 1.

[17] E. Weinstein and A. Weiss, “A general class of lower bounds in parameter estimation,” *IEEE Trans. Inform. Theory*, vol. 34, pp. 338–342, Mar. 1988.

[18] E. de Carvalho and D. T. M. Slock, “Cramér-Rao bounds for semi blind, blind and training sequence based channel estimation,” *Proc. First IEEE Signal Process. Workshop Signal Process. Adv. Wireless Commun.*, vol. 1, pp. 129–132, Apr. 1997.

[19] M. Dong and L. Tong, “Channel estimation and equalization with block interleavers,” *Proc. Third IEEE Signal Process. Workshop Signal Process. Adv. Wireless Commun.*, Mar. 2001.



Min Dong (S’00) received the B.S. degree from the Department of Automation, Tsinghua University, Beijing, China, in 1998. She is now pursuing the Ph.D. degree at the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY. Her research interests include statistical signal processing, wireless communications, and coomuunication networks.

Lang Tong (S’87–M’91–SM’01) received the B.E. degree from Tsinghua University, Beijing, China, in 1985 and the M.S. and Ph.D. degrees in electrical engineering in 1987 and 1990, respectively, from the University of Notre Dame, Notre Dame, IN.

After being a Postdoctoral Research Associate with the Information Systems Laboratory, Stanford University, Stanford, CA, he joined the Department of Electrical and Computer Engineering, West Virginia University, Morgantown, and was also with the University of Connecticut, Storrs. Since the Fall of 1998, he has been with the School of Electrical Engineering, Cornell University, Ithaca, NY, where he is an Associate Professor. He also held a Visiting Assistant Professor position at Stanford University in the Summer of 1992. His research interests include statistical signal processing, wireless communications, and system theory.

Dr. Tong received the Young Investigator Award from the Office of Naval Research in 1996 and the Outstanding Young Author Award from the IEEE Circuits and Systems Society.