INVITED PAPER *Special Section on Blind Signal Processing: ICA and BSS*

Optimal Pilot Placement for Semi-Blind Channel Tracking of Packetized Transmission over Time-Varying Channels*

Min DONG†, **Srihari ADIREDDY**†, ***and*** **Lang TONG**†a), ***Nonmembers***

SUMMARY The problem of optimal placement of pilot symbols is considered for single carrier packet-based transmission over time varying channels. Both flat and frequency-selective fading channels are considered, and the time variation of the channel is modeled by Gauss-Markov process. The semi-blind linear minimum mean-square error (LMMSE) channel estimation is used. Two different performance criteria, namely the maximum mean square error (MSE) of the channel tap state over a packet and the cumulative channel MSE over a packet, are used to compare different placement schemes. The pilot symbols are assumed to be placed in clusters of length $(2L + 1)$ where L is the channel order, and only one non-zero training symbols is placed at the center of each cluster. It is shown that, at high SNR, either performance metric is minimized by distributing the pilot clusters throughout the packet periodically. It is shown that at low SNR, the placement is in fact not optimal. Finally, the performance under the periodic placement is compared with that obtained with superimposed pilots.

key words: *pilot symbols, placement schemes, channel estimation, time varying, Gauss-Markov process*

1. Introduction

Due to the time-varying nature of the propagation channel, channel state acquisition is one of the main challenges to achieving high data rates in wireless communication. Pilot symbols are typically inserted in data packets to facilitate channel estimation and tracking. It has been recently shown that the optimized placement of pilot symbols enhances overall system performance [1], [6]. In a time-varying environment, the optimization of the pilot symbols placement in data packets is even more crucial.

The problem of optimal placement has been previously addressed under various settings. Optimal placement of training for maximizing ergodic capacity in the setting of frequency selective block fading has been considered in [1]. Under the assumption that the channel taps are i.i.d complex Gaussian, it was shown that periodic placement in frequency is optimal for OFDM where as a class of quasi periodic placement schemes were optimal for single carrier systems. Periodic placement in frequency turns out to be optimal for an OFDM system that maximizes ergodic capacity at high SNR and large coherence time regime in [12], too. From the channel estimation point of view, the optimal placement minimizing the Cramér-Rao Lower Bound (CRLB) for semi-blind estimation of frequency selective block fading channels in both single input single output (SISO) and multi input multi output (MIMO) systems was found in [6]. Periodic placement is shown to be one of the optimal placements in this frame work, too. Placement issues for channel estimation in multiple-antenna systems employing orthogonal space-time codes has been considered in [4].

Optimal training for time-varying channels has been previously explored into in [5], [7], [10], [11]. In [5], Cavers analyzed the pilot symbol assisted modulation (PSAM) under flat Rayleigh fading in terms of bit error rate, assuming a periodic training of cluster size 1. Also discussed is the effect of pilot symbol spacing and Doppler spread. In [10], for the flat Rayleigh fading modeled by a Gauss-Markov process, under the PSAM scheme mentioned above, the optimal spacing between the pilot symbols is determined numerically by maximizing the mutual information with binary inputs. In [11] the channel is once again assumed to be flat, Rayleigh but the time variation is modeled by a band-limited process. At high SNR and large block length regime, optimal parameters for pilots including the number and spacing of pilot symbols, power allocated to information and pilots are then determined by maximizing a lower bound on capacity. All the prior works start with the assumption that the pilot symbols are inserted one by one periodically in the data stream. Recently, we considered the optimal placement of pilots in an infinite data stream over time-varying channels with Gauss-Markov variation. Under the assumption that the Kalman Filter is used for channel tracking, it is shown that periodically placing pilot symbols one by one is optimal. See [8] for the details.

In this paper, we consider the problem of optimal pilot placement in finite length data packets that are being transmitted over time-varying channels. It is assumed that semi-blind linear minimum mean-square er-

Manuscript received July 26, 2002.
Manuscript revised October 31, 2002.
Final manuscript received November 28, 2002.

†The authors are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853, USA.
a) E-mail: ltong@ece.cornell.edu

*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 under Grant DAAD19-01-2-0011.

ror (LMMSE) channel estimator is used at the receiver. For simplicity, the estimator is first derived for flat fading channels, and then extended to the frequency selective channel with order L. For frequency selective fading channels, we assume that pilot symbols are constrained to be placed in clusters of length $(2L+1)$ with each cluster having only one non-zero symbol placed at the center. The presence of pilot symbols in the data stream makes the MSE of the channel estimator time varying. Two different performance criteria are considered: the maximum MSE of the channel tap state over a packet and cumulative channel MSE over a packet. The first criterion is particularly relevant for receivers using symbol-by-symbol techniques. We show that, at high SNR, both performance metrics are minimized by distributing the pilot clusters periodically in the packet. We also point out that this placement is in fact not optimal at low SNR. Finally, the performance under the periodic placement is compared with that obtained with superimposed pilots.

This paper is organized as follows. In Sect. 2 we introduce the system model, and formulate the problem. For both flat and frequency selective channels, we then obtain the expression of the semi-blind LMMSE channel estimator and its estimation error in Sect. 3. The optimization of pilot placement is then considered in Sect. 4, where the optimal placement schemes are derived for the high SNR regime. Channel estimation with superimposed pilots scheme is dealt with in Sect. 5. Section 6 contains some relevant simulations and the concluding remarks are delineated in Sect. 7.

2. System Model

In this section, we describe the system model. We first give the model assumed for the channel and then describe the model for the data packets.

2.1 Channel Model

The channel is modeled as a linear time-varying FIR filter of order L, and we denote the channel state vector at time instant k as $\mathbf{h}_k = [h_k[0], \cdots, h_k[L]]^T$. The time variation of the frequency selective fading channel within the duration of the data packet is modeled by the following first order vector Gauss-Markov process:

$$\mathbf{h}_k = a\mathbf{h}_{k-1} + \mathbf{u}_k, \tag{1}$$

where $\mathbf{u}_k \overset{i.i.d.}{\sim} \mathcal{CN}(0, (1-a^2)\sigma_h^2\mathbf{I})$ is the driving noise with the $\mathbf{u}_k$'s independent, and a is the correlation coefficient that may vary between zero and one according to the fading channel bandwidth f_m (Doppler spread). We assume that $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \sigma_h^2\mathbf{I})$. These assumptions imply that the channel taps are independent and identically distributed and each tap fades in a statistically identical fashion. The output at time instant k, y_k, can then be written as

$$y_k = \sum_{i=0}^{L} h_k[i]s_{k-i} + n_k, \;\; k = L+1, \cdots, N+P,$$

where s_k the transmitted symbols and $n_k \overset{i.i.d.}{\sim} \mathcal{CN}(0, \sigma_n^2)$ the complex circular white Gaussian noise at time k. Note that, as a special case when $L = 0$, we have the Rayleigh flat fading model.

We assume that each packet consists of of N data symbols and P pilot symbols. Data symbols are modeled as i.i.d random variables with zero mean and variance σ_d^2. We assume that each pilot symbol has the same power σ_p^2. We further assume that the data, the channel and noise are independent.

Finally, we assume that the receiver forms a semi-blind LMMSE estimate of the channel. It is also assumed that the estimate is formed independently from packet to packet. This is reasonable in systems where consecutive transmissions to a user are sufficiently separated in either time or frequency.

2.2 Pilot Symbol Placement

In general, the placement of pilot symbols in a packet can be described by a tuple $\mathbf{r} = (\boldsymbol{\nu}, \boldsymbol{\gamma})$, where $\boldsymbol{\nu} = [\nu_1, \cdots, \nu_{n+1}]$ is the data block lengths vector and $\boldsymbol{\gamma} = [\gamma_1, \cdots, \gamma_n]$ the pilot cluster lengths vector and n is the number of pilot clusters. An example is illustrated in Fig. 1. The vectors satisfy the following constraints

$$\sum_{i=1}^{n+1} \nu_i = N, \quad \sum_{i=1}^{n} \gamma_i = P. \tag{2}$$

Moreover, for those placements that start with pilot symbols, $\nu_1 = 0$, and for those that end with pilot symbols, $\nu_{n+1} = 0$. An equivalent way of specifying the placement is through the set $\mathcal{P}$ that contains the indexes of the positions of the pilot symbols in the packet. For example, for $N = 6$, $P = 2$ and $n = 2$, a placement described by $\mathbf{r} = ([2, 2, 2], [1, 1])$ can be equivalently specified by the index set $\mathcal{P} = \{3, 6\}$. We will use one of these two notations, depending on which one is most convenient, to specify the placement of pilot symbols in the packet. For convenience, we refer to the symbols between any two consecutive known symbol clusters as unknown symbol blocks.

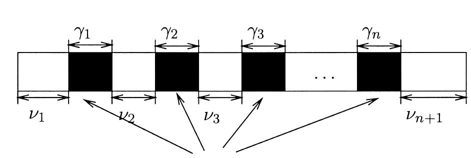

Fig. 1 An input sequence with multiple clusters.

2.3 Optimization Criteria

Both the metrics considered in this paper depend on the estimate of only those taps that are associated with data symbols. Given a placement $\mathcal{P}$, these channel taps are given by $h_k[l], (k-l) \notin \mathcal{P}$. The rationale behind such consideration is that the estimates of only these taps alone affect the performance of symbol decoder.

The first criterion considered is maximum MSE. The objective is to find the placement $\mathcal{P}^*$ that minimizes the maximum MSE of channel taps. If there exist multiple placements that have the same maximum MSE, we would like to obtain the placement scheme that minimizes the number of channel tap estimates with the maximum MSE. In other words, let $\tilde{h}_k[l]$ be the estimation error of $h_k[l]$, and let

$$\mathcal{E}(\mathcal{P}) = \max_{k,l:(k-l)\notin\mathcal{P}} \mathrm{E}\left\{\left|\tilde{h}_k[l]\right|^2\right\}, \tag{3}$$

$$\mathbb{P} = \{\mathcal{P}_\# : \mathcal{E}(\mathcal{P}_\#) = \min_{\mathcal{P}} \mathcal{E}(\mathcal{P})\}, \tag{4}$$

where the set $\mathbb{P}$ contains all those placements schemes that minimize the maximum MSE. The optimal placement scheme is the one that belongs to the set $\mathbb{P}$ and has the least instances of channel tap estimates with the maximum MSE. Let $\mathbf{I}_{\mathcal{P}_\#}$ be the index set of $\mathcal{P}_\# \in \mathbb{P}$, such that

$$\mathbf{I}_{\mathcal{P}_\#} = \left\{(k,l) : (k-l) \notin \mathcal{P}_\#, \mathrm{E}\left\{\left|\tilde{h}_k[l]\right|^2\right\} = \mathcal{E}(\mathcal{P}_\#)\right\}. \tag{5}$$

The cardinality of the set $\mathbf{I}_{\mathcal{P}_\#}$ gives the number of instances where the MSE of the channel tap is equal to the maximum MSE. Then the optimal placement is given by

$$\mathcal{P}^* = \arg\min_{\mathcal{P}_\# \in \mathbb{P}} \left|\mathbf{I}_{\mathcal{P}_\#}\right|, \tag{6}$$

where $|\cdot|$ in (6) denotes the cardinality of the set.

The other optimization criterion considered in this paper is the cumulative channel MSE over all those channel taps that affect the output due to data symbols. The optimal placement is then given by the one that minimizes this metric. Formally it is

$$\mathcal{P}^* = \arg\min_{\mathcal{P}} \sum_{k,l:(k-l)\notin\mathcal{P}} \mathrm{E}\left\{\left|\tilde{h}_k[l]\right|^2\right\}. \tag{7}$$

3. Semi-Blind LMMSE Channel Estimator

In this section we derive the structure and properties of the semi-blind LMMSE estimator for flat fading and frequency selective fading channels. If the vector $\mathbf{s} = [s_{N+P}, \cdots, s_1]^t$ is the transmitted data packet and $\mathbf{y}$ is the corresponding output, we have

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{8}$$

where

$$\mathbf{H} = \begin{bmatrix} h_{N+P}[0] & \cdots & h_{N+P}[L] & & & \\ & \ddots & & & \ddots & \\ & & & h_{L+1}[0] & \cdots & h_{L+1}[L] \end{bmatrix} \tag{9}$$

Denote the output that is due to the pilot symbols alone as $\mathbf{y}_t$, and the rest of the output as $\mathbf{y}_d$. Because of the unknown data sequence, the received data $\mathbf{y}$ and the channel $\mathbf{h}_k$ are not jointly Gaussian and hence the MMSE estimator of $\mathbf{h}_k$ does not have a closed-form expression. Therefore, the receiver forms the semi-blind LMMSE estimate $\hat{\mathbf{h}}_k$ of $\mathbf{h}_k$. The estimation error, denoted as $\tilde{\mathbf{h}}_k$ is defined as $\hat{\mathbf{h}}_k - \mathbf{h}_k$.

3.1 Flat Fading Channels

A important special case, corresponding to $L = 0$, is when the channel is assumed to under go Rayleigh flat fading. The fading process in (1) becomes a scalar Gauss-Markov model. The system equations for this scenario are given by

$$\mathbf{y}_t = \mathbf{S}_t\mathbf{h}_t + \mathbf{n}_t, \quad \mathbf{y}_d = \mathbf{S}_d\mathbf{h}_d + \mathbf{n}_d, \tag{10}$$

where $\mathbf{S}_t$ and $\mathbf{S}_d$ are diagonal matrices whose diagonal elements are equal to pilot symbols and data symbols respectively, and $\mathbf{h}_d$ and $\mathbf{h}_t$ are column vectors containing the channel states over data and pilot symbols, respectively. The resulting $\hat{\mathbf{h}}_d$ and its minimum MSE are then derived (see Appendix A) and given by

$$\hat{\mathbf{h}}_d = E\{\mathbf{h}_d\mathbf{y}^H\}E^{-1}\{\mathbf{y}\mathbf{y}^H\}\mathbf{y}, \tag{11}$$

$$\begin{aligned}\mathcal{M}(\hat{\mathbf{h}}_d) &\triangleq \mathrm{E}\{\tilde{\mathbf{h}}_d\tilde{\mathbf{h}}_d^H\} \\ &= \mathbf{R}_{\mathbf{h}_d} - \mathbf{R}_{\mathbf{h}_{dt}}\mathbf{S}_t^H(\mathbf{S}_t\mathbf{R}_{\mathbf{h}_t}\mathbf{S}_t^H + \sigma_n^2\mathbf{I})^{-1}\mathbf{S}_t\mathbf{R}_{\mathbf{h}_{dt}}^H\end{aligned} \tag{12}$$

where $\mathbf{R}_{\mathbf{h}_{dt}} = E\{\mathbf{h}_d\mathbf{h}_t^H\}$, $\mathbf{R}_{\mathbf{h}_t} = E\{\mathbf{h}_t\mathbf{h}_t^H\}$ and $\mathbf{R}_{\mathbf{h}_d} = E\{\mathbf{h}_d\mathbf{h}_d^H\}$, and these quantities are functions of placement $\mathcal{P}$.

It should be noted that for the flat fading channel, the training based MMSE channel estimator, $\hat{\mathbf{h}}_d = E\{\mathbf{h}_d\mathbf{y}_t^H\}E^{-1}\{\mathbf{y}_t\mathbf{y}_t^H\}\mathbf{y}_t$, has the same channel MSE as (12). Therefore, surprisingly, data observations do not provide any additional information to improve LMMSE estimation. In other words, the semi-blind LMMSE estimation is equivalent to the training based MMSE estimation. This is due to the statistical orthogonality of the channel $\mathbf{h}_d$ and data $\mathbf{s}_d$, which in turn results from the assumption of zero-mean data sequence and the independence of the channel and data.

3.2 Frequency Selective Channels

We assume that pilot symbols are inserted in delta like clusters each of which is of length $(2L+1)$. Each cluster contains only one non-zero symbol placed at the center of the cluster. It hence follows that for the types of training that we consider, P is of the form $r(2L+1)$, where r is an integer. The use of these pilot clusters leads to a separation of data and training observations which simplifies channel estimation, both for implementation and analysis. Such pilot clusters have also shown to be optimal in some sense in [1], [6].

Let $\mathbf{s}_t$ denote the length r column vector containing the r non-zero pilot symbols. Denote as $\mathbf{h}^{(l)}$ the column vector formed by $\{h_k[l]\}_{k=L+1}^{N+P}$, and $\mathbf{h}_d^{(l)}$ and $\mathbf{h}_t^{(l)}$ are column vectors formed from $\mathbf{h}^{(l)}$ by selecting those states corresponding to the data and non-zero pilot symbols, respectively. The training observation corresponding to $\mathbf{h}_t^{(l)}$ is given by

$$\mathbf{y}_t^{(l)} = \mathbf{S}_t \mathbf{h}_t^{(l)} + \mathbf{n}_t^{(l)}, \quad l = 0, \cdots, L; \tag{13}$$

where $\mathbf{S}_t = diag(\mathbf{s}_t)$. Due to the restriction on the structure of pilot clusters used and the model of channel, the estimation of $\mathbf{h}_d^{(l)}$ decouples for different l, and can be treated separately as if in a flat fading scenario (see Appendix B). The resulting minimum MSE for $\mathbf{h}_d^{(l)}$ is then given by

$$\begin{aligned}\mathcal{M}(\hat{\mathbf{h}}_d^{(l)}) = {}& \mathbf{R}_{\mathbf{h}_d^{(l)}} - \mathbf{R}_{\mathbf{h}_{dt}^{(l)}} \mathbf{S}_t^H (\mathbf{S}_t \mathbf{R}_{\mathbf{h}_t^{(l)}} \mathbf{S}_t^H \\ & + \sigma_n^2 \mathbf{I})^{-1} \mathbf{S}_t \mathbf{R}_{\mathbf{h}_{dt}^{(l)}}^H, \quad l = 0, \cdots, L.\end{aligned} \tag{14}$$

When $L = 0$, the MSE in (14) reduces to (12).

4. Optimal Placement of Pilot Symbols

In this section we formulate the problem of placement under both the performance metrics for the case of frequency selective channel with delta like training clusters. We also derive the structure of the optimal placement scheme for either performance metrics. A special case of this optimization gives the optimal placement of pilots for the flat fading channel.

4.1 The Minimax Optimization

The importance of this criterion stems from the fact that the maximum MSE of the channel tap estimate can be utilized to provide a lower bound on the performance of symbol-by-symbol detection schemes. This criterion is also important since the maximum MSE, as a worse case, is the limiting factor in any transmission designs.

4.1.1 The Maximum MSE

It can be seen from (14) that noise variance and the positions of the pilot symbol clusters affect the estimation performance of channel states associated with the data symbols. The resulting MSE is a complicated function

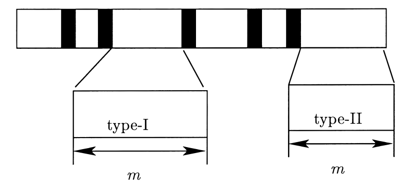

Fig. 2 Two types of data blocks in a packet.

of the placement $\mathcal{P}$, the channel correlation coefficient a and SNR. Obtaining the placement scheme that is optimal in general turns out to be a hard problem. Hence, in this paper, we limit ourselves to the case of high SNR (or $\sigma_n \to 0$). It has been shown that inserting training is an effective way of learning the channel only at high SNR [3]. At high SNR, some channel taps (the ones that are associated with non-zero pilots) can be estimated without any error. Therefore it is easy to see that no two training clusters should be placed consecutively. That is, between any two training clusters there should be at least one unknown symbol or equivalently $n = r$. But for different placements, the tracking performance can still be quite different and hence the optimization of the number of unknown symbols in each block still remains. If we let SNR tend to infinity i.e., $\lim_{\sigma_n \to 0} \hat{\mathbf{h}}_t(\sigma_n) = \mathbf{h}_t$, the expression in (14) reduces to

$$\mathcal{M}(\hat{\mathbf{h}}_d^{(l)}) = \mathbf{R}_{\mathbf{h}_d^{(l)}} - \mathbf{R}_{\mathbf{h}_{dt}^{(l)}} \mathbf{R}_{\mathbf{h}_t^{(l)}}^{-1} \mathbf{R}_{\mathbf{h}_{dt}^{(l)}}^H, \quad l = 0, \cdots, L. \tag{15}$$

As shown in Fig. 2, there are two types of unknown symbol blocks in a packet: those unknown symbol blocks that lie between two consecutive pilot clusters, which we will denote as type-I blocks, and those data blocks that reside at two ends of a packet, which we will denote as type-II blocks. We will first derive the maximum MSE and position at which this MSE is attained for each of these two block-types separately. This will help us find the maximum MSE over the whole packet along with the positions at which this MSE can be found and determine the optimal placement scheme. Before we proceed, we first simplify some notations. For an unknown symbol block with length m, we denote as $\mathbf{d}_m^{(l)}$ the column vector formed from $\mathbf{h}_d^{(l)}$ by selecting those states corresponding to the unknown data symbols in this block, and $\mathbf{t}^{(l)}$ the column vector formed from $\mathbf{h}_t^{(l)}$ by selecting the channel states over the non-zero symbols in those pilot clusters which are the immediate neighbors of the block.

Type-I blocks:
For a type-I unknown symbol block of length m, it has two immediate neighbor pilot clusters, and $\mathbf{t}^{(l)}$ is a 2-by-1 vector. By the Markov property of the channel in (1), given $\mathbf{t}^{(l)}$, $\mathbf{d}_m^{(l)}$ is independent of channel states corresponding to the rest of non-zero pilot symbols. Thus, $\mathbf{d}_m^{(l)}$ is only a function of $\mathbf{t}^{(l)}$. The minimum MSE in (15), in this case, can be rewritten as

$$\mathcal{M}(\hat{\mathbf{d}}_m^{(l)}) = \mathbf{R}_{\mathbf{d}_m} - \mathbf{R}_{\mathbf{d}_m\mathbf{t}}\mathbf{R}_{\mathbf{t}}^{-1}\mathbf{R}_{\mathbf{d}_m\mathbf{t}}^H, \quad (16)$$

where

$$\mathbf{R}_{\mathbf{d}_m} = \mathrm{E}\{\mathbf{d}_m^{(l)}\{\mathbf{d}_m^{(l)}\}^H\} = \sigma_h^2 \begin{bmatrix} 1 & \cdots & a^{m-1} \\ \vdots & \ddots & \vdots \\ a^{m-1} & \cdots & 1 \end{bmatrix} \quad (17)$$

$$\mathbf{R}_{\mathbf{d}_m\mathbf{t}} = \mathrm{E}\{\mathbf{d}_m^{(l)}\{\mathbf{t}^{(l)}\}^H\} = \sigma_h^2 a^{L+1} \begin{bmatrix} 1 & a^{m-1} \\ \vdots & \vdots \\ a^{m-1} & 1 \end{bmatrix} \quad (18)$$

$$\mathbf{R}_{\mathbf{t}} = \mathrm{E}\{\mathbf{t}^{(l)}\{\mathbf{t}^{(l)}\}^H\} = \sigma_h^2 \begin{bmatrix} 1 & a^{m+2L+1} \\ a^{m+2L+1} & 1 \end{bmatrix}. \quad (19)$$

It follows that the channel state MSE over each unknown symbol in the block is

$$\mathcal{M}(\hat{\mathbf{d}}_m^{(l)})_{ii} = \sigma_h^2 \Bigg(1 - a^{2(L+1)} \cdot \frac{a^{2(i-1)} - 2a^{2(m+L)} + a^{2(m-i)}}{1 - a^{2(m+2L+1)}}\Bigg) \quad i = 1, \cdots, m. \quad (20)$$

Note that $\mathcal{M}(\hat{\mathbf{d}}_m^{(l)})$ is not a function of l, this implies that the maximum MSE of the lth channel tap is the same for any l. Thus, the maximum MSE is given by

$$\mathcal{E}_1^{(m)} \triangleq \max_{l,i} \mathcal{M}(\hat{\mathbf{d}}_m^{(l)})_{ii} = \max_{1\le i\le m} \mathcal{M}(\hat{\mathbf{d}}_m^{(0)})_{ii} = \begin{cases} \dfrac{1 - a^{m+2L+1}}{1 + a^{m+2L+1}} & m \text{ odd} \\[2ex] \dfrac{2 - a^{m+2L} - a^{m+2L+2}}{1 - a^{2(m+2L+1)}} - 1 & m \text{ even} \end{cases} \quad (21)$$

and the position that gives the maximum MSE in the block is

$$i^* = \arg\max_{1\le i\le m} \mathcal{M}(\hat{\mathbf{d}}_m^{(0)})_{ii} \in \left\{ \left\lceil \frac{m+1}{2} \right\rceil, \left\lfloor \frac{m+1}{2} \right\rfloor \right\}. \quad (22)$$

Therefore, the maximum error appears in the middle of the data block, and is only a function of the block length m for fixed a. Thus, the maximum MSE can be calculated using (21) for any type-I unknown symbol block.

Type-II blocks:

Consider a type-II unknown symbol block with length m at the end of a packet. $\mathbf{t}^{(l)}$ in this case is a scalar. From (9), we note that the number of states of the lth tap corresponding to the unknown symbols in this block type is l-dependent, thus the length of $\mathbf{d}_m^{(l)}$ depends on different l. Again, by the Markov property of the channel, given $\mathbf{t}^{(l)}$, $\mathbf{d}_m^{(l)}$ is only a function of $\mathbf{t}^{(l)}$. Each channel state MSE $\mathcal{M}(\hat{\mathbf{d}}_m^{(l)})_{ii}$ in (16) is then given by

$$\mathcal{M}(\hat{\mathbf{d}}_m^{(l)})_{ii} = (1 - a^{2(L+i)})\sigma_h^2, \quad i = 1, \cdots, m - L + l. \quad (23)$$

The maximum MSE and position at which this MSE is attained is then given by

$$\mathcal{E}_2^{(m)} \triangleq \max_{i,l} \mathcal{M}(\hat{\mathbf{d}}_m^{(l)})_{ii} = \max_{i} \mathcal{M}(\hat{\mathbf{d}}_m^{(L)})_{ii} = 1 - a^{2(m+L)} \quad (24)$$

$$i^* = m. \quad (25)$$

For this block type, the maximum error occurs at the end of the packet, and appears on the last channel tap. By symmetry, for an unknown symbol block at the beginning of a packet, the maximum MSE occurs at the beginning of the packet and appears on the first channel tap. Again, the maximum MSE is only a function of data block size m for a given a.

4.1.2 The Optimal Placement

A. Packet starting and ending with pilot symbols

We constrain that every packet starts and ends with at least one pilot cluster. This implies $r \ge 2$, and also γ_1, $\gamma_r \ge (2L+1)$, and $\nu_1 = \nu_{r+1} = 0$. In this case, the packet contains only type-I blocks. Notice that $\mathcal{E}_1^{(m)}$ in (21) is a monotone increasing function of m. The optimal placement minimizing the maximum MSE is then obtained by minimizing the size of the longest unknown symbol block in a packet. The following theorem formalizes this result.

Theorem 1: If each data packet starts and ends with pilot clusters, i.e., $\nu_1 = \nu_{r+1} = 0$, and $P = r(2L+1)$, where $r \ge 2$. Under the assumed Rayleigh flat fading model, the optimal placement $(\boldsymbol{\nu}, \boldsymbol{\gamma})$ is given by

$$\gamma_i = 2L+1, \quad i = 2, \cdots, r;$$

$$\nu_i \in \left\{ \left\lceil \frac{N}{r-1} \right\rceil, \left\lceil \frac{N}{r-1} \right\rceil - 1 \right\}, \quad i = 2, \cdots, r. \quad (26)$$

Proof: See Appendix C.1.

Theorem 1 shows that, under the pilot cluster constraint, at high SNR, distributing pilot clusters periodically in the packet is optimal. Furthermore, the optimal placement is invariant under channel fading characteristics a. As an important case, letting $L = 0$, Theorem 1 gives the optimal placement for the flat fading channel, i.e., periodic pilot placement.

B. General case

In general, a packet contains both type-I and type-II blocks. In this case $\mathcal{E}(\mathcal{P})$ is obtained by comparing the maximum MSEs of the $(n+1)$ unknown symbol blocks, or equivalently comparing the maximum MSE of type-I blocks, denoted as $\mathcal{E}_I^{(m_I)}$ and that of type-II blocks, denoted as $\mathcal{E}_{II}^{(m_{II})}$. Intuitively, the optimal placement in

(6) is such that, the lengths of blocks in the same type are as equal as possible, and the maximum MSE of the two block types, i.e., $\mathcal{E}_I^{(m_I)}$ and $\mathcal{E}_{II}^{(m_{II})}$, are as equal as possible. This is verified in the following theorem.

Theorem 2: Under the assumed Rayleigh flat fading model, the optimal placement (ν, γ) is given by

$$\begin{aligned} &\gamma_i = 2L+1, \quad i = 1, \cdots, r; \\ &\nu_i \in \{v^*, v^*-1\}, i = 2, \cdots, r. \\ &\nu_1, \nu_{r+1} \in \left\{ \left\lceil \frac{N+q^*-(r-1)v^*}{2} \right\rceil, \right. \\ &\qquad \left. \left\lfloor \frac{N+q^*-(r-1)v^*}{2} \right\rfloor \right\} \end{aligned} \tag{27}$$

where

$$(v^*, q^*) = \begin{cases} (m_1^*, \bmod(\frac{N-2m_2^*}{r-1})) & \text{if } \mathcal{E}_1^{(m_1^*)} \leq \mathcal{E}_2^{(m_2^*+1)} \\ (m_1^*-1, \ 0), & \text{otherwise} \end{cases} \tag{28}$$

$$m_2^* = \mathop{\arg\min}_{0 \leq m_2 \leq \lfloor \frac{N-(r-1)}{2} \rfloor} \mathcal{E}_1^{(m_1)} - \mathcal{E}_2^{(m_2)} \quad \text{subject to } \mathcal{E}_1^{(m_1)} - \mathcal{E}_2^{(m_2)} \geq 0 \tag{29}$$

where

$$m_1 = \left\lceil \frac{N-2m_2}{r-1} \right\rceil. \tag{30}$$

Proof: See Appendix C.2.

Theorem 2 shows that at high SNR, in general, the optimal placement requires that each packet starts and ends with unknown symbol blocks of equal lengths, in between, pilot clusters comply with the optimal periodic placement as in the constrained case. The parameter for the optimal block length, $\{m^*, r^*\}$, is a function of the channel correlation coefficient a and the percentage of pilots. Finally, following Theorem 2, when $a \to 1$, ν_1, $\nu_{P+1} \to \frac{v^*+1}{4}$. This implies that for channels varies very slow, under the optimal placement, the lengths of the type-II blocks at two ends approach to $\frac{1}{4}$ of that of type-I blocks. Again, the optimal placement for flat fading channels is specified in Theorem 2 by letting $L = 0$.

4.2 The Cumulative MSE Optimization

In the previous section, we derived the placement scheme that minimizes the worst case performance, i.e., the maximum channel MSE in a packet. It is also important to find the placement scheme that minimizes the cumulative MSE.

In the following, we assume that every packet starts and ends with at least one pilot cluster, i.e., only type-I blocks are contained in the packet, and we consider the optimization in (7). Firstly, we show in Appendix D that for every placement $\mathcal{P}$,

$$\sum_{(k-i)\notin\mathcal{P}} \mathrm{E}\left\{ \left| \tilde{h}_k[i] \right|^2 \right\} = L \sum_{k\notin\mathcal{P}} \mathrm{E}\left\{ \left| \tilde{h}_k[0] \right|^2 \right\}. \tag{31}$$

Hence, the optimization can be rewritten as

$$\mathcal{P}^* = \arg\min_{\mathcal{P}} \sum_{k\notin\mathcal{P}} \mathrm{E}\left\{ \left| \tilde{h}_k[0] \right|^2 \right\}, \tag{32}$$

and we only need to consider the estimate of $\mathbf{h}_d^{(0)}$.

Let $\mathbf{J}_{\boldsymbol{\gamma}}$ be a selection matrix of size $r \times (N+P)$ such that

$$\mathbf{s}_t = \mathbf{J}_{\boldsymbol{\gamma}} \mathbf{s}. \tag{33}$$

Let $\mathbf{J}_{\boldsymbol{\nu}}$ be the $N \times (N+P)$ selection matrix such that

$$\mathbf{s}_d = \mathbf{J}_{\boldsymbol{\nu}} \mathbf{s}. \tag{34}$$

It is simple to show that, under the assumption of equi-powered non-zero pilot symbols, $\mathcal{M}(\hat{\mathbf{h}}_d^{(0)})$ in (14) can be equivalently rewritten as

$$\mathcal{M}(\hat{\mathbf{h}}_d^{(0)}) = \mathbf{J}_{\boldsymbol{\nu}} \left(\mathbf{R}_{\mathbf{h}_{(N+P)}}^{-1} + \frac{\sigma_p^2}{\sigma_n^2} \mathbf{J}_{\boldsymbol{\gamma}}^H \mathbf{J}_{\boldsymbol{\gamma}} \right)^{-1} \mathbf{J}_{\boldsymbol{\nu}}^H \tag{35}$$

where $\mathbf{R}_{\mathbf{h}_n}$ is the Toeplitz Hermitian matrix given by

$$\mathbf{R}_{\mathbf{h}_n} = \sigma_h^2 \begin{bmatrix} 1 & a & a^2 & \cdots & a^{n-1} \\ a & 1 & a & \cdots & a^{n-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \vdots \\ a^{n-1} & & & a & 1 \end{bmatrix}. \tag{36}$$

The optimal placement can then be obtained as

$$\begin{aligned} \mathbf{r}^* &= \arg\min_{\mathbf{r}} \operatorname{tr}\left(\mathcal{M}(\hat{\mathbf{h}}_d^{(0)}) \right) \\ &= \arg\min_{\mathbf{r}} \operatorname{tr} \mathbf{J}_{\boldsymbol{\nu}} \left(\mathbf{R}_{\mathbf{h}_{(N+P)}}^{-1} + \frac{\sigma_p^2}{\sigma_n^2} \mathbf{J}_{\boldsymbol{\gamma}}^H \mathbf{J}_{\boldsymbol{\gamma}} \right)^{-1} \mathbf{J}_{\boldsymbol{\nu}}^H. \\ &\triangleq \arg\min_{\mathbf{r}} M(N, P, \mathbf{r}, \sigma_p^2, \sigma_n). \end{aligned} \tag{37}$$

4.2.1 Problem Formulation

We now state some properties of $\mathbf{R}_{\mathbf{h}_n}$ that are crucial in the optimization. The inverse of $\mathbf{R}_{\mathbf{h}_n}$ is a tri-diagonal matrix [9] (page 409). For $0 < a < 1$, the matrix $\frac{(1-a^2)}{a}\mathbf{R}_{\mathbf{h}_n}^{-1}$ has an entry -1 in every position of the super-diagonal and sub-diagonal and has main diagonal entries $\frac{1}{a}, a+\frac{1}{a}, \cdots, a+\frac{1}{a}, \frac{1}{a}$.

From (35), it can be seen that the diagonal elements of the covariance matrix $\mathcal{M}(\hat{\mathbf{h}}_d^{(0)})$ are in fact the diagonal elements of the inverse of a symmetric tridiagonal matrix. The following lemma can be used in finding an expression for these terms.

Lemma 1: Let $\mathbf{A}_n$ is an $n \times n$ tridiagonal matrix with -1 in the sub-diagonal and super-diagonal places and main diagonal $(a_{11}, a_{22}, \cdots, a_{nn})$. If $diag(\mathbf{A}_n^{(-1)}) = (b_{11}, b_{22}, \cdots, b_{nn})$, then

$$b_{ii} = \frac{1}{a_{ii} - f^+(i-1) - f^-(n-i)}, \quad i = 1, \cdots, n, \tag{38}$$

where the functions $f^+(\cdot)$ and $f^-(\cdot)$ are defined by the following recursions

$$\begin{aligned} f^+(i) &= \frac{1}{a_{ii} - f^+(i-1)}, f^+(0) = 0,\ i = 1, \cdots, n \\ f^-(i) &= \frac{1}{a_{n-i+1,n-i+1} - f^-(i-1)}, \\ f^-(0) &= 0,\ i = 1, \cdots, n. \end{aligned} \tag{39}$$

Note that in order to define the functions $f^+(\cdot)$ and $f^-(\cdot)$, we only need $(a_{11}, \cdots, a_{nn})$.

Given a placement $\mathcal{P}$ we use the above lemma in order to obtain an expression for the cumulative MSE $M(N, P, \mathcal{P}, \sigma_p^2, \sigma_n)$. Let $m(i, N, P, \mathcal{P}, \sigma_p^2, \sigma_n)$ be the MSE of the channel estimate over the i^{th} symbol, so that

$$M(N, P, \mathcal{P}, \sigma_p^2, \sigma_n) = \sum_{i: i \notin \mathcal{P}} m(i, N, P, \mathcal{P}, \sigma_p^2, \sigma_n). \tag{40}$$

Given a placement $\mathcal{P}$, let $\mathcal{P}'$ be the set of positions at which the non-zero pilot symbols are present. We then define the functions $f^+(i, \mathcal{P})$ and $f^-(i, \mathcal{P})$ as in (39) with

$$\begin{aligned} a_{ii} &= \beta + \frac{1-a^2}{a} \frac{\sigma_p^2}{\sigma_n^2}, \quad i \in \mathcal{P}' \\ &= \beta, \quad i \notin \mathcal{P}' i \notin \{1, N+P\} \\ &= \frac{1}{a}, \quad i = 1, (N+P), \end{aligned} \tag{41}$$

where $\beta = a + \frac{1}{a}$. Then we obtain the following lemma quite easily from (35) and Lemma 1.

Lemma 2: The quantity $m(i, N, P, \mathcal{P}, \sigma_p^2, \sigma_n)$ is given by

$$\begin{aligned} &m(i, N, P, \mathcal{P}, \sigma_p^2, \sigma_n) \\ &\quad = \frac{1-a^2}{a} \frac{1}{a_{ii} - f^+(i-1, \mathcal{P}) - f^-(n-i, \mathcal{P})} \\ &\qquad i = 1, \cdots, n. \end{aligned} \tag{42}$$

where a_{ii} and the functions $f^+(i, \mathcal{P}), f^-(i, \mathcal{P})$ are defined as above.

In spite of this structure, it is in general quite difficult to obtain the optimal placement schemes. Heuristically, when we place the pilot symbols in clusters, the channel estimates over the pilot symbols is good where as the tracking (channel estimates over the unknown symbols) is poor. When we spread the pilot symbols, the channel estimates over the pilot symbols deteriorate but the tracking improves. Hence the optimal placement is a trade-off between these two quantities.

We again limit ourselves to the optimal placement problem at high SNR. The cumulative MSE at high SNR is denoted by $G(N, P, \mathcal{P})$. That is,

$$G(N, P, \mathcal{P}) = \lim_{\sigma_n \to 0} M(N, P, \mathcal{P}, \sigma_p^2, \sigma_n). \tag{43}$$

It is possible to simplify the expression obtained in Lemma 2 at high SNR by letting σ_n go to zero in $m(i, N, P, \mathcal{P}, \sigma_p^2, \sigma_n)$.

Lemma 3: The cumulative MSE is given by

$$G(N, P, \mathbf{r}) = \sum_{i=1}^{n} g(\nu_i), \tag{44}$$

where $g(\nu_i)$ is the cumulative MSE of the symbols in the i^{th} unknown symbol block. We have

$$g(\nu_i) = \sum_{j=0}^{\nu_i - 1} \frac{1}{\beta - f(j+L) - f(\nu_i + L - 1 - j)}, \tag{45}$$

where

$$\begin{aligned} f(j) &= \frac{1}{\beta - f(j-1)}, \quad \forall j \geq 1 \\ f(0) &= 0. \end{aligned} \tag{46}$$

4.2.2 The Optimal Placement

The optimal placement of pilot symbols for high SNR can now be found as

$$\mathbf{r}^* = \arg \min_{\mathbf{r}} G(N, P, \mathbf{r}). \tag{47}$$

As mentioned before, since perfect estimates of the channel tap over pilot symbols is obtained at high SNR, pilot clusters must always be separated, i.e., $n = r$ and $\gamma_i = (2L+1)$, $i = 1, \cdots, r$. Only $\boldsymbol{\nu}$, the number of unknown symbols in each block, is left to be optimized. We claim that the MSE is minimized by placing the unknown symbols such that the length of each block is as equal as possible. Due to Lemma 3, it is enough if we show that

$$2g(n) \leq g(n+1) + g(n-1), \quad \forall n \geq 0. \tag{48}$$

This is in fact true and we hence have the following theorem.

Theorem 3: If each packet starts and ends with pilot clusters, under the assumption that $P = r(2L+1)$, where $r \geq 2$, the placement $\mathbf{r}^*$, that is optimal with respect to (7) at high SNR is given by

$$\gamma_i = 2L + 1 \quad i = 1, \cdots, r \tag{49}$$

$$\nu_i \in \left\{ \left\lceil \frac{N}{r-1} \right\rceil, \left\lceil \frac{N}{r-1} \right\rceil - 1 \right\}. \tag{50}$$

Proof: See Appendix E.

For the flat fading case where $L = 0$, the above indicates that periodic pilot symbols placement is optimal. From Theorem 1 and Theorem 3, we notice that the optimization under both performance criteria results in the same optimal placement, which in certain degree demonstrates the generality and advantage of this placement. Finally, we want to point out that the optimal placements we have obtained are high SNR results, while their optimality at low SNR is not guaranteed. Indeed, there exist examples where the optimality does not hold at low SNR. We will give examples in Sect. 6 to emphasize this point.

5. Packets with Superimposed Training

In previous sections, we have considered optimizing the placement of pilot symbols when they are inserted in time. An alternate method of inserting pilots is by superposition. This is a technique that has recently attracted a lot of attention. Under superimposed training, the system equation can be written as

$$y_k = \sum_{i=0}^{L} h_k[i](s_{d,k-i} + s_{t,k-i}) + n_k, \quad k = L+1, \cdots, N+P, \tag{51}$$

where $s_{d,k}$ and $s_{t,k}$ are data and pilot symbols at time k, respectively. The power of data and pilot symbols are denoted by ρ_d^2 and ρ_t^2. For time-varying channels, it appears that this scheme might have the advantage of improving tracking capability due to the continuous presence of training in the data stream. Therefore, we are interested to compare the channel estimation performance between the superimposed training scheme and time-divisioned placement scheme.

When $L = 0$, let $\mathbf{h}$ be the column vector formed by $\{h_k[0]\}_{k=1}^{N+P}$. Then, the LMMSE estimator can be derived using the similar formula in (11), and the resulting MSE matrix (see Appendix F) is

$$\mathcal{M}(\hat{\mathbf{h}}) = \left(\mathbf{R}_{\mathbf{h}_{(N+P)}}^{-1} + \left(\frac{\rho_t^2}{\rho_d^2 \sigma_h^2} + \frac{\sigma_n^2}{\rho_t^2} \right) \mathbf{I} \right)^{-1} \tag{52}$$

where $\mathbf{R}_{\mathbf{h}_{(N+P)}}$ is the same defined in (36).

For frequency selective channels, the channel MSE at each time can also be obtained following the standard LMMSE derivation procedure. We will not give the detailed description here. To compare the performance between this scheme and the time-divisioned placement scheme, we should keep the total power allocated to data and training symbols in a packet the same in both schemes, i.e.,

$$\rho_t^2 = \frac{P}{N+P}\sigma_t^2, \quad \rho_d^2 = \frac{N}{N+P}\sigma_d^2. \tag{53}$$

The detailed comparison is given in the next section.

6. Simulations

We compared the estimation performance for the fading channel under different pilot placement strategies. The channel was Gaussian with variance $\sigma_h^2 = 1$ and $L = 0$. Figure 3 shows the maximum MSE vs. channel correlation coefficient a for different placement schemes at SNR=30 dB. The percentage of pilots is 33%. Notice that when $a = 0$ (the channel varies independently) or $a = 1$ (the constant channel), no gain in the optimal placement as expected. The efficiency of the optimal placement becomes apparent for a between 0 and 1. We see that there is a significant gain by placing pilot symbols optimally. Also, further performance improvement can be obtained by using the placement in Theorem 2, comparing to the placement in Theorem 1. However, we also notice that when a approaches to 1, the performance gap under the optimal placements in the two cases is very small. This implies that the constraint of starting and ending with pilot symbols results in negligible performance loss when a is close to 1. For bandwidths in the 10 kHz range and Doppler spreads of order 100 Hz, the parameter a typically ranges between 0.9 and 0.99 [10]. Figure 4 shows an example of non-optimality of the placement we obtained at low SNR. We see that a is close to 1, clustering pilot symbols results in better performance. This indicates that at high noise level and relatively slow fading, good training estimation performance takes an important place in the trade-off between channel estimated over pilots and tracking ability over data.

Also, we compared the performance under superimposed training with the above time-divisioned placements in both figures. In Fig. 3, we observe that when a is small, superimposed scheme outperforms the time-divisioned scheme, while when a closes to 1, the later scheme outperforms the former one. This shows that,

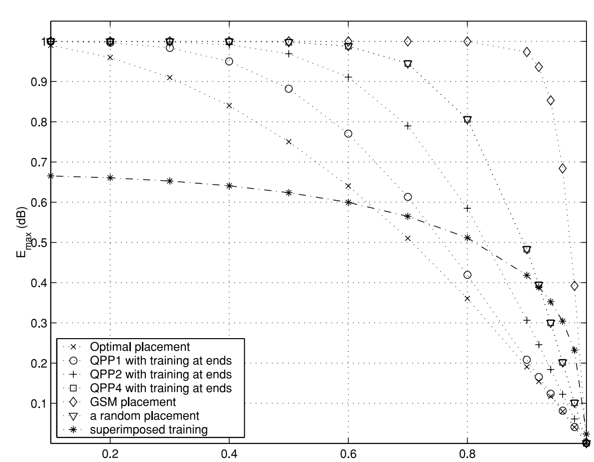

Fig. 3 $\mathcal{E}_{max}(\mathcal{P})$ vs. a. $L = 0$, SNR=30 dB, $N + P = 120$, $\eta = 33\%$, $\sigma_h^2 = 1$.

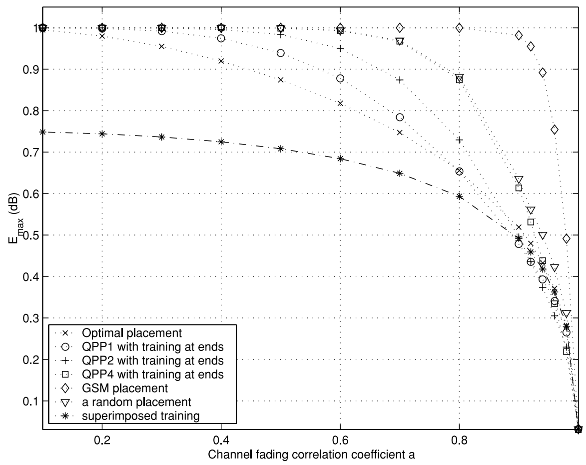

Fig. 4 $\mathcal{E}_{max}(\mathcal{P})$ vs. η. $L=0$, SNR=0 dB, $N+P=120$, $\eta=33\%$, $\sigma_h^2=1$.

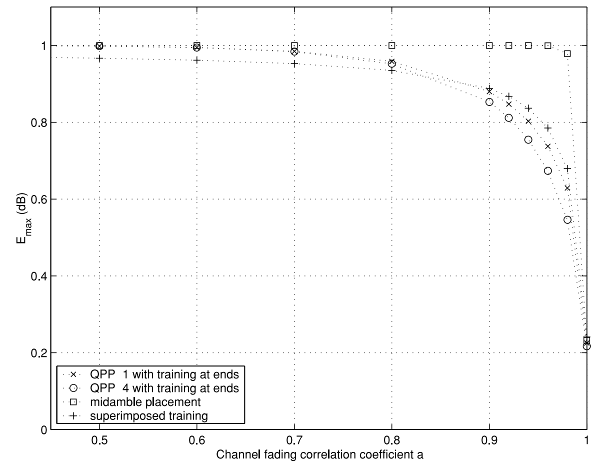

Fig. 6 $\mathcal{E}_{max}(\mathcal{P})$ vs. η. $L=2$, SNR=−10 dB, $N+P=120$, $\eta=33\%$, $\sigma_h^2=1$.

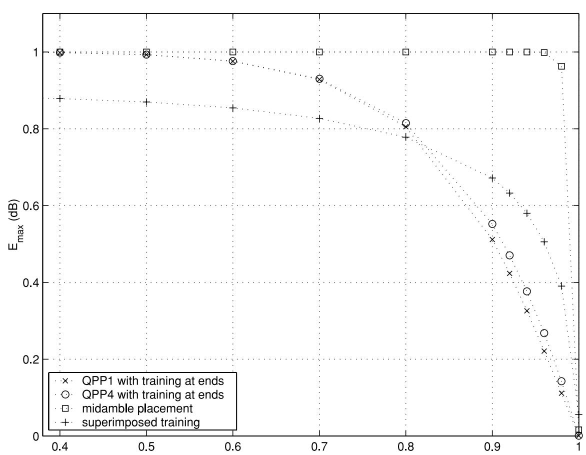

Fig. 5 $\mathcal{E}_{max}(\mathcal{P})$ vs. a. $L=2$, SNR=30 dB, $N+P=120$, $\eta=33\%$, $\sigma_h^2=1$.

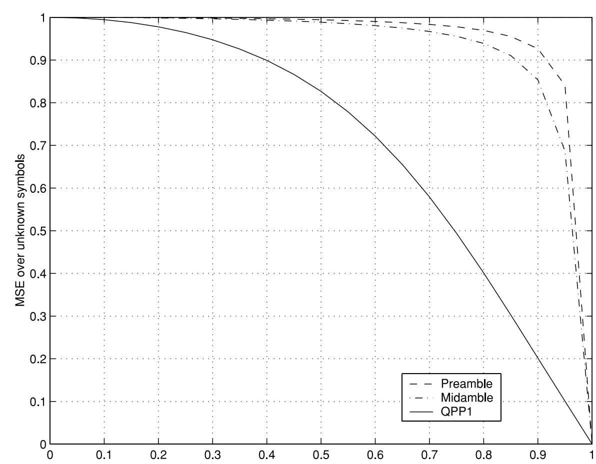

Fig. 7 Variation of MSE with a for $L=0$ and SNR $=25$ dB.

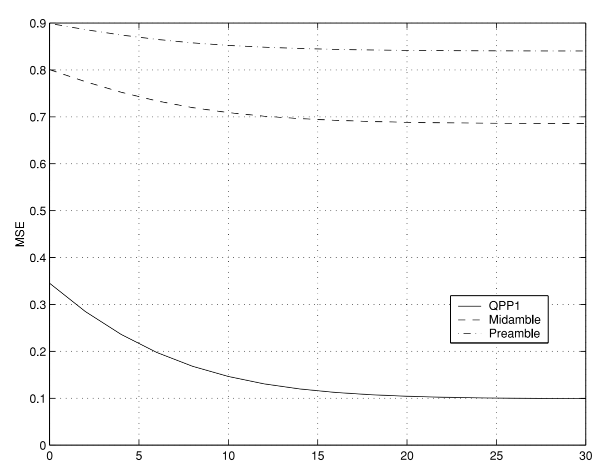

Fig. 8 Variation of MSE with SNR for $L=0$ and $a=0.95$.

for fast varying channels, the superimposed scheme shows its advantage of constant presence of training which helps tracking the channel state. However, the presence of data interference prevent accurate estimation performance, which shows its disadvantage when the channel varies slow. In Fig. 4, when the noise level becomes higher, we see that the region that the superimposed scheme outperforms the time-divisioned scheme becomes larger.

Finally, for frequency selective channels, we relaxed the restricted pilot cluster structure described in Sect. 3 to general pilot sequences, and compare the performance for different placements. In Figs. 5 and 6, we plotted the maximum MSE vs. a when $L=2$ and the pilot sequence consists of constant modulus symbols. We observe that similar performance behaviors as in Figs. 3 and 4 are also shown in this case.

For cumulative MSE criterion, Fig. 7 shows the variation of the cumulative MSE of the channel estimate over the data symbols with the correlation parameter a. We choose $N=116$, $P=32$, $L=0$ and SNR $=$ 25 dB. The QPP-1 placement scheme is optimal for this scenario. As expected, we find that there is no gain in the optimal placement for $a = 0$ and $a = 1$. Even if a falls slightly below one, we see that the optimal placement gives a large gain. Figure 8 plots the variation of cumulative MSE of the channel estimate over data symbols with SNR. As before, we choose $N = 116$, $P = 32$, $L = 0$ and $a = 0.95$. We find that there is a significant gain to be obtained by placing pilots optimally.

7. Conclusion

In this paper, we considered the placement of pilot symbols for packet transmission over a time-varying fading channel. Both flat and frequency selective fading channels were considered and the time-variation of the channels was modeled by a Gauss-Markov process. For frequency selective fading channels, we constrained the pilot clusters of length $(2L+1)$ with each cluster containing only one non-zero pilot symbol placed at the center. It was shown that at high SNR, pilot symbols should be placed periodically to minimize both the maximum MSE and the cumulative MSE over data symbols in a packet. We also showed that the optimal placement under flat fading channel is a special case of the above when letting $L = 0$, i.e., periodic pilot symbol placement. In spite of the generality of this placement at high SNR, we present an example to emphasize that this optimality is not guaranteed to hold at low SNR. Finally, we also compared this periodic scheme with superimposed scheme through simulations. The later scheme shows the advantage under fast fading scenario or high noise environment, while the former scheme is better when the channel fades slowly and the noise level is low.

References

[1] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for frequency selective block-fading channels," IEEE Trans. Inf. Theory, vol.48, no.8, pp.2338–2353, Aug. 2002.
[2] A.W. Marshall and I. Olkin, Inequalities: Theory of Majorization and its Applications, Academic Press, 111 Fifth Avenue, New York, NY 10003, 1979.
[3] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links," Submitted to IEEE Trans. Inf. Theory, Aug. 2000.
[4] C. Budianu and L. Tong, "Channel estimation for space-time orthogonal block code," IEEE Trans. Signal Process., vol.50, no.10, pp.2515–2528, Oct. 2002.
[5] J.K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," IEEE Trans. Veh. Technol., vol.40, no.4, pp.686–693, Nov. 1991.
[6] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," IEEE Trans. Signal Process., vol.50, no.12, pp.3055–3069, Dec. 2002.
[7] M. Dong and L. Tong, "Optimal placement of training for channel estimation and tracking," MILCOM, vol.2, pp.1195–1199, Vienna, Virginia, Oct. 2001.
[8] M. Dong, L. Tong, and B.M. Sadler, "Optimal pilot placement for time-varying channels," submitted to IEEE Trans. Signal Process., Jan. 2003. Available at http://acsp.ece.cornell.edu/pubJ.html
[9] R.A. Horn and C.R. Johnson, Matrix Analysis, Cambridge University Press, New York, NY, 1985.
[10] M. Medard, I. Abou-Faycal, and U. Madhowf, "Adaptive coding with pilot signals," 38th Allerton Conference, Oct. 2000.
[11] S. Ohno and G.B. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," IEEE Trans. Wireless Commun., vol.1, no.4, pp.712–720, Oct. 2002.
[12] S. Ohno and G.B. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," ICASSP, vol.4, pp.2389–2392, May 2001.

Appendix A: Derivation of (11) and (12)

Based on observation vector $\mathbf{y}$ from a packet, the semi-blind LMMSE estimator can be derived using the orthogonality principle

$$E\{(\hat{\mathbf{h}}_d - \mathbf{h}_d)\mathbf{y}^H\} = 0. \tag{A$\cdot$1}$$

Therefore, we obtain Eq. (11). The MSE of $\hat{\mathbf{h}}_d$ is then given by

$$\mathcal{M}(\hat{\mathbf{h}}_d) = E\{\tilde{\mathbf{h}}_d\tilde{\mathbf{h}}_d^H\} = \mathbf{R}_{\mathbf{h}_d} - \mathbf{R}_{\mathbf{h}_d\mathbf{y}}\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{h}_d\mathbf{y}}^H \tag{A$\cdot$2}$$

where $\mathbf{R}_{\mathbf{h}_d\mathbf{y}} = E\{\mathbf{h}_d\mathbf{y}^H\}$ and $\mathbf{R}_{\mathbf{y}} = E\{\mathbf{y}\mathbf{y}^H\}$. Without loss of generality (w.l.o.g.), let

$$\mathbf{h} = \begin{bmatrix}\mathbf{h}_t \\ \mathbf{h}_d\end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix}\mathbf{y}_t \\ \mathbf{y}_d\end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix}\mathbf{S}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_d\end{bmatrix}.$$

Then,

$$\mathbf{R}_{\mathbf{h}_d\mathbf{y}} = [E\{\mathbf{h}_d\mathbf{y}_t^H\},\ E\{\mathbf{h}_d\mathbf{y}_d^H\}] = [\mathbf{R}_{\mathbf{h}_{dt}}\mathbf{S}_t^H,\ \mathbf{0}]$$

where we use the zero mean property of the data sequence. Furthermore,

$$\mathbf{R}_{\mathbf{y}}^{-1} = \begin{bmatrix}\mathbf{R}_{\mathbf{y}_t} & \mathbf{R}_{\mathbf{y}_{td}} \\ \mathbf{R}_{\mathbf{y}_{dt}} & \mathbf{R}_{\mathbf{y}_d}\end{bmatrix}^{-1} \tag{A$\cdot$3}$$

where $\mathbf{R}_{\mathbf{y}_t} = E\{\mathbf{y}_t\mathbf{y}_t^H\}$, $\mathbf{R}_{\mathbf{y}_d} = E\{\mathbf{y}_d\mathbf{y}_d^H\}$, and $\mathbf{R}_{\mathbf{y}_{td}} = E\{\mathbf{y}_t\mathbf{y}_d^H\}$. Since

$$\begin{aligned}\mathbf{R}_{\mathbf{y}_{td}} &= E\{(\mathbf{S}_t\mathbf{h}_t + \mathbf{n}_t)(\mathbf{S}_d\mathbf{h}_d + \mathbf{n}_d)^H\} \\ &= E\{\mathbf{S}_t\mathbf{h}_t\mathbf{h}_d^H\mathbf{S}_d\} \\ &= \mathbf{0}\end{aligned} \tag{A$\cdot$4}$$

where we use the zero mean property of $\mathbf{S}_d$ and the independence of $\mathbf{h}$, $\mathbf{S}$ and $\mathbf{n}$. Thus, Eq. (A$\cdot$3) is block diagonal and from (A$\cdot$2), we have

$$\begin{aligned}\mathcal{M}(\hat{\mathbf{h}}_d) &= \mathbf{R}_{\mathbf{h}_d} - [\mathbf{R}_{\mathbf{h}_{dt}}\mathbf{S}_t^H, \mathbf{0}]\begin{bmatrix}\mathbf{R}_{\mathbf{y}_t}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{y}_d}^{-1}\end{bmatrix} \\ &\quad \cdot [\mathbf{R}_{\mathbf{h}_{dt}}\mathbf{S}_t^H, \mathbf{0}]^H \\ &= \mathbf{R}_{\mathbf{h}_d} - \mathbf{R}_{\mathbf{h}_{dt}}\mathbf{R}_{\mathbf{y}_t}^{-1}\mathbf{R}_{\mathbf{h}_{dt}}^H.\end{aligned} \tag{A$\cdot$5}$$

Therefore, we have Eq. (12).

Appendix B: Decoupling of the Estimation of $\mathbf{h}_d^{(l)}$ for Different l, and Derivation of (14)

Similar as in (11), by the orthogonality principle, the semi-blind LMMSE estimator for $\mathbf{h}_d^{(l)}$ is given by

$$\hat{\mathbf{h}}_d^{(l)} = \mathbf{R}_{\mathbf{h}_d^{(l)}\mathbf{y}}\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{y}, \quad l = 0, \cdots, L, \tag{A$\cdot$6}$$

where $\mathbf{R}_{\mathbf{h}_d^{(l)}\mathbf{y}} = \mathrm{E}\{\mathbf{h}_d^{(l)}\mathbf{y}^H\}$ and $\mathbf{R}_{\mathbf{y}} = \mathrm{E}\{\mathbf{y}\mathbf{y}^H\}$. Under the structure of delta-like pilot clusters of length $(2L+1)$, data and training observations are separated.

W.o.l.g, let

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_d \end{bmatrix}, \ \mathbf{y}_t = \begin{bmatrix} \mathbf{y}_t^{(0)} \\ \vdots \\ \mathbf{y}_t^{(L)} \end{bmatrix}$$

where $\mathbf{y}_t^{(l)}$ is the column vector consisting of training observations corresponding to the lth channel tap. Then, for $\mathbf{h}_d^{(0)}$,

$$\begin{aligned} \mathbf{R}_{\mathbf{h}_d^{(0)}\mathbf{y}} &= \left[\mathrm{E}\left\{\mathbf{h}_d^{(0)}\mathbf{y}_t^H\right\}, \mathrm{E}\left\{\mathbf{h}_d^{(0)}\mathbf{y}_d^H\right\}\right] \\ &= \left[\mathrm{E}\left\{\mathbf{h}_d^{(0)}[\mathbf{y}_t^{(0)}, \cdots, \mathbf{y}_t^{(L)}]^H\right\}, \mathbf{0}\right] \qquad (\mathrm{A}\cdot 7) \\ &= \left[\mathrm{E}\left\{\mathbf{h}_d^{(0)}{\mathbf{y}_t^{(0)}}^H\right\}, \mathbf{0}, \cdots, \mathbf{0}\right] \qquad (\mathrm{A}\cdot 8) \end{aligned}$$

where (A·7) follows from the assumption that data symbols have zero mean, and are independent of the channel taps, and (A·8) follows from the assumption that channel taps are i.i.d. with zero mean.

$\mathbf{R}_{\mathbf{y}}$ can be written as

$$\mathbf{R}_{\mathbf{y}} = \begin{bmatrix} \mathbf{R}_{\mathbf{y}_t} & \mathbf{R}_{\mathbf{y}_{td}} \\ \mathbf{R}_{\mathbf{y}_{dt}} & \mathbf{R}_{\mathbf{y}_d} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{\mathbf{y}_t} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{y}_d} \end{bmatrix}, \tag{A$\cdot$9}$$

where $\mathbf{R}_{\mathbf{y}_{td}} = \mathrm{E}\{\mathbf{y}_t\mathbf{y}_d^H\} = \mathbf{y}_t E\{\mathbf{y}_d\}^H = \mathbf{0}$, which results from the assumption of zero mean data symbols. And also

$$\mathbf{R}_{\mathbf{y}_t} = \mathrm{E}\{\mathbf{y}_t\mathbf{y}_t^H\} = \begin{bmatrix} \mathbf{R}_{\mathbf{y}_t^{(0)}} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{R}_{\mathbf{y}_t^{(L)}} \end{bmatrix}, \tag{A$\cdot$10}$$

where $\mathrm{E}\{\mathbf{y}_t^{(i)}\{\mathbf{y}_t^{(j)}\}^H\} = 0$ due to the i.i.d. and zero mean channel taps. Hence, $\hat{\mathbf{h}}_d^{(0)}$ becomes

$$\hat{\mathbf{h}}_d^{(0)} = \mathbf{R}_{\mathbf{h}_d^{(0)}\mathbf{y}_t^{(0)}}\mathbf{R}_{\mathbf{y}_t^{(0)}}^{-1}\mathbf{y}_t^{(0)} \tag{A$\cdot$11}$$

Therefore, the estimation of $\mathbf{h}_d^{(0)}$ is only a function of $\mathbf{y}_t^{(0)}$, thus decouples from other channel taps. The same derivation can be obtained for any channel tap $\mathbf{h}_d^{(l)}$, $l = 0, \cdots, L$. The MSE of $\mathbf{h}_d^{(l)}$ is then given by

$$\begin{aligned} \mathcal{M}(\hat{\mathbf{h}}_d^{(l)}) &= \mathrm{E}\{\tilde{\mathbf{h}}_d^{(l)}\{\tilde{\mathbf{h}}_d^{(l)}\}^H\} \\ &= \mathbf{R}_{\mathbf{h}_d^{(l)}} - \mathbf{R}_{\mathbf{h}_d^{(l)}\mathbf{y}_t^{(l)}}\mathbf{R}_{\mathbf{y}_t^{(l)}}^{-1}\mathbf{R}_{\mathbf{h}_d^{(l)}\mathbf{y}_t^{(l)}}^H \\ &= \mathbf{R}_{\mathbf{h}_d^{(l)}} - \mathbf{R}_{\mathbf{h}_{dt}^{(l)}}\mathbf{S}_t^H(\mathbf{S}_t\mathbf{R}_{\mathbf{h}_t^{(l)}}\mathbf{S}_t^H + \sigma_n^2\mathbf{I})^{-1} \\ &\quad \cdot \mathbf{S}_t\mathbf{R}_{\mathbf{h}_{dt}^{(l)}}^H. \end{aligned} \tag{A$\cdot$12}$$

Appendix C

C.1 Proof of Theorem 1

Only type-I blocks are presented in a packet in this case. As we have discussed, at high SNR, $n = r$, and there are $(r-1)$ data blocks. Let $\nu_k(\mathcal{P})$ be the length of the ith block for a given placement $\mathcal{P}$ ($\nu_1 = \nu_{r+1} = 0$).

W.o.l.g., let $\nu_1(\mathcal{P}) \le \cdots \le \nu_r(\mathcal{P})$, equivalently, let

$$\begin{aligned} &\nu_k(\mathcal{P}) = \nu_r(\mathcal{P}) - i_k(\mathcal{P}), \\ &i_k(\mathcal{P}) \ge 0, \ k = 2, \cdots, r-1. \end{aligned} \tag{A$\cdot$13}$$

Since $\sum_{k=2}^r \nu_k(\mathcal{P}) = N$, we have

$$(r-1)\nu_r(\mathcal{P}) = N + \sum_{k=2}^{r-1} i_k(\mathcal{P}). \tag{A$\cdot$14}$$

It is easy to verify that $\mathcal{E}_1^{(m)}$ in (21) is a monotone increasing function of m. Hence, $\mathbb{P}$ in (4) can be rewritten as

$$\begin{aligned} \mathbb{P} &= \{\mathcal{P}_\# : \mathcal{E}(\mathcal{P}_\#) = \arg\min_{\mathcal{P}} \nu_r(\mathcal{P})\} \\ &= \left\{\mathcal{P}_\# : \mathcal{E}(\mathcal{P}_\#) = \arg\min_{\mathcal{P}} \sum_{k=2}^{r-1} i_k(\mathcal{P})\right\}. \end{aligned} \tag{A$\cdot$15}$$

It is straight forward to see that

$$\min_{\mathcal{P}} \sum_{k=2}^{r-1} i_k(\mathcal{P}) = (r-1)\left\lceil \frac{N}{r-1} \right\rceil - N. \tag{A$\cdot$16}$$

Therefore,

$$\mathbb{P} = \left\{\mathcal{P}_\# : \sum_{k=2}^{r-1} i_k(\mathcal{P}_\#) = (r-1)\left\lceil \frac{N}{r-1} \right\rceil - N\right\}. \tag{A$\cdot$17}$$

Recall that $\mathbf{I}_{\mathcal{P}_\#}$ is the index set of the maximum MSE in a packet. Also, by the property of type-I blocks, the number of positions at which the maximum MSE of the estimate in $\mathbf{h}_d^{(l)}$ is attained is the same for any l. It follows that

$$\begin{aligned} |\mathbf{I}_{\mathcal{P}_\#}| = (L+1)|\{&i_k(\mathcal{P}_\#) : i_k(\mathcal{P}_\#) = 0, \\ &\text{subject to (A}\cdot\text{16)}\}| \end{aligned}$$

and thus

$$\begin{aligned} \mathcal{P}^* &= \arg\min_{\mathcal{P}_\# \in \mathbb{P}} |\mathbf{I}_{\mathcal{P}_\#}| \\ &= \arg\min_{\mathcal{P}_\# \in \mathbb{P}} |\{i_k(\mathcal{P}_\#) : i_k(\mathcal{P}_\#) = 0, \end{aligned}$$

subject to (A· 16)}|

Then, we have

$$\min_{\mathcal{P}_\# \in \mathbb{P}} |\{i_k(\mathcal{P}_\#) : i_k(\mathcal{P}_\#) = 0, \text{ subject to (A· 16)}\}| = \max_{\mathcal{P}_\# \in \mathbb{P}} |\{i_k(\mathcal{P}_\#) : i_k(\mathcal{P}_\#) > 0, \text{ subject to (A· 16)}\}| = (r-1)\left\lceil \frac{N}{r-1} \right\rceil - N, \quad \text{(A· 18)}$$

where

$$i_k(\mathcal{P}^*) = \begin{cases} 1 & k = 1, \cdots, (r-1)\lceil \frac{N}{r-1} \rceil - N \\ 0 & \text{otherwise} \end{cases} \quad \text{(A· 19)}$$

and

$$\nu_r = \left\lceil \frac{N}{r-1} \right\rceil. \quad \text{(A· 20)}$$

Therefore,

$$\nu_k \in \left\{ \left\lceil \frac{N}{r-1} \right\rceil, \left\lceil \frac{N}{r-1} \right\rceil - 1 \right\}, \quad k = 2, \cdots, r.$$

And let $m = \lceil \frac{N}{r-1} \rceil$ in (21), we obtain the maximum MSE under the optimal placement. □

C.2 Proof of Theorem 2

Both type-I and type-II blocks are presented in a packet in this case. Denote m_2 the maximum type-II block length. W.o.l.g., let $m_2 = \nu_{r+1} \geq \nu_1$, and $q_2 = \nu_{r+1} - \nu_1$. Given m_2 and q_2, if we only consider the optimization problem for the rest of data blocks, we reduce the to the constraint problem in Theorem 1. And

$$m_1 = \left\lceil \frac{N - 2m_2 - q_2}{r-1} \right\rceil \quad \text{(A· 21)}$$

is the maximum type-I block length under the optimal placement in that case.

To utilize this result, we first restrict m_2 such that

$$\begin{cases} \mathcal{E}_1^{(m_1)} \geq \mathcal{E}_2^{(m_2)} \\ (r-1)m_1 - q_1 + 2m_2 - q_2 = N \end{cases} \quad \text{(A· 22)}$$

where $q_1 = \sum_{k=2}^{r-1} i_k(\mathcal{P})$ as in (A· 14), and $q_1 \in \{0, \cdots, r-2\}$. The second constraint is by $\sum_{k=1}^{r+1} \nu_k = N$.

Under the above constraints, given an m_2, $\mathcal{E}_1^{(m_1)} = \min \mathcal{E}(\mathcal{P})$. Recall that $\mathcal{E}_1^{(m_1)}$ monotone increases with m_1, thus

$$q_{2\#} = \arg\min_{q_2} \mathcal{E}_1^{(m_1)} = \arg\min_{q_2} (N - 2m_2 + q_2) = 0.$$

Then, given m_2, $m_1 = \lceil \frac{N-2m_2}{r-1} \rceil$. Since $\mathcal{E}_2^{(m_2)}$ also monotone increases with m_2, to minimize $\mathcal{E}_1^{(m_1)}$,

$$m_{2\#} = \arg\min_{0 \leq m_2 \leq \lfloor \frac{N-(r-1)}{2} \rfloor} \mathcal{E}_1^{(m_1)}, \text{ subject to } \mathcal{E}_1^{(m_1)} - \mathcal{E}_2^{(m_2)} \geq 0$$
$$= \arg\min_{0 \leq m_2 \leq \lfloor \frac{N-(r-1)}{2} \rfloor} \mathcal{E}_1^{(m_1)} - \mathcal{E}_2^{(m_2)}, \text{ subject to } \mathcal{E}_1^{(m_1)} - \mathcal{E}_2^{(m_2)} \geq 0 \quad \text{(A· 23)}$$

Denote the above placement we obtained as $\mathcal{P}_\#$. Now, we move one data symbol to the last data block, i.e., increase ν_{r+1} by 1. Denote the new placement $\mathcal{P}'_\#$, where $m'_{2\#} = m_{2\#} + 1$. Notice that according to the above minimization, the following is true

$$\mathcal{E}_2^{(m'_{2\#})} > \mathcal{E}_1^{(m'_{1\#})} \quad \text{(A· 24)}$$

where $m'_{1\#} = \left\lceil \frac{N-2m_{2\#}-1}{r-1} \right\rceil$. This implies $\mathcal{E}(\mathcal{P}'_\#) = \mathcal{E}_2^{(m_{2\#'})}$.

- If $\mathcal{E}(\mathcal{P}_\#) > \mathcal{E}(\mathcal{P}'_\#)$, i.e., $\mathcal{E}_1^{(m_1^\#)} > \mathcal{E}_2^{(m'_{2\#})}$, then

$$\min_{\mathcal{P}} \mathcal{E}(\mathcal{P}) = \mathcal{E}_2^{(m'_{2\#})}. \quad \text{(A· 25)}$$

And

$$\min_{\mathcal{P}'_\#} \left| \mathbf{I}_{\mathcal{P}'_\#} \right| = 1, \quad \text{(A· 26)}$$

where $\nu_{r+1} = m_{2\#} + 1$, $\nu_1 = m_{2\#}$, and $\nu_i \in \{m'_{1\#}, m'_{1\#} - 1\}$, $i = 2, \cdots, r-1$.
Notice that, in this case, by (A· 24), we have $\mathcal{E}_1^{(m_{1\#})} > \mathcal{E}_1^{(m'_{1\#})}$, and we conclude

$$\begin{cases} v^* = m'_{1\#} = m_1^\# - 1 \\ q^* = q'_{1\#} = 0 \end{cases} \quad \text{(A· 27)}$$

- If $\mathcal{E}(\mathcal{P}_\#) \leq \mathcal{E}(\mathcal{P}'_\#)$, i.e., $\mathcal{E}_1^{(m_{1\#})} \leq \mathcal{E}_2^{(m'_{2\#})}$, then

$$\min_{\mathcal{P}} \mathcal{E}(\mathcal{P}) = \mathcal{E}_1^{(m_{1\#})}. \quad \text{(A· 28)}$$

Therefore,

$$\min_{\mathcal{P}_\#} \left| \mathbf{I}_{\mathcal{P}_\#} \right| = q_{1\#} = \mod \left(\frac{N - 2m_{2\#}}{r-1} \right), \quad \text{(A· 29)}$$

where $\nu_{r+1} = \nu_1 = m_{2\#}$, $\nu_i \in \{m_{1\#}, m_{1\#} - 1\}$, $i = 2, \cdots, r-1$ □

Appendix D: Proof of (31)

Notice that

$$\sum_{k,l:(k-l)\notin\mathcal{P}} \mathrm{E}\left\{ \left| \tilde{h}_k[l] \right|^2 \right\} = \sum_{l,i} \mathcal{M}(\hat{\mathbf{h}}_d^{(l)})_{ii}, \quad \text{(A· 30)}$$

where $\mathcal{M}(\hat{\mathbf{h}}_d)^{(l)}$ is given in (A· 12). Therefore, to show (31), it is equivalently to show

$$\sum_i \mathcal{M}(\hat{\mathbf{h}}_d^{(k)})_{ii} = \sum_i \mathcal{M}(\hat{\mathbf{h}}_d^{(l)})_{ii}, \tag{A· 31}$$

for $0 \le k, l \le L$, $k \ne l$.

Under the structure of the pilot clusters, it is easy to see that the time duration between $h_t^{(k)}[i]$ and $h_t^{(k)}[j]$ is equivalent to that between $h_t^{(l)}[i]$ and $h_t^{(l)}[j]$. Since the fading of each channel tap is a stationary process, and all taps are identically distributed, this gives

$$\mathbf{R}_{\mathbf{h}_t^{(k)}} = \mathbf{R}_{\mathbf{h}_t^{(l)}}. \tag{A· 32}$$

Define

$$\mathbf{D}^{(l)} \triangleq \mathbf{S}_t^H(\mathbf{S}_t\mathbf{R}_{\mathbf{h}_t^{(l)}}\mathbf{S}_t^H + \sigma_n^2\mathbf{I})^{-1}\mathbf{S}_t. \tag{A· 33}$$

Then,

$$\mathbf{D}^{(l)} = \mathbf{D}^{(k)}, \quad l \ne k. \tag{A· 34}$$

Similarly, under the assumption of packet starts and ends with pilot clusters of length $(2L+1)$, the correlation between $\mathbf{h}_d^{(l)}$ and $\mathbf{h}_t^{(l)}$ is the same for any l. Thus, we have

$$\mathbf{R}_{\mathbf{h}_d^{(l)}} = \mathbf{R}_{\mathbf{h}_d^{(k)}}, \quad \text{and} \quad \mathbf{R}_{\mathbf{h}_{dt}^{(l)}} = \mathbf{R}_{\mathbf{h}_{dt}^{(k)}}. \tag{A· 35}$$

Thus, from (31)

$$\mathcal{M}(\hat{\mathbf{h}}_d^{(l)}) = \mathcal{M}(\hat{\mathbf{h}}_d^{(0)}), \tag{A· 36}$$

and (A· 31) follows. □

Appendix E: Proof of Theorem 3

To proof the theorem, it is enough to show (48). The function $g(n)$ is defined as

$$g(n) = \sum_{i=0}^{n-1} \phi(\beta - f(i) - f(n-1-i)), \tag{A· 37}$$

where $\phi(x) = \frac{1}{x}$. Now the function $\phi(x)$ is a continuous decreasing convex function in the region $(0, \infty)$. If $\mathbf{x} = (x_0, \cdots, x_{(n-1)})$ and $\mathbf{y} = (y_0, \cdots, y_{(n-1)})$, we have [2] (page 10)

$$\mathbf{x} \prec^w \mathbf{y} \Rightarrow \sum_{i=0}^{n-1} \phi(x_i) \le \sum_{i=0}^{n-1} \phi(y_i). \tag{A· 38}$$

Here $\mathbf{x} \prec^w \mathbf{y}$ if

$$\sum_{i=0}^{k} x_{(i)} \ge \sum_{i=0}^{k} y_{(i)}, \quad k = 0, \cdots, n-1, \tag{A· 39}$$

where $x_{(0)} \le x_{(2)} \le \cdots \le x_{(n-1)}$ denote the components of $\mathbf{x}$ in increasing order.

Our aim is to show that

$$g(n) + g(n) \le g(n-1) + g(n+1), \tag{A· 40}$$

that is

$$\begin{aligned}&\sum_{i=0}^{n-1} \phi(\beta - f(i) - f(n-1-i)) \\ &\quad + \sum_{i=0}^{n-1} \phi(\beta - f(i) - f(n-1-i)) \\ &\quad \le \sum_{i=0}^{n-2} \phi(\beta - f(i) - f(n-2-i)) \\ &\quad + \sum_{i=0}^{n} \phi(\beta - f(i) - f(n-i)).\end{aligned} \tag{A· 41}$$

We claim that $(\beta - f(i) - f(n-1-i), \beta - f(i-1) - f(n-i)) \prec^w (\beta - f(i-1) - f(n-1-i), \beta - f(i) - f(n-i))$ for $i = 1, \cdots, (n-1)$. This combined with the fact that

$$\phi(\beta - f(0) - f(n-1)) \le \phi(\beta - f(0) - f(n)), \tag{A· 42}$$

proves the required result. □

Appendix F: Derivation of Eq. (52)

When $L = 0$, within a packet, we have

$$\mathbf{y} = (\rho_d\mathbf{S}_d + \rho_t\mathbf{S}_t)\mathbf{h} + \mathbf{n}.$$

similar as in (A· 2), we have

$$\mathcal{M}(\hat{\mathbf{h}}) = \mathbf{R_h} - \mathbf{R_{hy}}\mathbf{R_y}^{-1}\mathbf{R_{hy}}^H \tag{A· 43}$$

where

$$\begin{aligned}\mathbf{R_{hy}} &= E\{\mathbf{h}\mathbf{y}^H\} = \rho_t\mathbf{R_h}\mathbf{S}_t^H \\ \mathbf{R_y} &= \rho_d^2 E\{(\mathbf{S}_d\mathbf{h})(\mathbf{h}^H\mathbf{S}_d^H)\} + \rho_t^2\mathbf{S}_t\mathbf{R}_h\mathbf{S}_t^H + \sigma_n^2\mathbf{I} \\ &= \rho_t^2\mathbf{S}_t\mathbf{R}_h\mathbf{S}_t^H + (\rho_d^2\sigma_h^2\mathbf{I} + \sigma_n^2\mathbf{I}).\end{aligned} \tag{A· 44}$$

Therefore, using matrix inversion lemma, we have Eq. (52).

Min Dong received the B.Eng. degree from 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, New York. Her research interests include statistical signal processing, wireless communications and communication networks.

Srihari Adireddy was born in India in 1977. He received the B.Tech degree from the Department of Electrical Engineering, Indian Institute of Technology, Madras, India in 1998 and a special M.S from the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY in 2001. Currently, he is working towards his PhD degree at Cornell University, Ithaca. His research interests include signal processing, information theory and random access protocols.

Lang Tong received the B.E. degree from Tsinghua University, Beijing, China, in 1985, and M.S. and Ph.D. degrees in electrical engineering in 1987 and 1990, respectively, from the University of Notre Dame, Notre Dame, Indiana. He was a Postdoctoral Research Affiliate at the Information Systems Laboratory, Stanford University in 1991. Currently, he is an Associate Professor in the School of Electrical and Computer Engineering, Cornell University, Ithaca, New York. Dr. Tong received *Young Investigator Award* from the Office of Naval Research in 1996, and the *Outstanding Young Author Award* from the IEEE Circuits and Systems Society. His areas of interest include statistical signal processing, adaptive receiver design for communication systems, signal processing for communication networks, and information theory.