Optimal Insertion of Pilot Symbols for Transmissions Over Time-Varying Flat Fading Channels

Min Dong, *Student Member, IEEE*, Lang Tong, *Senior Member, IEEE*, and Brian M. Sadler, *Member, IEEE*

***Abstract*—Two major training techniques for wireless channels are time-division multiplexed (TDM) training and superimposed training. For the TDM schemes with regular periodic placements (RPPs), the closed-form expression for the steady-state minimum mean square error (MMSE) of the channel estimate is obtained as a function of placement for Gauss–Markov flat fading channels. We then show that among all periodic placements, the single pilot RPP scheme (RPP-1) minimizes the maximum steady-state channel MMSE. For binary phase-shift keying (BPSK) and quadrature phase-shift keying (QPSK) signaling, we further show that the optimal placement that minimizes the maximum uncoded bit error rate (BER) is also RPP-1. We next compare the MMSE and BER performance under the superimposed training scheme with those under the optimal TDM scheme. It is shown that while the RPP-1 scheme performs better at high SNR and for slowly varying channels, the superimposed scheme outperforms RPP-1 in the other regimes. This demonstrates the potential for using superimposed training in relatively fast time-varying environments.**

***Index Terms*—Channel tracking, Gauss–Markov, Kalman filter, pilot symbols, placement schemes, PSAM, superimposed, time-varying.**

I. Introduction

CHANNEL estimation is a major challenge for reliable wireless transmissions. Often, in practice, pilot symbols known to the receiver are multiplexed with data symbols for channel acquisition. Two major types of training for single carrier systems are time division multiplexed (TDM) training and superimposed training. Pilot symbols in a TDM system are inserted into the data stream according to a certain placement pattern, and the channel estimate is updated using these pilot symbols. For superimposed training, on the other hand, pilot and data symbols are added and transmitted together, and the channel estimate is updated at each symbol.

The way that pilot symbols are multiplexed into the data stream affects the system performance for time-varying channels. Under TDM training, the presence of pilot symbols makes channel estimation accurate at some periods of time and coarse at others. If the percentage of pilot symbols is fixed, we then have to choose between obtaining accurate estimations infrequently or frequent but less accurate estimates. Is it better to cluster pilot symbols as in the case of the Global System for Mobile Communications (GSM) systems or to spread pilot symbols evenly in the data stream as in the pilot symbol assisted modulation (PSAM) [1]? What is the optimal placement that minimizes the mean square error (MSE) of the channel estimator? Does the MSE-minimizing training also minimize the bit error rate (BER)? In choosing the optimal training scheme, do we need to know the rate of channel variation and the level of signal-to-noise ratio (SNR)? How does TDM training compare with superimposed training? Intuitively, superimposed training may have the advantage when the channel fades rapidly, but the superimposed data interferes with pilot-aided channel estimation, which may lead to an undesirable performance floor in the high SNR regime.

In this paper, we address these issues systematically. We model the time-varying flat fading channel by a Gauss–Markov process and use the minimum mean square error (MMSE) channel estimator along with the symbol-by-symbol maximum likelihood (ML) detector. The MMSE channel estimator is implemented using the Kalman filter. For TDM training we show that, among all periodic placements, the regular periodic placement with pilot cluster size one (referred to as RPP-1) minimizes the maximum steady-state channel MMSE and uncoded BER for both binary phase-shift keying (BPSK) and quaternary phase-shift keying (QPSK) signaling, regardless of the SNR level or the rate of channel variation. Given the constraint of the minimum length of pilot clusters γ, we show that RPP-γ is optimal. Performance comparisons between the optimal TDM scheme and the superimposed scheme are given both analytically and numerically. We show that the optimal TDM scheme performs better at high SNR and for slowly varying channels, whereas the superimposed scheme is superior for many situations of practical importance. In the process of establishing the optimality of RPP-1, we also provide the closed-form expression for steady-state channel MMSE at each data symbol position, which is useful to evaluate the performance of coded transmissions.

Pilot symbol-assisted modulation (PSAM), which has been proposed in [2] and [3], includes the periodic TDM training with cluster size one—the RPP-1 scheme. Cavers first analyzed the performance of PSAM [1]. Although the optimality of RPP-1 has never been shown for either channel MMSE or BER until now, it has been applied and studied in various settings [4]–[8]. The optimality of RPP-1 may not be surprising in retrospect, but the fact that the optimality holds uniformly across all fading

Manuscript received January 29, 2003; revised June 9, 2003. This work was supported in part by the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564 and the Army Research Laboratory CTA on Communication and Networks under Grant DAAD19-01-2-0011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Zhi Ding.

M. Dong and L. Tong are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: mdong@ece.cornell.edu; ltong@ece.cornell.edu).

B. M. Sadler is with the Army Research Laboratory, Adelphi, MD 20783 USA (e-mail: bsadler@arl.army.mil).

Digital Object Identifier 10.1109/TSP.2004.826182

and SNR levels may seem unexpected. Establishing the optimality formally and uniformly across a wide range of channel conditions, however, does not come from a direct application of the standard Kalman filtering theory. In particular, we need to examine all possible training patterns and their corresponding maximum steady-state MMSEs, which would not have been possible without characterizing the MSE behavior with respect to the placement pattern. Under TDM training, the channel estimator switches between the Kalman updates using pilot symbols and the Kalman predictions during data transmissions, and the switching occurs before the steady-state in either phase has been reached.

Optimal training has been previously considered for block-fading channels from a channel-estimation perspective under both TDM and superimposed trainings [9]–[11], as well as from an information theoretic viewpoint [12], [13]. For time-varying channels, existing results tend to assume the RPP-1 scheme and optimize parameters such as pilot symbol spacing and power and rate allocations [1], [6], [7]. In [6], for flat Rayleigh fading modeled by a Gauss–Markov process and the PSAM scheme, the optimal spacing between the pilot symbols is determined numerically by maximizing the mutual information with binary inputs. In [7], with the flat fading channel modeled by a bandlimited process, under some assumptions on the channel and data symbols, optimal parameters for pilots, including pilot symbol spacing and power allocation, are determined by maximizing a lower bound on capacity. In [14], the performance in various aspects of code division multiple access (CDMA) systems under two pilot-assisted schemes is analyzed. In [15], we addressed the problem of optimal placement of pilot symbols in TDM schemes for packetized transmission over time-varying channels at high SNR.

This paper is organized as follows. In Section II, we study the optimal pilot placement for TDM schemes. We first introduce the system model and formulate the problem and then obtain the optimal placement for both channel tracking and uncoded BER performance. We then consider superimposed training in Section III, where we derive the steady-state channel MSE with Kalman tracking and the BER. In Section IV, we provide both analytical and numerical performance comparison under the optimal TDM scheme and the superimposed scheme. Finally, we conclude in Section V.

II. Optimal Placement for TDM Training

A. Channel Model

We model a time-varying flat Rayleigh fading channel as

$$y_k = s_k h_k + w_k, \quad k = 1, 2, \ldots \tag{1}$$

where y_k is the received observation, s_k the transmitted symbol, $h_k \sim \mathcal{CN}(0, \sigma_h^2)$ the zero mean complex Gaussian channel state with variance σ_h^2, and $w_k \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0, \sigma_w^2)$ the complex circular additive white Gaussian noise (AWGN) at time k. We assume that data s_k, channel h_k, and noise w_k are jointly independent.

The dynamics of the channel state h_k are modeled by a first-order Gauss–Markov process

$$h_k = a h_{k-1} + u_k, \quad u_k \overset{\text{i.i.d.}}{\sim} \mathcal{CN}\left(0, (1-a^2)\sigma_h^2\right) \tag{2}$$

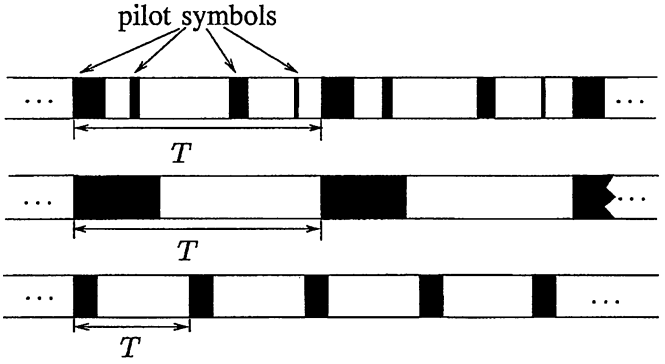


Fig. 1. Data streams with periodic placements.

where u_k is the white Gaussian driving noise. Parameter $a \in [0, 1]$ is the fading correlation coefficient that characterizes the degree of time variation; small a models fast fading, and large a corresponds to slow fading. The Gauss–Markov model is widely adopted as a simple and effective model to characterize the fading process [16]–[19]. The first-order Gauss–Markov model is parameterized by the fading correlation coefficient a. The value of a can be determined by the channel Doppler spread and the transmission bandwidth, where the relation among the three is found in [19]. It can be accurately obtained at the receiver for a variety of channels [6], [17], [18]. Here, we assume that a is known.

B. Periodic TDM Placements

We consider the class of periodic placements, as shown in Fig. 1, where the placement pattern of pilot symbols repeats periodically in the data stream. The restriction to periodic placements is mild; a system with aperiodic training will not reach a steady state and is seldom considered in practice. We define the period of a placement, denoted by T, to be the length of the smallest block over which the placement pattern repeats. Note that the starting point of a period can be arbitrarily chosen. Without loss of generality, we assume that each period starts with a pilot symbol and ends with a data symbol.

In general, any periodic placement with n clusters of pilot symbols in a period of length T can be specified by a 2-tuple $\boldsymbol{\mathcal{P}} = (\boldsymbol{\gamma}, \boldsymbol{\nu})$, where $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_n]$ is the pilot cluster length vector and $\boldsymbol{\nu} = [\nu_1, \ldots, \nu_n]$ the data block length vector, as illustrated in Fig. 2. Note that $T = \sum_{i=1}^{n} (\gamma_i + \nu_i)$. We further denote $\mathcal{I}_p(\boldsymbol{\mathcal{P}})$ as the index set containing positions (relative to the beginning of the period) of the pilot symbols within one period.

For different placement schemes, we also assume the following.

- A1) All pilot symbols have equal power and are denoted by σ_p^2; the power for data symbols is denoted by σ_d^2.
- A2) The percentage of pilot symbols in a data stream $\eta = \sum_{i=1}^{n} \gamma_i / T$ is fixed.

C. Receiver

We consider a typical receiver structure shown in Fig. 3, where the channel estimator provides the channel estimate $\hat{h}_k$ to the demodulator, and the data symbol s_k is detected based

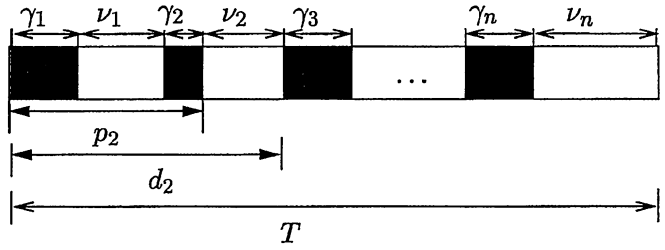


Fig. 2. Representation of placement within one placement period.

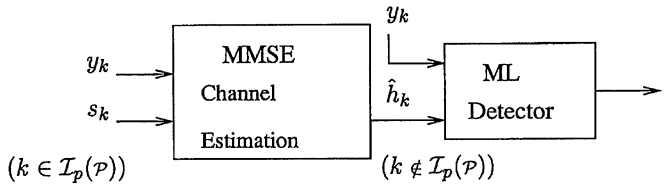


Fig. 3. Receiver structure.

on the received sample y_k and $\hat{h}_k$ using the symbol-by-symbol ML detector.[1]

For a given placement $\boldsymbol{\mathcal{P}}$, the observations over pilot symbols are given by $\{y_{lT+k} : k \in \mathcal{I}_p(\boldsymbol{\mathcal{P}}), l = 0, 1, \ldots\}$. We consider the MMSE channel estimator based on the current and all past pilot symbols and their corresponding observations. The MMSE channel estimator at time $(lT+k)$, which is denoted by $\hat{h}_{lT+k}$, is then given by

$$\hat{h}_{lT+k}(\boldsymbol{\mathcal{P}}) = \mathrm{E}\{h_{lT+k} \,|\, \{y_{lT+j} : j \le k, j \in \mathcal{I}_p(\boldsymbol{\mathcal{P}})\}, \{y_{(l-m)T+j} : j \in \mathcal{I}_p(\boldsymbol{\mathcal{P}}), \quad m = 1, 2, \ldots\}\}$$

which, for a channel with Gaussian statistics, is equivalent to the linear MMSE (LMMSE) estimator. The direct use of the MMSE estimator requires information storage on all the past pilot symbol observations at the receiver, as well as computing a matrix inversion at each time to estimate the channel. As the number of observations grows, this excessively burdens the receiver with both storage and computation. In practice, an adaptive filter is desired, especially for time-varying channels. The above MMSE estimator can be implemented recursively by the Kalman filter. At each time, one only needs to store the most recent channel estimate and estimates the channel with simple scalar operations. The Kalman filter has been widely used due to its optimality and adaptivity. For the transmission with TDM pilot placements we consider in Fig. 1, the Kalman filter switches between two modes: It updates the channel estimate using the pilot symbols during the training period and predicts the channel state during data transmission.

Given the estimated channel state $\hat{h}_k$, the optimal detection for equally probable data symbol alphabets is given by the ML detector. For a fixed placement $\boldsymbol{\mathcal{P}}$, let

$$M[k; \boldsymbol{\mathcal{P}}] \triangleq \mathrm{E}\{|\hat{h}_k(\boldsymbol{\mathcal{P}}) - h_k(\boldsymbol{\mathcal{P}})|^2\}$$

be the MMSE of the channel estimate at time k produced by the Kalman filter. Conditioned on any data symbol s_k, y_k, and $\hat{h}_k(\boldsymbol{\mathcal{P}})$ are jointly Gaussian with zero mean and covariance

$$\boldsymbol{\Sigma}_k = \begin{bmatrix} \sigma_w^2 + |s_k|^2 \sigma_h^2 & s_k\left(\sigma_h^2 - M[k; \boldsymbol{\mathcal{P}}]\right) \\ s_k^*\left(\sigma_h^2 - M[k; \boldsymbol{\mathcal{P}}]\right) & \sigma_h^2 - M[k; \boldsymbol{\mathcal{P}}] \end{bmatrix}.$$

For any phase-shift keying (PSK) constellation, we have $|s_k|^2 = \sigma_d^2$. The ML decision rule is given by

$$\begin{aligned} \hat{s}_k &= \arg\max_{s_k} p(y_k, \hat{h}_k \,|\, s_k) \\ &= \arg\min_{s_k} [y_k^* \; \hat{h}_k^*(\boldsymbol{\mathcal{P}})] \boldsymbol{\Sigma}_k^{-1} [y_k \; \hat{h}_k(\boldsymbol{\mathcal{P}})]^T \\ &= \arg\max_{s_k} \mathrm{Re}\{s_k^* \hat{h}_k^*(\boldsymbol{\mathcal{P}}) y_k\} \\ &= \arg\min_{s_k} |y_k - \hat{h}_k(\boldsymbol{\mathcal{P}}) s_k|^2 \end{aligned} \tag{3}$$

which shows that the same ML detector for the known channel can be used by substituting the channel estimate.

D. Optimization Criteria

The MSE and BER performance of TDM schemes are not stationary. During a training block, the Kalman filter uses pilot symbols to produce increasingly more accurate channel estimates until the data transmission starts, during which time, the Kalman filter can only predict the channel state based on the Gauss–Markov model and, therefore, produces increasingly inaccurate estimates.

The Kalman filter update algorithm can be obtained from the standard Kalman filter theory [20], [21] and is detailed in Appendix A. Here, we only present the channel MMSE update needed for analysis.

During a training block, we obtain the recursive expressions for the channel MMSE as, for $k \in \mathcal{I}_p(\boldsymbol{\mathcal{P}})$ and all integer l

$$M[lT+k; \boldsymbol{\mathcal{P}}] = \frac{\sigma_w^2\left(a^2 M[lT+k-1; \boldsymbol{\mathcal{P}}] + (1-a^2)\sigma_h^2\right)}{\sigma_w^2 + \left(a^2 M[lT+k-1; \boldsymbol{\mathcal{P}}] + (1-a^2)\sigma_h^2\right)\sigma_p^2}. \tag{4}$$

Once the ith training cluster in a placement period ends, of which the index (relative to the beginning of the period) is denoted by p_i, as shown in Fig. 2, the Kalman filter predicts the channel state during data transmissions of duration ν_i. This MMSE is given by

$$M[lT + p_i + k; \boldsymbol{\mathcal{P}}] = a^{2k} M[lT + p_i; \boldsymbol{\mathcal{P}}] + \sigma_h^2(1 - a^{2k}) \tag{5}$$

for $k = 1, \ldots, \nu_i$.

As $l \to \infty$, the system converges to a periodic steady state, and we are naturally interested in the steady-state performance

$$M_k(\boldsymbol{\mathcal{P}}) \triangleq \lim_{l \to \infty} M[lT + k; \boldsymbol{\mathcal{P}}], \quad k = 1, \ldots, T. \tag{6}$$

[1]Note that the globally optimum detector is the ML sequence detector.

Furthermore, we will only be interested in the MSE of the channel estimator during data transmission. Thus, from (5) and (6)

$$M_{p_i+k}(\boldsymbol{\mathcal{P}}) = a^{2k} M_{p_i}(\boldsymbol{\mathcal{P}}) + (1-a^{2k})\sigma_h^2, \quad k = 1, \ldots, \nu_i \quad (7)$$

which monotonically increases with k.

In a placement period, let $d_i \triangleq p_i + \nu_i$ be the index of the position of the last symbol in the ith data block, as shown in Fig. 2. Then, the maximum steady-state MMSE in this block is reached at d_i. The maximum steady-state channel MMSE over data symbols is then given by

$$\mathcal{E}(\boldsymbol{\mathcal{P}}) \triangleq \max_{k:k\notin\mathcal{I}_p(\boldsymbol{\mathcal{P}})} M_k(\boldsymbol{\mathcal{P}}) = \max_{1\le i\le n} M_{d_i}(\boldsymbol{\mathcal{P}}). \quad (8)$$

The optimal placement $\boldsymbol{\mathcal{P}}^*_{\mathrm{MMSE}}$ that minimizes the maximum steady-state channel MMSE can then be obtained by[2]

$$\begin{aligned}\boldsymbol{\mathcal{P}}^*_{\mathrm{MMSE}} &= \arg\min_{\boldsymbol{\mathcal{P}}} \mathcal{E}(\boldsymbol{\mathcal{P}}) \\ &= \arg\min_{\boldsymbol{\mathcal{P}}} \max_{1\le i\le n} M_{d_i}(\boldsymbol{\mathcal{P}}). \end{aligned} \quad (9)$$

The BER performance is directly affected by channel MMSE, and our goal is to find the optimal placement that minimizes the maximum steady-state BER. Specifically, let $P_e(k;\boldsymbol{\mathcal{P}})$ be the steady-state BER at the kth position relative to the beginning of the placement period. We are interested in the following optimization:

$$\boldsymbol{\mathcal{P}}^*_{\mathrm{BER}} = \arg\min_{\boldsymbol{\mathcal{P}}} \max_{k:k\notin\mathcal{I}_p(\boldsymbol{\mathcal{P}})} P_e(k;\boldsymbol{\mathcal{P}}). \quad (10)$$

We show next that $\boldsymbol{\mathcal{P}}^*_{\mathrm{MMSE}} = \boldsymbol{\mathcal{P}}^*_{\mathrm{BER}}$ for the BPSK and QPSK constellations.

Proposition 1: Under the Gauss–Markov channel model with BPSK or QPSK input symbols, if the MMSE channel estimator is used along with the ML detector, then

$$\boldsymbol{\mathcal{P}}^*_{\mathrm{MMSE}} = \boldsymbol{\mathcal{P}}^*_{\mathrm{BER}}.$$

Proof: See Appendix B.

E. Optimal TDM Placement

We first find the optimal placement for a special class of placements called *regular periodic placements*. The extension to the general class follows.

The *regular periodic placement* RPP-γ has only one pilot cluster of size γ and one data cluster of size ν, with $T = \gamma/\eta$. In Fig. 1, the second and third examples are placements belonging to this class.

From (8) and (9), it follows that for RPP-γ, the optimal placement is obtained by

$$\gamma_* = \arg\min_{\gamma} \mathcal{E}(\gamma) = \arg\min_{\gamma} M_T(\gamma). \quad (11)$$

[2]We assume there are n pilot clusters for a given placement $\boldsymbol{\mathcal{P}}$.

Our problem now is to find the explicit expression of the steady-state MMSE $M_T(\gamma)$ and analyze its behavior as a function of pilot cluster size γ.

A useful quantity in the sequel is the steady-state MMSE when all symbols are pilots. It is the solution to the steady-state Riccati equation

$$\begin{aligned} M_\infty &\triangleq \lim_{\gamma\to\infty} M_\gamma(\gamma) = \frac{\sigma_w^2\left(a^2 M_\infty + \sigma_u^2\right)}{\sigma_w^2 + \left(a^2 M_\infty + \sigma_u^2\right)\sigma_p^2} \\ &= \frac{\sigma_h^2}{\frac{1}{2}\left(1+\sigma_h^2\mathrm{SNR}_p\right) + \sqrt{\left[\frac{1}{2}\left(1+\sigma_h^2\mathrm{SNR}_p\right)\right]^2 + \frac{a^2}{1-a^2}\sigma_h^2\mathrm{SNR}_p}} \end{aligned} \quad (12)$$

where the signal-to-noise ratio $\mathrm{SNR}_p \triangleq \sigma_p^2/\sigma_w^2$. Obviously, M_∞ is a lower bound on MMSE for any placement.

The following Lemma provides the closed-form MMSE expression for the RPP-γ scheme.

Lemma 1: For any RPP-γ scheme, the steady-state channel MMSE is given by

$$M_T(\gamma) = \delta_T(\gamma) + M_\infty \quad (13)$$

$$M_k(\gamma) = \frac{M_T(\gamma) - \sigma_h^2(1-a^{2(T-k)})}{a^{2(T-k)}} \qquad k = \gamma+1, \ldots, T-1 \quad (14)$$

where $\delta_T(\gamma)$ is computed as follows:

$$\delta_T(\gamma) = -b_\gamma + \sqrt{b_\gamma^2 + c_\gamma} \quad (15)$$

$$b_\gamma \triangleq \left(\frac{1-\left(\frac{a^{2\frac{1-\eta}{\eta}}}{\alpha}\right)^\gamma}{1-\frac{1}{\alpha^\gamma}}\right)\left(\frac{\alpha-1}{2\beta}\right) - \frac{\left(1-a^{2\frac{1-\eta}{\eta}\gamma}\right)}{2}\rho_\infty\sigma_h^2 \quad (16)$$

$$c_\gamma \triangleq \left(\frac{1-a^{2\frac{1-\eta}{\eta}\gamma}}{1-\frac{1}{\alpha^\gamma}}\right)\left(\frac{\alpha-1}{\beta}\right)\rho_\infty\sigma_h^2 \quad (17)$$

$$\alpha \triangleq \frac{1}{a^2}\left(1+(1-a^2\rho_\infty)\sigma_h^2\mathrm{SNR}_p\right)^2$$

$$\beta \triangleq \mathrm{SNR}_p\left(1+(1-a^2\rho_\infty)\sigma_h^2\mathrm{SNR}_p\right) \quad (18)$$

$$\rho_\infty \triangleq 1 - \frac{M_\infty}{\sigma_h^2}. \quad (19)$$

Proof: See Appendix C.

From (13), because M_∞ is not a function of γ, the optimization in (11) can now be rewritten as

$$\gamma_* = \arg\min_{\gamma} \delta_T(\gamma)$$

where the explicit expression of $\delta_T(\gamma)$ is obtained in (15) as a function of γ. By analyzing the behavior of $\delta_T(\gamma)$ as a function of γ, we obtain the optimal placement for RPP schemes in the following theorem.

Theorem 1: For the class of RPP schemes, under A1) and A2), the maximum MMSE $\mathcal{E}(\gamma)$ of the channel estimates during

data transmission is a monotone increasing function of γ. Thus, RPP-1 is optimal among all RPP placements, and the minimum $\mathcal{E}(\gamma)$ is given by

$$\begin{aligned}\mathcal{E}^{\mathrm{TDM}} &\triangleq \min_{\gamma} \mathcal{E}(\gamma) = \mathcal{E}(1) \\ &= \sigma_h^2 - a^{2\frac{1-\eta}{\eta}} \left(\sigma_h^2 - M_1\right) \end{aligned} \tag{20}$$

where

$$M_1 = \frac{\sigma_h^2}{\frac{1}{2}(1+\sigma_h^2 \mathrm{SNR}_p) + \sqrt{\left[\frac{1}{2}(1+\sigma_h^2 \mathrm{SNR}_p)\right]^2 + \frac{a^{\frac{2}{\eta}}}{1-a^{\frac{2}{\eta}}}\sigma_h^2 \mathrm{SNR}_p}}. \tag{21}$$

Proof: See Appendix D.

Theorem 1 demonstrates that decreasing the training cluster length and training the channel more frequently results in decreased steady-state maximum channel MMSE and, thus, lower BER. This immediately implies that if there is a constraint on the minimum pilot cluster size γ_o, RPP-γ_o is optimal.

We next show that RPP-1 is in fact optimal among all periodic placements. We outline a few steps required to prove the optimality of RPP-1, leaving the details to Appendix E. Consider first the case with two pilot clusters of lengths $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$ and two data blocks of lengths $\boldsymbol{\nu} = (\nu_1, \nu_2)$ present in each period. Let $d_1 = \gamma_1 + \nu_1$ and $d_2 = d_1 + \gamma_2 + \nu_2$ be the end positions of the two data blocks, where the MMSE $M_{d_i}(\boldsymbol{\mathcal{P}})$ reaches the maximum. Intuition suggests that moving the second pilot cluster away from the first, i.e., increasing ν_1 and decreasing ν_2, increases $M_{d_1}(\boldsymbol{\mathcal{P}})$ and decreases $M_{d_2}(\boldsymbol{\mathcal{P}})$. (This is not obvious, however, because moving the second pilot cluster will also affect the initial MMSE $M_1(\boldsymbol{\mathcal{P}})$.) It follows that to minimize the maximum MMSE for the entire period suggests that the equalization rule that forces $M_{d_1}(\boldsymbol{\mathcal{P}}) = M_{d_2}(\boldsymbol{\mathcal{P}})$, which leads to making pilot clusters equal and eventually results in the reduction to the RPP-γ scheme. Extending to any placement with n pilot clusters in a period, using the similar equalization rule and applying the above result to each two consecutive pilot clusters repeatedly, leads to the same reduction to RPP-γ. Combining Theorem 1 and Proposition 1, we then have the optimality of RPP-1.

Theorem 2: Given a fixed percentage of pilot symbols η, the optimal placement for periodic TDM training that minimizes the maximum steady-state MMSE and BER for any first order Gauss–Markov channel is RPP-γ_o, where γ_o is the minimum pilot cluster size allowed.

Proof: See Appendix E.

We point out that this optimality holds, regardless of the values of SNR_p and a.

III. Superimposed Training Scheme

The pilot design for superimposed training takes the form of allocating power to pilot and data symbols at each time index. For the stationary Gauss–Markov channel considered here, it is reasonable to consider the time-invariant power allocation where the transmitted symbol $s_k = \rho_t t_k + \rho_d d_k$ is the superposition of pilot and data symbols. The observation is given by

$$y_k = (\rho_t t_k + \rho_d d_k) h_k + w_k \tag{22}$$

where $\{t_k\}$ is the pilot sequence, and $\{d_k\}$ is the data sequence that is drawn from an independent and identically distributed (i.i.d.) zero mean sequence. We assume that t_k and d_k have unit powers, i.e., $\mathrm{E}\{t_k^2\} = \mathrm{E}\{d_k^2\} = 1$, and we denote ρ_t and ρ_d as the pilot and data power allocation coefficients, respectively.

A. Kalman Tracking with Superimposed Training

A complication of superimposed training is that h_k and $(y_k, y_{k-1}, \ldots)$ are not jointly Gaussian. Therefore, the MMSE channel estimator based on the conditional expectation is difficult to compute and implement. Instead, we consider the LMMSE channel estimator implemented via the Kalman filter. Rewrite (22) as

$$y_k = \rho_t t_k h_k + v_k$$

where $v_k \triangleq \rho_d d_k h_k + w_k$. Because $\{d_k\}$ is i.i.d. zero mean and independent of h_k, we have

$$\mathrm{E}\{v_i v_j^*\} = \left(\rho_d^2 \sigma_h^2 + \sigma_w^2\right) \delta_{ij}, \quad \mathrm{E}\{h_i v_j^*\} = 0 \quad \forall i, j.$$

Let $\hat{h}_k$ be the LMMSE estimator of h_k based on the current and all past observations $\{y_k, y_{k-1}, \ldots\}$; then, the Kalman filter can again be used as the optimal LMMSE estimator to track the channel. Let $M[k] \triangleq \mathrm{E}\{|h_k - \hat{h}_k|^2\}$. Following the Kalman filter derivation [20], [21], we then have the Kalman filtering algorithm for channel tracking under superimposed training with

$$\begin{aligned} K[k] &\triangleq \frac{\mathrm{E}\{h_k (y_k - a\rho_t t_k \hat{h}_{k-1})^*\}}{\mathrm{E}\{|y_k - a\rho_t t_k \hat{h}_{k-1}|^2\}} \\ &= \frac{\left(a^2 M[k-1] + (1-a^2)\sigma_h^2\right)\rho_t t_k}{\rho_t^2\left(a^2 M[k-1] + (1-a^2)\sigma_h^2\right) + \rho_d^2 \sigma_h^2 + \sigma_w^2} \end{aligned} \tag{23}$$

$$\hat{h}_k = a\hat{h}_{k-1} + K[k](y_k - a\rho_t t_k \hat{h}_{k-1})$$

$$M[k] = (1 - K[k]\rho_t t_k)\left(a^2 M[k-1] + (1-a^2)\sigma_h^2\right). \tag{24}$$

Combining (23) and (24), we have

$$M[k] = \frac{\left(\frac{\sigma_w^2}{\sigma_h^2 \rho_t^2} + \frac{\rho_d^2}{\rho_t^2}\right)\left(\frac{a^2}{1-a^2} M[k-1] + \sigma_h^2\right)}{\frac{1}{1-a^2}\left(\frac{\sigma_w^2}{\sigma_h^2 \rho_t^2} + \frac{\rho_d^2}{\rho_t^2}\right) + \left(\frac{a^2}{(1-a^2)\sigma_h^2} M[k-1] + 1\right)}.$$

The steady-state MSE, which is defined as $\mathcal{E}^{\mathrm{sup}} \triangleq \lim_{k\to\infty} M[k]$, is then given by

$$\mathcal{E}^{\mathrm{sup}} = \frac{\left(\frac{a^2}{1-a^2}\mathcal{E}^{\mathrm{sup}} + \sigma_h^2\right)}{\frac{1}{1-a^2} + \left(\frac{a^2}{(1-a^2)\sigma_h^2}\mathcal{E}^{\mathrm{sup}} + 1\right)\kappa}$$

where

$$\kappa \triangleq \frac{\sigma_h^2 \rho_t^2}{\sigma_h^2 \rho_d^2 + \sigma_w^2} \tag{25}$$

is the received signal-to-interference plus noise ratio (SINR).

The solution of the above equation is

$$\mathcal{E}^{\mathrm{sup}} = \frac{\sigma_h^2}{\frac{1}{2}(1+\kappa) + \sqrt{\left[\frac{1}{2}(1+\kappa)\right]^2 + \frac{a^2}{1-a^2}\kappa}}. \tag{26}$$

Note that at the steady-state, in contrast to the periodic placement scheme, the channel MSE in this case is time-invariant.

B. BER Performance

We again consider BPSK signaling. The detector estimates d_k based on $\hat{h}_k$ and y_k by

$$\hat{d}_k = \text{sign}\{\text{Re}\{\hat{h}_k^*(y_k - \rho_t t_k \hat{h}_k)\}\}. \tag{27}$$

Notice that this detector is not the true ML detector based on y_k and $\hat{h}_k$; it is a pseudo ML that assumes the estimated $\hat{h}_k$ has no error.

If $(d_k, d_{k-1}, \ldots)$ are transmitted, $\hat{h}_k$ and $(y_k - \rho_t t_k \hat{h}_k)$ are correlated, zero mean complex Gaussian random variables. Therefore, from the error probability calculation in [22, App. C], for a system using BPSK at the steady state, the bit error probability conditioned on d_k is

$$\begin{aligned}\Pr(\hat{d}_k \neq d_k \,|\, d_k) &= \mathrm{E}_{(d_{k-1},\ldots)}\{\Pr(\hat{d}_k \neq d_k \,|\, d_k, d_{k-1}, \ldots)\} \\ &= \frac{1}{2}\left[1 - \sqrt{\frac{\text{Re}^2\{\zeta[k]\}}{1 - \text{Im}^2\{\zeta[k]\}}}\right] \end{aligned} \tag{28}$$

where

$$\zeta[k] = \frac{\mathrm{E}\{\hat{h}_k^*(y_k - \rho_t t_k \hat{h}_k)\}}{\sqrt{\mathrm{E}|\hat{h}_k|^2 \mathrm{E}|y_k - \rho_t t_k \hat{h}_k|^2}}.$$

The terms in the denominator and the numerator are derived as follows.

$$\begin{aligned}&\mathrm{E}|\hat{h}_k|^2 = \sigma_h^2 - \mathcal{E}^{\text{sup}} \\ &\mathrm{E}|y_k - \rho_t t_k \hat{h}_k|^2 = \rho_d^2 \sigma_h^2 + \rho_t^2 \mathcal{E}^{\text{sup}} + \sigma_w^2 \\ &\qquad + 2\rho_t t_k \left(\rho_d d_k \mathcal{E}^{\text{sup}} - K[\infty]\sigma_w^2\right) \\ &\qquad = \left[\frac{1}{\kappa} + \left(2\frac{\rho_d^2}{\rho_t^2}\kappa + 2\frac{\rho_d}{\rho_t}\frac{d_k}{t_k} - 1\right)\frac{\mathcal{E}^{\text{sup}}}{\sigma_h^2}\right]\rho_t^2 \sigma_h^2 \\ &\mathrm{E}\{\hat{h}_k^*(y_k - \rho_t t_k \hat{h}_k)\} \\ &\qquad = \rho_d d_k \left(\sigma_h^2 - \mathcal{E}^{\text{sup}}\right) + K[\infty]\sigma_w^2 \\ &\qquad = \left[\frac{\rho_d}{\rho_t}\frac{d_k}{t_k}\left(1 - \frac{\mathcal{E}^{\text{sup}}}{\sigma_h^2}\right) + \frac{\mathcal{E}^{\text{sup}}}{\sigma_h^2}\left(1 - \frac{\rho_d^2}{\rho_t^2}\kappa\right)\right]\rho_t t_k \sigma_h^2 \end{aligned}$$

where $K[\infty] = \lim_{k\to\infty} K[k]$, and it can be obtained from (24) as

$$K[\infty] = \frac{1 - \frac{\mathcal{E}^{\text{sup}}}{\sigma_h^2}}{\rho_t t_k \left(\frac{a^2}{1-a^2}\frac{\mathcal{E}^{\text{sup}}}{\sigma_h^2} + 1\right)}.$$

The BER is then given by

$$\begin{aligned}P_e &= \mathrm{E}\{\Pr(\hat{d}_k \neq d_k \,|\, d_k)\} = \mathrm{E}\left\{\frac{1}{2}\left[1 - |\zeta[k]|\right]\right\} \\ &= \frac{1}{2} - \frac{1}{4}\frac{\frac{\mathcal{E}^{\text{sup}}}{\sigma_h^2}\left(1 - \frac{\rho_d^2}{\rho_t^2}\kappa\right) + \frac{\rho_d}{\rho_t}\left(1 - \frac{\mathcal{E}^{\text{sup}}}{\sigma_h^2}\right)}{\sqrt{\left[\frac{1}{\kappa} + \left(2\frac{\rho_d^2}{\rho_t^2}\kappa + 2\frac{\rho_d}{\rho_t} - 1\right)\frac{\mathcal{E}^{\text{sup}}}{\sigma_h^2}\right]\left(1 - \frac{\mathcal{E}^{\text{sup}}}{\sigma_h^2}\right)}} \\ &\quad - \frac{1}{4}\frac{\frac{\mathcal{E}^{\text{sup}}}{\sigma_h^2}\left(1 - \frac{\rho_d^2}{\rho_t^2}\kappa\right) - \frac{\rho_d}{\rho_t}\left(1 - \frac{\mathcal{E}^{\text{sup}}}{\sigma_h^2}\right)}{\sqrt{\left[\frac{1}{\kappa} + \left(2\frac{\rho_d^2}{\rho_t^2}\kappa - 2\frac{\rho_d}{\rho_t} - 1\right)\frac{\mathcal{E}^{\text{sup}}}{\sigma_h^2}\right]\left(1 - \frac{\mathcal{E}^{\text{sup}}}{\sigma_h^2}\right)}}. \end{aligned} \tag{29}$$

IV. TDM versus Superimposed Schemes: Performance Comparison

In this section, we compare the optimal TDM training (RPP-1) with superimposed training under the same transmission power P. We thus need to impose the following power constraints:

$$P = \rho_t^2 + \rho_d^2 = \eta\sigma_p^2 + (1-\eta)\sigma_d^2, \quad \frac{\rho_d^2}{\rho_t^2} = \frac{\sigma_d^2}{\sigma_p^2}\frac{1-\eta}{\eta}. \tag{30}$$

The first constraint keeps the transmission power P used in each scheme the same, and the second one keeps the ratio of power allocated to pilots and data in each scheme the same. Then, for a TDM scheme with the percentage of pilot symbols η, pilot power σ_p^2, and data power σ_d^2, the corresponding power allocation coefficients ρ_d and ρ_t for the superimposed scheme are given by

$$\rho_t^2 = \eta\sigma_p^2, \quad \rho_d^2 = (1-\eta)\sigma_d^2$$

and κ in (25) can be rewritten by

$$\kappa = \frac{1}{\frac{\sigma_d^2}{\sigma_p^2}\frac{1-\eta}{\eta} + \frac{1}{\eta\sigma_h^2 \text{SNR}_p}}. \tag{31}$$

The normalized MMSEs (NMMSEs) corresponding to those in (20) and (26) are given by (32) and (33), shown at the bottom of the page. We note that for the superimposed scheme, channel tracking benefits from constant presence of pilot symbols. However, it is affected by both noise and the interference from data. This effect is evident from (33), where $\bar{\mathcal{E}}^{\text{sup}}(\eta, a, \sigma_h^2\text{SNR}_p, \sigma_d^2/\sigma_p^2)$ is a function of κ, which indicates the SINR level. The higher κ is, the smaller $\bar{\mathcal{E}}^{\text{sup}}(\eta, a, \sigma_h^2\text{SNR}_p, \sigma_d^2/\sigma_p^2)$ will be. On the other hand, the

$$\bar{\mathcal{E}}^{\text{TDM}}\left(\eta, a, \sigma_h^2\text{SNR}_p\right) \triangleq \frac{\mathcal{E}^{\text{TDM}}}{\sigma_h^2} = 1 - a^{2\frac{1-\eta}{\eta}}\left(1 - \frac{1}{\frac{1}{2}\left(1 + \sigma_h^2\text{SNR}_p\right) + \sqrt{\left[\frac{1}{2}\left(1 + \sigma_h^2\text{SNR}_p\right)\right]^2 + \frac{a^{\frac{2}{\eta}}}{1 - a^{\frac{2}{\eta}}}\sigma_h^2\text{SNR}_p}}\right) \tag{32}$$

$$\bar{\mathcal{E}}^{\text{sup}}\left(\eta, a, \sigma_h^2\text{SNR}_p, \frac{\sigma_d^2}{\sigma_p^2}\right) \triangleq \frac{\mathcal{E}^{\text{sup}}}{\sigma_h^2} = \frac{1}{\frac{1}{2}(1+\kappa) + \sqrt{\left[\frac{1}{2}(1+\kappa)\right]^2 + \frac{a^2}{1-a^2}\kappa}}. \tag{33}$$

RPP-1 scheme has the advantage of updating the channel state during training with no data interference, but there is no information sent to facilitate tracking during data transmission. For a given η, differing from the RPP-1 scheme where the NMMSE in (32) is only affected by SNR_p, the NMMSE for the superimposed scheme is also a function of data/pilot power ratio σ_d^2/σ_p^2. Thus, the performance under superimposed training varies with data power, whereas that under RPP-1 does not.

A. Limiting Cases

To gain insights into the fundamental differences of the two schemes, we consider the limiting performance in various cases.

1) Fast Fading ($a \to 0$): For this case, we have almost i.i.d. fading. For fixed $\sigma_h^2\text{SNR}_p$

$$\bar{\mathcal{E}}^{\text{TDM}}\left(\eta, a, \sigma_h^2\text{SNR}_p\right) = 1 - O\left(a^{2\frac{1-\eta}{\eta}}\right)$$
$$\bar{\mathcal{E}}^{\text{sup}}\left(\eta, a, \sigma_h^2\text{SNR}_p, \frac{\sigma_d^2}{\sigma_p^2}\right) = \frac{1}{1+\kappa} - O(a^2). \quad (34)$$

As expected, due to the constant presence of pilot symbols in the data stream, the superimposed scheme provides a better tracking ability than that of the RPP-1 scheme.

2) Slow Fading ($a \to 1$): As $a \to 1$, the channel becomes constant. In the limit, channel estimation becomes perfect for both schemes, and we have

$$\bar{\mathcal{E}}^{\text{TDM}}\left(\eta, a, \sigma_h^2\text{SNR}_p\right) = O\left(\sqrt{1-a^{\frac{1}{\eta}}}\right) \quad (35)$$

$$\bar{\mathcal{E}}^{\text{sup}}\left(\eta, a, \sigma_h^2\text{SNR}_p, \frac{\sigma_d^2}{\sigma_p^2}\right) = O(\sqrt{1-a}). \quad (36)$$

For constant channels, it is intuitive that both of the schemes should give perfect estimation at the steady state. However, it is evident from (35) and (36) that, as a becomes less than 1, the variation rate of $\bar{\mathcal{E}}^{\text{TDM}}(\eta, a, \sigma_h^2\text{SNR}_p)$ is faster than that of $\bar{\mathcal{E}}^{\text{sup}}(\eta, a, \sigma_h^2\text{SNR}_p, \sigma_d^2/\sigma_p^2)$. In other words, the NMMSE of TDM training deteriorates at a much more rapid rate than that of superimposed training, as the channel varies faster. This again demonstrates that the superimposed scheme provides better tracking performance by constant presence of pilot symbols.

3) High SNR ($SNR_p \to \infty$): This corresponds to the noiseless case. For RPP-1, the channel state over each pilot symbol can be perfectly estimated. Estimation errors during data transmission are due to the tracking ability of the Kalman filter, and we have

$$\bar{\mathcal{E}}^{\text{TDM}}(\eta, a, \text{SNR}_p) = 1 - a^{2\frac{1-\eta}{\eta}} + O\left(\frac{1}{\text{SNR}_p}\right). \quad (37)$$

For the superimposed scheme, although there is no noise, the interference from data is always present. Therefore, the tracking error is due to the data symbol interference, and we have

$$\bar{\mathcal{E}}^{\text{sup}}\left(\eta, a, \text{SNR}_p, \frac{\sigma_d^2}{\sigma_p^2}\right) = \frac{1}{\frac{1}{2}(1+\kappa) + \sqrt{\left[\frac{1}{2}(1+\kappa)\right]^2 + \frac{a^2}{1-a^2}\kappa}} + O\left(\frac{1}{\text{SNR}_p}\right) \quad (38)$$

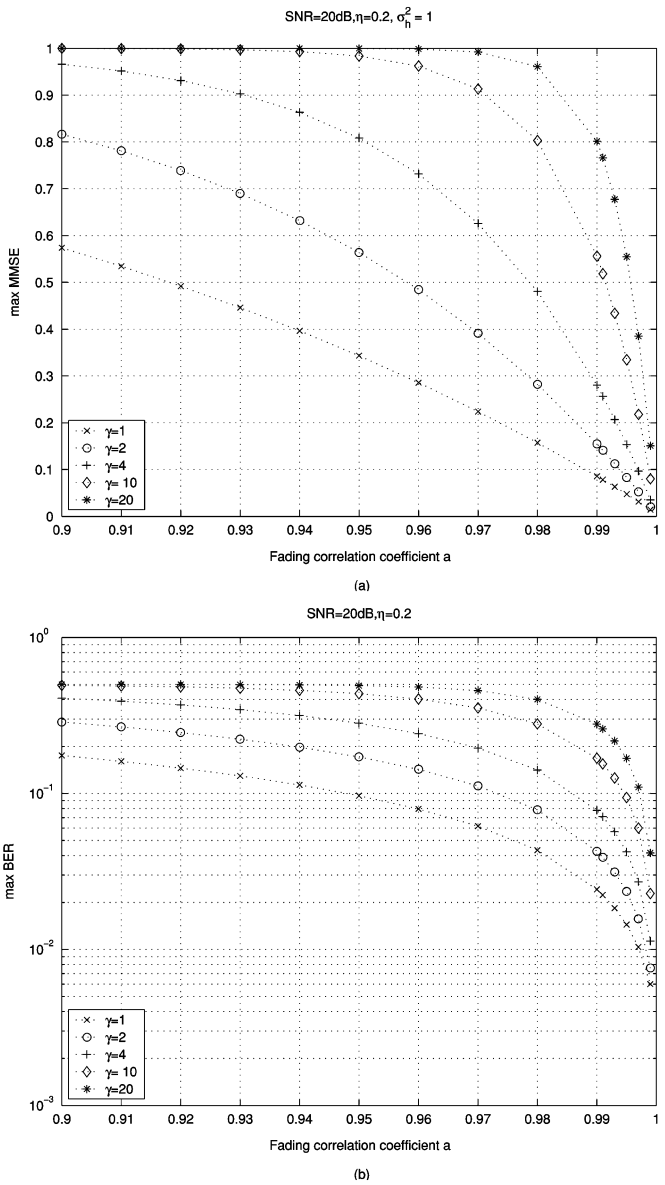


Fig. 4. (a) Maximum steady-state channel MMSE $\mathcal{E}(\gamma)$ versus a. (b) Maximum steady-state BER versus a. ($\eta = 20\%$, SNR = 20 dB).

where κ in this case is given by

$$\kappa = \frac{\sigma_p^2}{\sigma_d^2}\frac{\eta}{1-\eta}.$$

The NMMSEs for both schemes vary with SNR_p on the same order. The limiting NMMSEs depend on the channel fading rate, which is characterized by a. Because it is difficult to directly compare (37) and (38), we resort to numerical comparisons.

B. Numerical Comparisons

1) Optimal versus Suboptimal TDM Schemes: We compare the performance under different TDM RPP-γ placement schemes. The received SNR is defined as $\text{SNR} \triangleq \sigma_h^2 P/\sigma_w^2$. The MMSE and BER were calculated using MMSE expressions in Lemma 1 and the BER expression in (40), respectively. Fig. 4(a) and (b) shows the maximum steady-state MMSE and BER performance, respectively, under the variation of a for

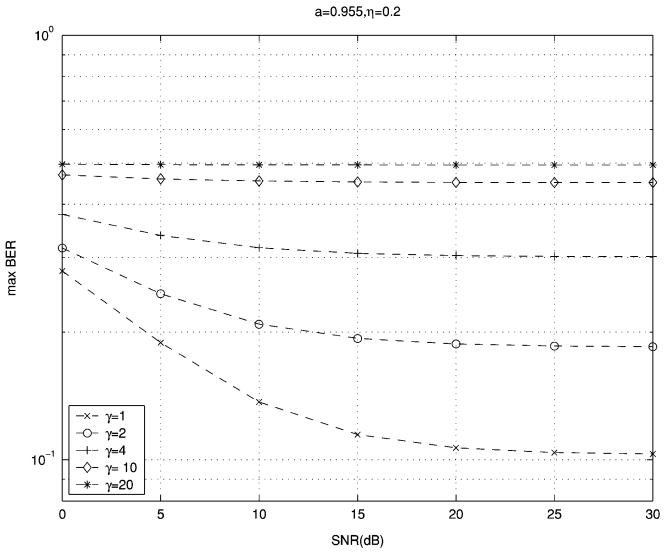


Fig. 5. Maximum steady-state BER versus SNR ($a = 0.985, \eta = 20\%$).

SNR $= 20$ dB. The percentage of pilot symbols in the stream was $\eta = 20\%$. The power of data and pilot symbols were set to be equal ($\sigma_d^2 = \sigma_p^2$). We observe that the largest gain obtained by placing pilot symbols optimally occurs when a is in the range from 0.9 to 1, which is a common range of channel time variation.[3] Fig. 5 shows the maximum steady-state BER performance under the variation of SNR at $a = 0.985$. Note that the gain of the optimal placement increases with SNR. Furthermore, placing pilot symbols optimally can result in a several decibel gain and achieve a much lower error floor.

2) Superimposed versus RPP-1 Schemes: Under the power constraints in (30), we calculate the MMSE and BER under superimposed training using expressions in (26) and (29), respectively, and compare them with those under RPP-1.

Figs. 6 and 7 show the MMSE and BER performance versus fading rate a for superimposed and RPP-1 schemes with $\eta = 10\%$ when SNR $= 20$ and 5 dB, respectively. We set half of the total transmission power P to pilot symbols, i.e., $\rho_d^2 = \rho_t^2$. As a comparison, average BER is also shown for the RPP-1 scheme. It is obtained by averaging the steady-state BER at all data positions in one placement period. Again, the steady-state BER at any data position can be calculated by MMSE expressions in Lemma 1 and the BER formula in (40).

For high SNR (20 dB), we observe in Fig. 6(b) that RPP-1 performs better than the superimposed scheme for slowly varying channels (a above 0.98). For such cases, the TDM scheme gives more accurate channel estimates during training than the superimposed training. However, as the channel varies more rapidly, the TDM training deteriorates at a more rapid rate than that of the superimposed scheme. It is apparent that even for the common fade rates of $0.9 \leq a \leq 0.97$, the superimposed scheme that offers better tracking is preferred. The advantage of the superimposed training is more pronounced when SNR is lowered to 5 dB, as shown in Fig. 7(b), where the effect of interference from data is less significant compared with the noise.

[3]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 [6].

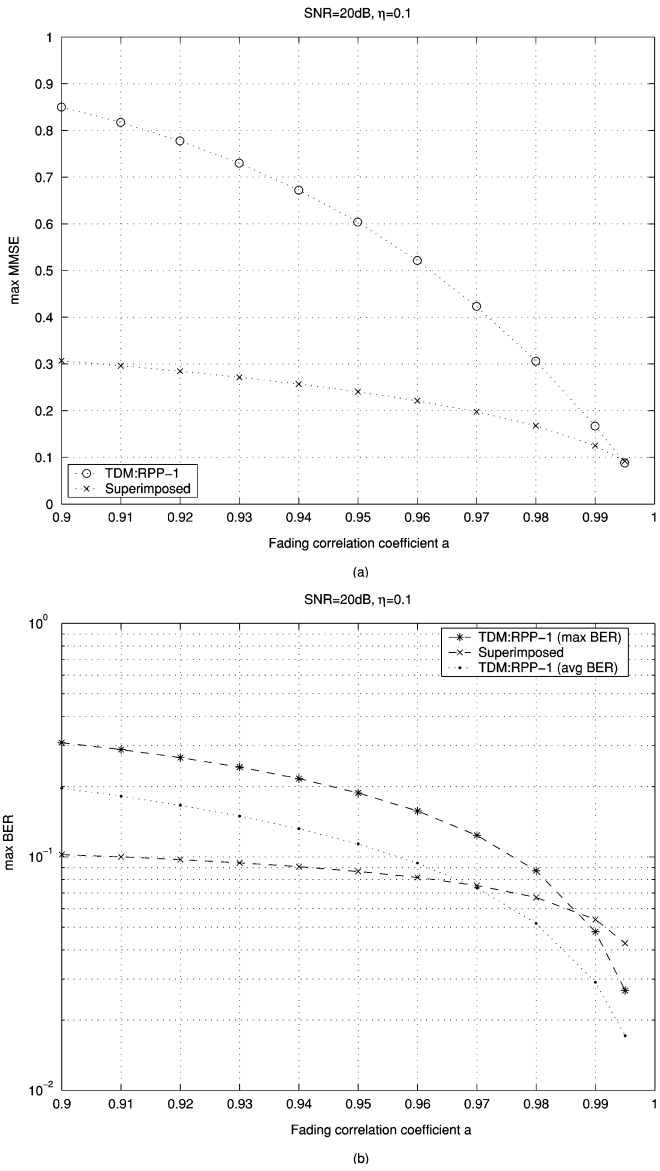


Fig. 6. (a) Maximum steady-state MMSE versus a. (b) Maximum steady-state BER versus a. (SNR $= 20$ dB, $\eta = 10\%$. Dotted line: Average BER for RPP-1.)

Fig. 8(a) and (b) shows the BER performance against SNR for $a = 0.99$ and $a = 0.95$, respectively. Similar performance gain regimes for each scheme can be seen. For $a = 0.99$ (very slow variation) at low SNR, we see that there is little difference in the performance under the two schemes. At high SNR, the RPP-1 scheme provides better performance. For $a = 0.95$, however, we see that the superimposed scheme uniformly performs better than RPP-1 at different SNR.

Under the power constraints in (30), Fig. 9 shows the BER versus the percentage of pilot symbols in the RPP-1 scheme, with SNR $= 20$ dB, and $a = 0.9$. Notice that RPP-1 benefits from a high percentage of pilot symbols, resulting in small tracking error. Therefore, in a high η regime, the BER is lower than that of the superimposed scheme.

Finally, we notice from the comparisons that the difference of the BER under different pilot-insertion strategies does not appear as big as that of the channel MMSE. This is due to the large decision region for BPSK signals (only the sign of the deci-

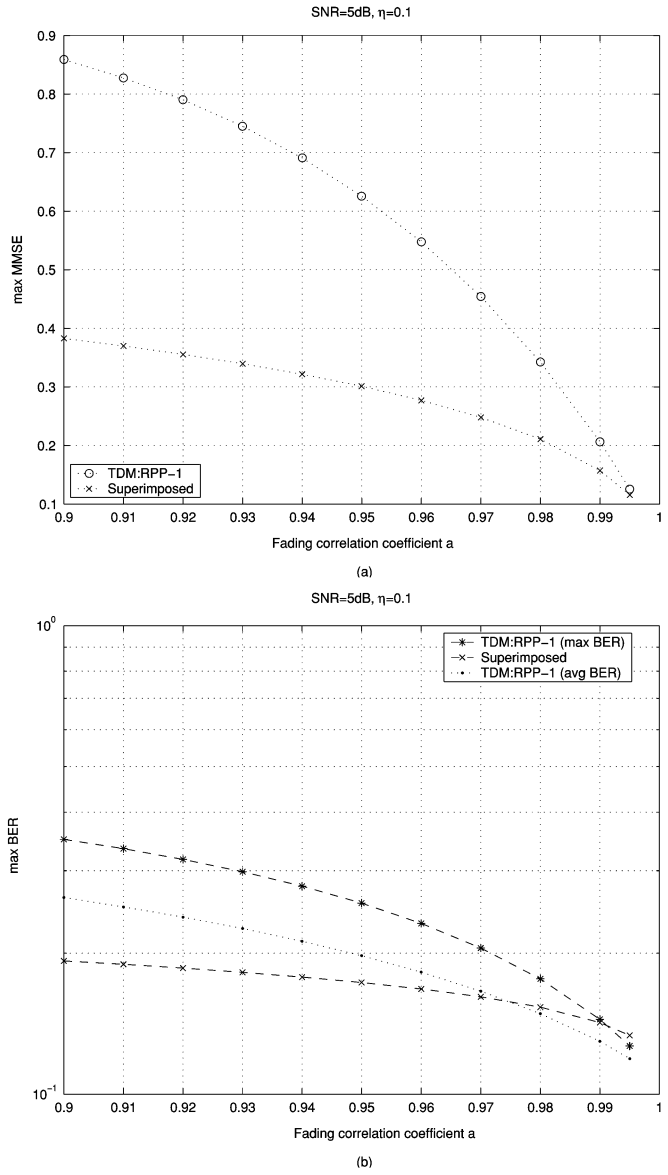


Fig. 7. (a) Maximum steady-state MMSE versus a. (b) Maximum steady-state BER versus a. [SNR $= 5$ dB, $\eta = 10\%$. Dotted line: Average BER for RPP-1.)]

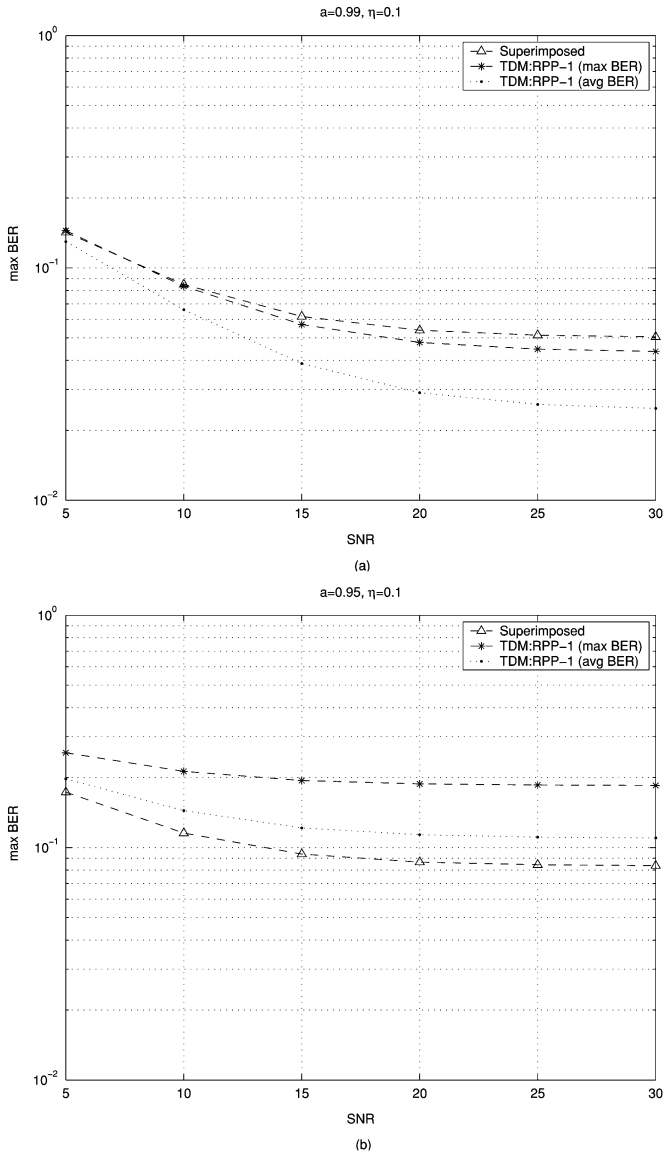


Fig. 8. (a) Maximum steady-state BER versus SNR. ($\mathrm{a} = 0.99, \eta = 10\%$.) (b) Maximum steady-state BER versus SNR. ($a = 0.95, \eta = 10\%$. Dotted line: Average BER for RPP-1.)

sion variable matters); therefore, BPSK is more forgiving with respect to a relatively inaccurate channel estimate. For higher order signal constellations, however, as the decision region becomes smaller for each symbol, we expect the channel MMSE to be more tightly coupled to the BER performance, and the difference in pilot-insertion strategies will result in a larger difference of BER performance.

3) Kalman Filter Convergence Rate: The comparison of pilot-insertion strategies in this paper is focused on the steady-state analysis. A practical issue arises on how long it takes for the receiver process to converge to its steady state. For continuous data transmissions, the process will eventually reach the steady state, and the beginning process has a negligible impact on the performance. For a packet transmission system, however, this question is particularly relevant.

For TDM periodic training, the time for convergence is the limit over the number of placement periods. From the Kalman filter analysis, similarly as in the training case [as in (42)], for the TDM training, the convergence rate over placement periods is exponential. However, the time for reaching the steady state depends on the fading rate a, SNR, and the initial values and, therefore, varies from application to application. We have tested the convergence time for different levels of fading rate and SNR (the typical range of fading rate $a \in [0.9, 0.99]$, and SNR from 0 to 20 dB), and the process converges typically within three to five placement periods. Therefore, the steady state can be reached in a short time. For a packet transmission system, such as GSM, if a packet contains 150 symbols with 20% pilots, there can be as many as 30 placement periods. The receiver often receives multiple packets continuously at a time. Therefore, our steady-state analysis is suitable for the system. Furthermore, from the above observations, we point out that besides the steady-state performance, the RPP-1 scheme has the additional advantage of faster convergence than other RPP-γ schemes, because it has shorter placement periods. Compared with TDM training, the steady state under the superimposed training can be reached within a few iterations (ten to 20 steps in our sim-

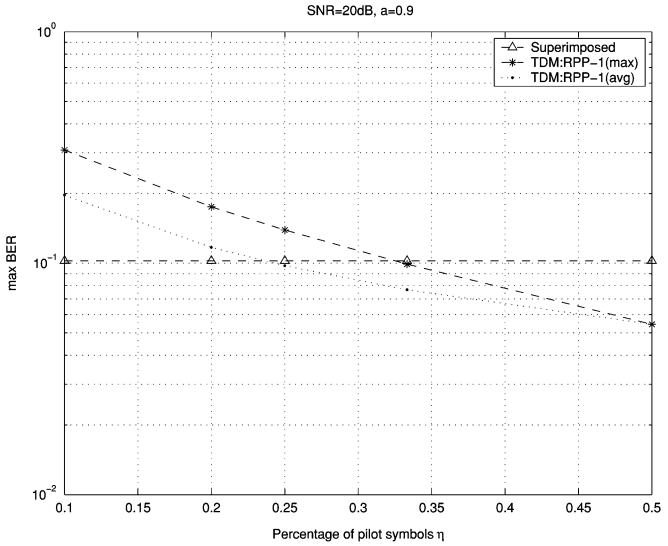


Fig. 9. BER versus η for $a = 0.9$, SNR = 20 dB. (Dotted line: Average BER for RPP-1.)

ulations). Thus, the convergence for the superimposed training can be faster than that of TDM training, where it needs several placement periods, especially for the case with longer periods.[4]

V. Conclusion

In this paper, we have studied two different forms of training schemes, using the MMSE of channel estimation and uncoded BER as the figures of merit. For Gauss–Markov fading channels, we have established the optimality of the single-pilot periodic training (RPP-1) among all periodic TDM training schemes. The optimality of RPP-1 holds uniformly across all SNR levels and all fade rates. This result allows us to compare the best TDM training with superimposed training. We showed that while the traditional TDM training performs better for slow fading channels at high SNR, the superimposed scheme outperforms the best TDM scheme in regimes of practical importance.

The performance metrics chosen in this paper are practical but limited from an information theoretic perspective. Although we have shown the connection between MMSE and BER, we have not considered coding. To this end, the work by Medard *et al.* [6] is the most relevant. In their work, adaptive modulation and coding for channels with PSAM is considered, and the spacing between the pilot symbols is optimized numerically by maximizing the mutual information with binary inputs, resulting in improved channel capacity. From our results, the complete characterization of MMSE and BER performance should provide guidelines on code design, rate allocation, and power allocation.

Appendix A
Kalman Filer for TDM Training

- During the pilot cluster transmissions ($k \in \mathcal{I}_p(\boldsymbol{\mathcal{P}})$):

Kalman Gain:

$$K[lT+k] = \frac{\left(a^2 M[lT+k-1;\boldsymbol{\mathcal{P}}] + (1-a^2)\sigma_h^2\right) s_{lT+k}}{\sigma_w^2 + \left(a^2 M[lT+k-1;\boldsymbol{\mathcal{P}}] + (1-a^2)\sigma_h^2\right)\sigma_p^2}.$$

Channel estimation update:

$$\hat{h}_{lT+k} = a\hat{h}_{lT+k-1} + K[k](y_{lT+k} - a\hat{h}_{lT+k-1}s_{lT+k}).$$

MMSE update:

$$M[lT+k;\boldsymbol{\mathcal{P}}] = (1 - K[lT+k]s_{lT+k}) \times \left(a^2 M[lT+k-1;\boldsymbol{\mathcal{P}}] + (1-a^2)\sigma_h^2\right).$$

- During the data block transmissions ($k \notin \mathcal{I}_p(\boldsymbol{\mathcal{P}})$):

Channel estimation update:

$$\hat{h}_{lT+k} = a\hat{h}_{lT+k-1}.$$

MMSE update:

$$M[lT+k;\boldsymbol{\mathcal{P}}] = a^2 M[lT+k-1;\boldsymbol{\mathcal{P}}] + (1-a^2)\sigma_h^2.$$

Appendix B
Proof of Proposition 1

Proof: The ML detection rule is given in (3). For a system using BPSK, it can be simplified as

$$\hat{s}_k = \text{sign}\{\text{Re}\{\hat{h}_k^* y_k\}\}\sigma_d.$$

Under the system equation in (1), the decision statistic is

$$\begin{aligned} b_k &= \text{Re}\{\hat{h}_k^* y_k\} = \text{Re}\{s_k \hat{h}_k^*(\hat{h}_k + \tilde{h}_k) + \hat{h}_k^* w_k\} \\ &= s_k|\hat{h}_k|^2 + s_k \text{Re}\{\hat{h}_k^* \tilde{h}_k\} + \text{Re}\{\hat{h}_k^* w_k\} \end{aligned} \quad (39)$$

where $\tilde{h}_k \triangleq h_k - \hat{h}_k$. Conditioned on $\hat{h}_k$ and s_k, the second and third terms in (39) are independent zero mean Gaussian random variables. At the steady state

$$b_k \sim \mathcal{N}\left(s_k|\hat{h}_k|^2, \frac{1}{2}\left(\sigma_d^2 M_k(\boldsymbol{\mathcal{P}}) + \sigma_w^2\right)|\hat{h}_k|^2\right).$$

Therefore, for a system using BPSK, the bit error probability conditioned on $\hat{h}_k$ and s_k is

$$\Pr(\hat{s}_k \neq s_k \,|\, \hat{h}_k, s_k) = Q\left(\sqrt{\frac{2|\hat{h}_k|^2\sigma_d^2}{\sigma_d^2 M_k(\boldsymbol{\mathcal{P}}) + \sigma_w^2}}\right).$$

Define $\text{SNR}_d \triangleq \sigma_d^2/\sigma_w^2$. The BER for data symbols at the kth position of a placement period is thus given by

$$\begin{aligned} P_e(k;\boldsymbol{\mathcal{P}}) &= \text{E}\left\{Q\left(\sqrt{\frac{2|\hat{h}_k|^2\sigma_d^2}{\sigma_d^2 M_k(\boldsymbol{\mathcal{P}}) + \sigma_w^2}}\right)\right\} \\ &= \frac{1}{2}\left[1 - \sqrt{\frac{1 - \frac{M_k(\boldsymbol{\mathcal{P}})}{\sigma_h^2}}{1 + \frac{1}{\sigma_h^2 \text{SNR}_d}}}\right] \end{aligned} \quad (40)$$

[4]Due to space limitations, we do not provide plots on the MMSE update trajectory over iterations but only the result on convergence time, which is the information needed here. The plots can be found in the technical report in [23].

where to obtain (40), we use the following result:

$$\int_0^\infty Q(\sqrt{x})(1/a)e^{-x/a}\,dx = \frac{1}{2}[1-\sqrt{a/(2+a)}]$$

and $x \sim \exp(a) = (1/a)e^{-x/a}$.[5]

For QPSK signaling ($s_k = (\pm 1 \pm j)\sigma_d/\sqrt{2}$), the decision rule is

$$\mathrm{Re}\{\hat{s}_k\} = \mathrm{sign}\{\mathrm{Re}\{\hat{h}_k^* y_k\}\}\frac{\sigma_d}{\sqrt{2}}$$
$$\mathrm{Im}\{\hat{s}_k\} = \mathrm{sign}\{\mathrm{Im}\{\hat{h}_k^* y_k\}\}\frac{\sigma_d}{\sqrt{2}}.$$

The bit error probability conditioned on $\hat{h}_k$ and s_k can be derived similarly as in the BPSK case

$$\Pr(\text{bit error}\,|\,\hat{h}_k, s_k) = Q\left(\sqrt{\frac{|\hat{h}_k|^2\sigma_d^2}{\sigma_d^2 M_k(\boldsymbol{\mathcal{P}})+\sigma_w^2}}\right)$$

and the BER at the kth position of a period is given by

$$P_e(k;\boldsymbol{\mathcal{P}}) = \mathrm{E}\left\{Q\left(\sqrt{\frac{|\hat{h}_k|^2\sigma_d^2}{\sigma_d^2 M_k(\boldsymbol{\mathcal{P}})+\sigma_w^2}}\right)\right\} = \frac{1}{2}\left[1-\sqrt{\frac{1-\frac{M_k(\boldsymbol{\mathcal{P}})}{\sigma_h^2}}{1+\frac{M_k(\boldsymbol{\mathcal{P}})}{\sigma_h^2}+\frac{2}{\sigma_h^2\mathrm{SNR}_d}}}\right]. \quad (41)$$

The BER expressions for BPSK and QPSK signaling are now obtained as functions of the steady-state channel MMSE with placement $\boldsymbol{\mathcal{P}}$ in (40) and (41), respectively. In both cases, it is clear that increasing $M_k(\mathcal{P})$ results in increased $P_e(k;\boldsymbol{\mathcal{P}})$. It immediately follows that in either case, the optimization in (10) is equivalent to that in (9). ■

APPENDIX C
PROOF OF LEMMA 1

Proof: At the steady state of an RPP-γ scheme, the channel MMSE attains a periodic steady state. During a training period ($1 \le k \le \gamma$), M_k obeys the same update recursion as in (4)

$$M_k(\gamma) = \frac{\sigma_w^2\left(a^2 M_{k-1}(\gamma)+\sigma_u^2\right)}{\sigma_w^2+\left(a^2 M_{k-1}(\gamma)+\sigma_u^2\right)\sigma_p^2}$$

[5]Note that $\hat{h}_k$ is zero mean Gaussian random variable, and therefore, $|\hat{h}_k|^2$ is exponentially distributed.

where $\sigma_u^2 \triangleq \sigma_h^2(1-a^2)$. Define $\delta_k(\gamma)$ as the difference of M_k and M_∞; then, we have the equation shown at the bottom of the page. We then have the following first-order differential equation for $\delta_k(\gamma)$

$$\frac{1}{\delta_k(\gamma)} = \alpha\frac{1}{\delta_{k-1}(\gamma)}+\beta, \quad k=1,\ldots,\gamma$$

where α and β are defined in (18). Note that in the above recursion, when $k=1$, $\delta_0(\gamma)$ corresponds to the value of $\delta_k(\gamma)$ over the end position of the previous placement period, i.e., $\delta_0(\gamma) = \delta_T(\gamma)$. Therefore, we can express $\delta_k(\gamma)$ in terms of $\delta_T(\gamma)$ by

$$\frac{1}{\delta_k(\gamma)} = \alpha^k\frac{1}{\delta_T(\gamma)}+\beta\frac{1-\alpha^k}{1-\alpha}, \quad 1\le k\le\gamma. \quad (42)$$

During data transmission ($\gamma+1 \le k \le T$), the updating recursion for $M_k(\gamma)$ is in (7). Therefore

$$\delta_k(\gamma) = a^{2(k-\gamma)}\delta_\gamma(\gamma)+(1-a^{2(k-\gamma)})\rho_\infty\sigma_h^2, \quad k=\gamma+1,\ldots,T. \quad (43)$$

From (42) and (43), $\delta_\gamma(\gamma)$ and $\delta_T(\gamma)$ satisfy the following relations:

$$\begin{cases}\delta_\gamma(\gamma) = \dfrac{\delta_T(\gamma)}{1+\delta_T(\gamma)\frac{\beta\left(1-\frac{1}{\alpha^\gamma}\right)}{\alpha-1}}\left(\frac{1}{\alpha}\right)^\gamma \\ \delta_T(\gamma) = a^{2\gamma\frac{1-\eta}{\eta}}\delta_\gamma(\gamma)+\left(1-a^{2\gamma\frac{1-\eta}{\eta}}\right)\rho_\infty\sigma_h^2\end{cases} \quad (44)$$

where for a given η, we have used the relation $T=\gamma/\eta$. The steady-state equation for $\delta_T(\gamma)$ is then given by

$$\frac{\beta\left(1-\frac{1}{\alpha^\gamma}\right)}{\alpha-1}\delta_T^2(\gamma) + \left[1-\left(\frac{a^{2\frac{1-\eta}{\eta}}}{\alpha}\right)^\gamma-\frac{\beta\left(1-\frac{1}{\alpha^\gamma}\right)}{\alpha-1}\left(1-a^{2\gamma\frac{1-\eta}{\eta}}\right)\rho_\infty\sigma_h^2\right] \times\delta_T(\gamma)-\left(1-a^{2\gamma\frac{1-\eta}{\eta}}\right)\rho_\infty\sigma_h^2 = 0. \quad (45)$$

Solving the above equation, we have the expression of $\delta_T(\gamma)$, as a function of γ, given in (15).

From (7), for any RPP-γ scheme, the expression of the steady-state channel MMSE $M_k(\gamma)$ over each data symbol is then obtained in (13) and (14). ■

$$\begin{aligned}\delta_k(\gamma) &\triangleq M_k(\gamma)-M_\infty \\ &= \frac{\sigma_w^2\left(a^2M_{k-1}(\gamma)+\sigma_u^2\right)}{\sigma_w^2+\left(a^2\delta_{k-1}(\gamma)+a^2M_\infty+\sigma_u^2\right)\sigma_p^2}-\frac{\sigma_w^2\left(a^2M_\infty+\sigma_u^2\right)}{\sigma_w^2+\left(a^2M_\infty+\sigma_u^2\right)\sigma_p^2} \\ &= \frac{a^2\sigma_w^4\delta_{k-1}(\gamma)}{\left(a^2\sigma_p^2\delta_{k-1}(\gamma)+\sigma_w^2+\left(a^2M_\infty+\sigma_u^2\right)\sigma_p^2\right)\left(\sigma_w^2+\left(a^2M_\infty+\sigma_u^2\right)\sigma_p^2\right)}.\end{aligned}$$

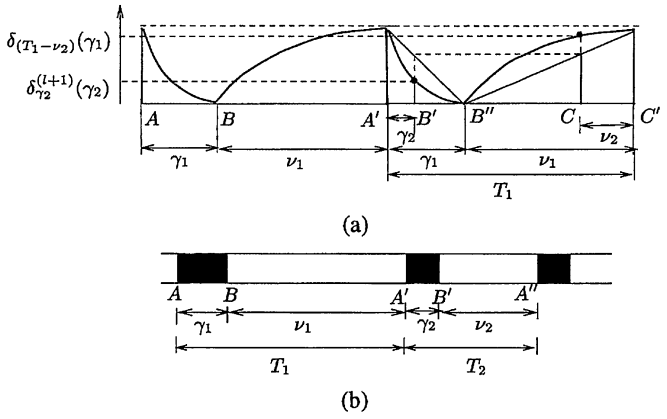


Fig. 10. Proof of Theorem 1.

APPENDIX D
PROOF OF THEOREM 1

Proof: The algebraic proof of Theorem 1, based on the expression for $\delta_T(\gamma)$ in (15) as a function of γ, can be found in the technical report in [23]. Here, we give a more intuitive graphic-aided proof using Fig. 10.

The basic idea in this proof is the following. We first let the data stream contain larger pilot clusters (thus longer period). After the process reaches its steady state, we change the pilot placement to the one with smaller pilot clusters (thus shorter period). After this rearrangement, we show that the channel MMSE at the last position of a period (i.e., $M_T(\gamma)$) is smaller in the new placement than it is in the previous placement. This eventually results in the decreased MMSE when the new process goes to its steady state.

For a RPP-γ scheme, during training, from (42), we have

$$\delta_k(\gamma) = \frac{\delta_T(\gamma)}{1+\delta_T(\gamma)\frac{\beta\left(1-\frac{1}{\alpha^k}\right)}{\alpha-1}}\left(\frac{1}{\alpha}\right)^k, \quad k=1,\ldots,\gamma.$$

Thus, $\delta_k(\gamma)$ exponentially decreases with rate $(1/\alpha)^k$. During a data block, it follows from (43) that $(\rho_\infty\sigma_h^2 - \delta_k(\gamma))$ exponentially decreases at rate $a^{2(k-\gamma)}$.

Fig. 10(b) and (a) describe the placement pattern $\boldsymbol{\mathcal{P}}$ and the corresponding steady-state trajectory of $\delta_k(\gamma)(=M_k(\gamma)-M_\infty)$, respectively. For fixed pilot percentage η, let us consider two schemes: RPP-γ_1 with $\boldsymbol{\mathcal{P}}_1 = (\gamma_1, \nu_1)$ and RPP-γ_2 with $\boldsymbol{\mathcal{P}}_2 = (\gamma_2, \nu_2)$, where $\gamma_1 > \gamma_2$. The RPP-γ_1 scheme, where the placement period is T_1, is shown in the left part of Fig. 10(b). Indexes A and A' denote the end positions of the data blocks in the $(l-1)$th and lth placement periods under RPP-γ_1, respectively. Index B denotes the end position of the pilot cluster in the lth period. Assume that in the $(l-1)$th and lth placement periods, the channel MMSE is at its steady state. The corresponding $\delta_k(\gamma_1)$ at A and A' is $\delta_{T_1}(\gamma_1)$. The trajectory curve $C_{A\to B\to A'}$ of $\delta_k(\gamma_1)$ is shown in Fig. 10(a). Because the changing rates of$\delta_k(\gamma_1)$ during pilot and data cluster are exponential, $C_{A\to B}$ and $C_{B\to A'}$ are both exponential. If RPP-γ_1 is still used in the $(l+1)$th placement period, then the trajectory of $\delta_k(\gamma_1)$ is the curve $C_{A'\to B''\to C'}$ in Fig. 10(a). It is equivalent to $C_{A\to B\to A'}$. Now, after the lth period, we change the placement to RPP-γ_2, which is shown in the right part of Fig. 10(b), where indexes B' and A'' denote the new end positions of pilot and data cluster in the period, respectively. The change of placement results in the new MMSE. Denote $\delta_{\gamma_2}^{(l+1)}(\gamma_2) = M[lT+\gamma_2] - M_\infty$ over B' and, similarly, $\delta_{T_2}^{(l+1)}(\gamma_2)$ over A''. Note that $\delta_{\gamma_2}^{(l+1)}(\gamma_2)$ is still on the trajectory curve $C_{A'\to B'}$ in Fig. 10(a), i.e., $\delta_{\gamma_2}^{(l+1)}(\gamma_2) = \delta_{\gamma_2}(\gamma_1)$. Now, let C in Fig. 10(a) be the point such that the length $CC' = \nu_2$. Because for fixed pilot percentage $\eta, \gamma_1/\gamma_2 = \nu_1/\nu_2$, we have $A'B'/A'B'' = CC'/B''C'$, which is shown in Fig. 10(a). Then, because $C_{A'\to B''}$ and $C_{B''\to C'}$ are exponential, where the former one is convex and the latter one is concave, from the geometry, we have

$$\delta_{\gamma_2}^{(l+1)}(\gamma_2) = \delta_{\gamma_2}(\gamma_1) \le \delta_{(T_1-\nu_2)}(\gamma_1) \tag{46}$$

which is shown in Fig. 10(a). From (43), since

$$\delta_{T_2}^{(l+1)}(\gamma_2) = a^{2\nu_2}\delta_{\gamma_2}^{(l+1)}(\gamma_2) + (1-a^{2\nu_2})\rho_\infty\sigma_h^2$$
$$\delta_{T_1}(\gamma_1) = a^{2\nu_2}\delta_{T_1-\nu_2}(\gamma_1) + (1-a^{2\nu_2})\rho_\infty\sigma_h^2$$

it follows from (46) that $\delta_{T_2}^{(l+1)}(\gamma_2) \le \delta_{T_1}(\gamma_1)$. Consequently, $\delta_{T_2}^{(j)}(\gamma_2) \le \delta_{T_1}(\gamma_1)$, for $j > l$. At the steady state of RPP-γ_2, we have

$$\delta_{T_2}(\gamma_2) = \lim_{l\to\infty}\delta_{T_2}^{(l+1)}(\gamma_2) \le \delta_{T_1}(\gamma_1).$$

Because $\mathcal{E}(\gamma_i) = \delta_{T_i}(\gamma_i) + M_\infty$ and M_∞ is not a function of γ, we have $\mathcal{E}(\gamma_1) \ge \mathcal{E}(\gamma_2)$, for $\gamma_1 > \gamma_2$. The minimum $\mathcal{E}(\gamma)$ can be obtained using Lemma 1. ■

APPENDIX E
PROOF OF THEOREM 2

Recall that d_i is the index for the end position of the ith data block relative to the beginning of a period. We have the following lemma.

Lemma 2: Given $\boldsymbol{\mathcal{P}} = (\boldsymbol{\gamma}, \boldsymbol{\nu})$, with n pilot clusters in a placement period, for any $2 \le i \le n$

$$M_{d_i}(\boldsymbol{\gamma}, \boldsymbol{\nu}) \le M_{d_i}(\boldsymbol{\gamma}, \boldsymbol{\nu} - \mathbf{e}_{i-1} + \mathbf{e}_i) \tag{47}$$
$$M_{d_{i-1}}(\boldsymbol{\gamma}, \boldsymbol{\nu}) \ge M_{d_{i-1}}(\boldsymbol{\gamma}, \boldsymbol{\nu} - \mathbf{e}_{i-1} + \mathbf{e}_i) \tag{48}$$

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

Proof: We use Fig. 11 to assist in our proof. The figure describes the placement pattern $\boldsymbol{\mathcal{P}}$ in a period, and the steady-state trajectory of $\delta_k(\boldsymbol{\mathcal{P}})(=M_k(\boldsymbol{\mathcal{P}}) - M_\infty)$. Relative to the beginning of a period, let A and C denote the end positions of the $(i-1)$th and ith data block and B the end position of the ith pilot cluster. Let $\boldsymbol{\mathcal{P}}' = (\boldsymbol{\gamma}, \boldsymbol{\nu}')$ be the new placement satisfying $\boldsymbol{\nu}' = \boldsymbol{\nu} + \mathbf{e}_{i-1} - \mathbf{e}_i$. Then, showing (47) is equivalent to showing

$$\delta_C(\boldsymbol{\gamma}, \boldsymbol{\nu}') < \delta_C(\boldsymbol{\gamma}, \boldsymbol{\nu}).$$

For simplicity, we denote δ_C' as $\delta_C(\boldsymbol{\gamma}, \boldsymbol{\nu}')$ and δ_C as $\delta_C(\boldsymbol{\gamma}, \boldsymbol{\nu})$ and, similarly, those of other points. Assume in the $(l-1)$th placement period that the MMSE is at its steady state. In the lth period, we move the ith pilot cluster right by one step. Let

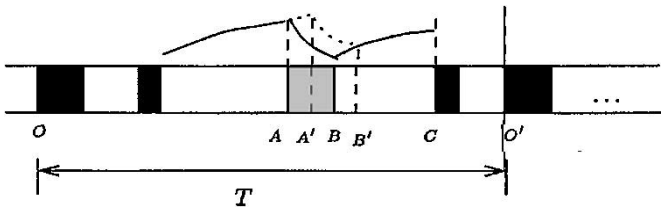


Fig. 11. Proof of Lemma 2.

$A' = A + 1$ and $B' = B + 1$. This results in the new MMSE at $(l-1)T + A'$ and thereafter. Denote $\delta'^{(l)}_{A'} = M[(l-1)T + A'] - M_\infty$ and, similarly, $\delta'^{(l)}_{B'}$ and $\delta'^{(l)}_{C}$. Then, from (42) and (43), we have

$$\begin{cases} \delta'^{(l)}_{A'} = a^2\delta_A + (1-a^2)\rho_\infty\sigma_h^2 \\ \frac{1}{\delta'^{(l)}_{B'}} = \frac{\alpha^{\gamma_i}}{\delta_{A'}} + \beta\frac{\alpha^{\gamma_i}-1}{\alpha-1} \\ \frac{1}{\delta_B} = \frac{\alpha^{\gamma_i}}{\delta_A} + \beta\frac{\alpha^{\gamma_i}-1}{\alpha-1}. \end{cases} \tag{49}$$

The one-step increment on the trajectory $B \to C$ is $\epsilon_B = (1-a^2)(\rho_\infty\sigma_h^2 - \delta_B)$. If we can show

$$\delta'^{(l)}_{B'} < \delta_B + \epsilon_B \tag{50}$$

then $\delta'^{(l)}_{C} < \delta_C$. Consequently, $\delta'^{(j)}_{C} < \delta_C, \forall j > l$, and $\delta'_C = \lim_{j\to\infty}\delta'^{(j)}_{C} < \delta_C$. Therefore, we only need to show (50). Let $\delta'^{(l)}_{A'} = \delta_A + \epsilon_A$, where $\epsilon_A > 0$. From the first equation of (49), we have $\epsilon_A = (1-a^2)(\rho_\infty\sigma_h^2 - \delta_A)$. Let $\bar\beta \triangleq \beta\frac{\alpha^{\gamma_i}-1}{\alpha-1}$. From the second and third equations of (49), we have

$$\begin{aligned} &\delta'^{(l)}_{B'} - (\delta_B + \epsilon_B) \\ &= \frac{\alpha^{\gamma_i}\epsilon_A}{[\alpha^{\gamma_i} + \bar\beta(\delta_A+\epsilon_A)][\alpha^{\gamma_i} + \bar\beta\delta_A]} - \epsilon_A \\ &= (1-a^2)\left(\frac{\alpha^{\gamma_i}(\rho_\infty - \delta_A)}{[\alpha^{\gamma_i} + \bar\beta(\delta_A+\epsilon_A)][\alpha^{\gamma_i} + \bar\beta\delta_A]}\right. \\ &\quad \left. - \left(\rho_\infty - \frac{\delta_A}{\alpha^{\gamma_i} + \bar\beta\delta_A}\right)\right) \\ &< (1-a^2)\left((\rho_\infty - \delta_A) - \left(\rho_\infty - \frac{\delta_A}{\alpha^{\gamma_i} + \bar\beta\delta_A}\right)\right) \\ &< 0 \end{aligned}$$

where the inequalities are due to $\alpha > 1$. Therefore, we have proved (47). By a similar argument, we can show (48). ∎

Lemma 3: Given η, for any $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_n]$, $\exists \boldsymbol{\nu}^* \in \mathbf{R}^{n\times 1}$ such that

$$M_{d_i}(\boldsymbol{\gamma}, \boldsymbol{\nu}^*) = M_{d_j}(\boldsymbol{\gamma}, \boldsymbol{\nu}^*) = \min_{\boldsymbol{\nu}} \mathcal{E}(\boldsymbol{\gamma}, \boldsymbol{\nu}). \tag{51}$$

Proof: Let $\boldsymbol{\nu}_i = [\nu_1, \ldots, \nu_i], 1 \le i \le n$. For fixed $\nu_{i+1}, \ldots, \nu_n$, let

$$\mathcal{E}^{(i)}(\boldsymbol{\gamma}, \boldsymbol{\nu}) \triangleq \max_{1\le j\le i} M_{d_j}(\boldsymbol{\gamma}, \boldsymbol{\nu}), \quad \boldsymbol{\nu}_i^* = \arg\min_{\boldsymbol{\nu}_i} \mathcal{E}^{(i)}(\boldsymbol{\gamma}, \boldsymbol{\nu}).$$

We first show that

$$\exists \nu_n^*, \text{s.t.} \quad M_{d_n}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n^*) = \mathcal{E}^{(n-1)}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n^*). \tag{52}$$

where

$$\mathcal{E}^{(n-1)}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n^*) = \min_{\boldsymbol{\nu}_{n-1}} \mathcal{E}^{(n-1)}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}, \nu_n^*). \tag{53}$$

It is not hard to see that $M_{d_n}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n)$ and $\mathcal{E}^{(n-1)}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n)$ are both continuous function of ν_n when $\nu_n \in \mathbf{R}$. Therefore, we only need to check if the two functions $M_{d_n}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n)$ and $\mathcal{E}^{(n-1)}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n)$ have a cross point. When $\nu_n \to T - \|\boldsymbol{\gamma}\|_1$,[6] all n pilot clusters are clustered together. Note that

$$\lim_{\nu_n \to T-\|\boldsymbol{\gamma}\|_1} M_{d_i}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n) = \lim_{\nu_n \to T-\|\boldsymbol{\gamma}\|_1} M_{p_i}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n)$$

where we recall that p_i is the index of the end of the ith pilot cluster. Because the channel MMSE monotonically decreases during training, we have

$$\begin{aligned} &\lim_{\nu_n \to T-\|\boldsymbol{\gamma}\|_1} \mathcal{E}^{(n-1)}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n) \\ &\quad = \lim_{\nu_n \to T-\|\boldsymbol{\gamma}\|_1} M_{p_1}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n) \\ &\quad \le \lim_{\nu_n \to T-\|\boldsymbol{\gamma}\|_1} M_{d_n}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n) \end{aligned} \tag{54}$$

where the inequality is due to the channel MMSE decrease during the first pilot cluster. Similarly

$$\begin{aligned} \lim_{\nu_n \to 0} M_{d_n}(\boldsymbol{\gamma}, \boldsymbol{\nu}^*, \nu_n) &= \lim_{\nu_n \to 0} M_{p_n}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n) \\ &\le \lim_{\nu_n \to 0} M_{d_{n-1}}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n) \\ &\le \lim_{\nu_n \to 0} \mathcal{E}^{(n-1)}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n). \end{aligned} \tag{55}$$

Combining (54) and (55), we have (52).

Now, fixing this ν_n^*, we again show that $\exists \nu_{n-1}^* \in (0, T - \|\boldsymbol{\gamma}\|_1 - \nu_n^*)$, s.t.

$$\mathcal{E}^{(n-2)}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-2}^*, \nu_{n-1}^*, \nu_n^*) = M_{d_{n-1}}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-2}^*, \nu_{n-1}^*, \nu_n^*). \tag{56}$$

Again, it suffices to check if the two functions have a cross point by checking the two end point values that ν_{n-1} can take. Similar to (55) above, we have

$$\begin{aligned} &\lim_{\nu_{n-1} \to 0} M_{d_{n-1}}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-2}^*, \nu_{n-1}, \nu_n^*) \\ &\qquad \le \lim_{\nu_{n-1} \to 0} \mathcal{E}^{(n-2)}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-2}^*, \nu_{n-1}, \nu_n^*). \end{aligned}$$

Assume there is no cross point, i.e.,

$$M_{d_{n-1}}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-2}^*, \nu_{n-1}, \nu_n^*) < \mathcal{E}^{(n-2)}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-2}^*, \nu_{n-1}, \nu_n^*)$$

for $\nu_{n-1} \in (0, T - \|\boldsymbol{\gamma}\|_1 - \nu_n^*)$. Then, $\mathcal{E}^{(n-1)}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-1}^*, \nu_n^*) = \min_{\nu_{n-1}} \mathcal{E}^{(n-2)}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-2}^*, \nu_{n-1}, \nu_n^*)$. By Lemma 2, it follows that

$$\nu_{n-1}^* = \arg\min_{\nu_{n-1}} \mathcal{E}^{(n-2)}(\boldsymbol{\gamma}, \boldsymbol{\nu}_{n-2}^*, \nu_{n-1}, \nu_n^*) = T - \|\boldsymbol{\gamma}\|_1 - \nu_n^*$$

[6]We define $\|\boldsymbol{\gamma}\|_1 = \sum_{i=1}^n \gamma_i$.

i.e., when all first $(n-1)$ pilot clusters are together. Then, we have

$$\begin{aligned}\mathcal{E}^{(n-1)}(\boldsymbol{\gamma},\boldsymbol{\nu}^*_{n-1},\nu^*_n) &= \lim_{\nu_{n-1}\to T-\|\boldsymbol{\gamma}\|_1-\nu^*_n} M_{d_1}(\boldsymbol{\gamma},\boldsymbol{\nu}^*_{n-2},\nu_{n-1},\nu^*_n)\\ &= M_{p_1}(\boldsymbol{\gamma},\boldsymbol{\nu}^*_{n-2},\nu_{n-1},\nu^*_n)\\ &< M_{d_n}(\boldsymbol{\gamma},\boldsymbol{\nu}^*_{n-1},\nu^*_n).\end{aligned}$$

This contradicts our earlier conclusion in (52). Thus, we have (56). From (52) and (53)

$$\begin{aligned}M_{d_{n-1}}(\boldsymbol{\gamma},\boldsymbol{\nu}^*_{n-2},\nu^*_{n-1},\nu^*_n) &= \mathcal{E}^{(n-2)}(\boldsymbol{\gamma},\boldsymbol{\nu}^*_{n-2},\nu^*_{n-1},\nu^*_n)\\ &\geq \mathcal{E}^{(n-1)}(\boldsymbol{\gamma},\boldsymbol{\nu}^*_{n-1},\nu^*_n)\\ &= M_{d_n}(\boldsymbol{\gamma},\boldsymbol{\nu}^*_{n-1},\nu^*_n). \qquad (57)\end{aligned}$$

Recursively, for fixed $\nu^*_{i+1},\ldots,\nu^*_n$, we can show that $\exists \nu^*_i \in (0, T-\|\boldsymbol{\gamma}\|_1-\sum_{j=i+1}^n \nu^*_j)$, s.t. for $2 \leq i \leq n, M_{d_i}(\boldsymbol{\gamma},\boldsymbol{\nu}^*_{i-1},\nu^*_i,\ldots,\nu^*_n) = \mathcal{E}^{(i-1)}(\boldsymbol{\gamma},\boldsymbol{\nu}^*_{i-1},\nu^*_i,\ldots,\nu^*_n)$.

Let $\boldsymbol{\nu}^* = [\nu^*_1,\ldots,\nu^*_n]$, where ν^*_i is obtained as above. Then, similar to (57), we have $M_{d_i}(\boldsymbol{\gamma},\boldsymbol{\nu}^*) \geq M_{d_{i+1}}(\boldsymbol{\gamma},\boldsymbol{\nu}^*)$, for $2 \leq i \leq n-1$, and $M_{d_1}(\boldsymbol{\gamma},\boldsymbol{\nu}^*)(=\mathcal{E}^{(1)}(\boldsymbol{\gamma},\boldsymbol{\nu}^*)) = M_{d_2}(\boldsymbol{\gamma},\boldsymbol{\nu}^*)$. By Lemma 2, it is easy to see that

$$\begin{aligned}M_{d_2}(\boldsymbol{\gamma},\boldsymbol{\nu}^*) &= \min_{\boldsymbol{\nu}_2} \mathcal{E}^{(2)}(\boldsymbol{\gamma},\boldsymbol{\nu}_2,\nu^*_3,\ldots,\nu^*_n)\\ &= \mathcal{E}^{(2)}(\boldsymbol{\gamma},\boldsymbol{\nu}^*) = M_{d_3}(\boldsymbol{\gamma},\boldsymbol{\nu}^*).\end{aligned}$$

Using the same argument recursively, we finally have $M_{d_i}(\boldsymbol{\gamma},\boldsymbol{\nu}^*) = M_{d_j}(\boldsymbol{\gamma},\boldsymbol{\nu}^*)$, for $1 \leq i,j, \leq n$. By Lemma 2, it follows that under $\boldsymbol{\nu}^*$, $\mathcal{E}(\gamma,\boldsymbol{\nu}^*)$ reaches the minimum, and we have (51). ■

Proof of Theorem 2: By Lemma 3, we have the optimal $\boldsymbol{\nu}(\in \mathbf{R}^n)$ for each choice of $\boldsymbol{\gamma}$, whichis denoted as $\nu^*_{\boldsymbol{\gamma}}$. The optimization of the placement can be rewritten as

$$(\boldsymbol{\gamma}^*,\boldsymbol{\nu}^*) = \arg\min_{\boldsymbol{\gamma}} \mathcal{E}(\boldsymbol{\gamma},\boldsymbol{\nu}^*_{\boldsymbol{\gamma}}). \qquad (58)$$

Given η and T, we fix the number of pilot clusters in a period and show that the optimal placement is the one that reduces to RPP-γ. By Theorem 1, the result eventually follows.

1) Two-Cluster Case $(n=2)$: By Lemma 3, for a given $\boldsymbol{\gamma} = [\gamma_1,\gamma_2]$, there exist $\boldsymbol{\nu}^*_{\boldsymbol{\gamma}} = [\nu^*_1,\nu^*_2] \in \mathbf{R}^{2\times 1}$, such that

$$M_{d_1}(\boldsymbol{\gamma},\boldsymbol{\nu}^*_{\boldsymbol{\gamma}}) = M_{d_2}(\boldsymbol{\gamma},\boldsymbol{\nu}^*_{\boldsymbol{\gamma}}) = \min_{\boldsymbol{\nu}} \mathcal{E}(\boldsymbol{\gamma},\boldsymbol{\nu}). \qquad (59)$$

Now, we need to show that for $\boldsymbol{\mathcal{P}}^* = ([\gamma,\gamma],[\nu,\nu])$

$$M_{d_1}(\boldsymbol{\mathcal{P}}^*) = M_{d_2}(\boldsymbol{\mathcal{P}}^*) = \mathcal{E}(\boldsymbol{\mathcal{P}}^*) \leq \mathcal{E}(\boldsymbol{\gamma},\boldsymbol{\nu}^*_{\boldsymbol{\gamma}}). \qquad (60)$$

This can be shown using an argument similar to that in the graphic-aided proof of Theorem 1, and we will not elaborate here. The detail can be found in the technical report in [23].

From (60), because $\boldsymbol{\gamma}$ is arbitrary, we have

$$\mathcal{E}(\boldsymbol{\mathcal{P}}^*) = \min_{\boldsymbol{\mathcal{P}}} \mathcal{E}(\boldsymbol{\mathcal{P}}).$$

2) General n-Cluster Case: For a placement period with n pilot clusters, we use the equalizing rule in Lemma 3 and apply the result in the two-cluster case to the placement of each two consecutive pilot clusters.

Given any placement $\boldsymbol{\mathcal{P}} = (\boldsymbol{\gamma},\boldsymbol{\nu})$, we construct the following procedure. Denote $(\boldsymbol{\gamma},\boldsymbol{\nu})$ at the ith step by $(\boldsymbol{\gamma}^{(i)},\boldsymbol{\nu}^{(i)})$; then, we have the following.

1) At step i, by Lemma 3, there exists $\boldsymbol{\nu}^*_{\boldsymbol{\gamma}}$, such that, for $i = 1,\ldots,n$,

$$\mathcal{E}(\boldsymbol{\gamma}^{(i)},\boldsymbol{\nu}^*_{\boldsymbol{\gamma}}) = M_{d_i}(\boldsymbol{\gamma}^{(i)},\boldsymbol{\nu}^*_{\boldsymbol{\gamma}}) = \min_{\boldsymbol{\nu}^{(i)}} \mathcal{E}(\boldsymbol{\gamma}^{(i)},\boldsymbol{\nu}^{(i)}). \qquad (61)$$

Denote $\boldsymbol{\gamma}^{(i)}_0 = \boldsymbol{\gamma}^{(i)}, \boldsymbol{\nu}^{(i)}_0 = \boldsymbol{\nu}^*_{\boldsymbol{\gamma}}$.

2) Let $k = 1$. Using the result in the two-cluster case, we equalize each two consecutive pilot cluster lengths in order.

i) Define the "averaging" matrix $\mathbf{U}_k$ as

$$\mathbf{U}_k(i,j) = \begin{cases} 1, & i=j, \quad i \notin \{k,k+1\} \\ \frac{1}{2}, & i,j \in \{k,k+1\} \\ 0, & \text{otherwise.}\end{cases}$$

We average the lengths of the kth and $(k+1)$th pilot clusters and the same for the kth and $(k+1)$th data blocks, while keeping the lengths of the rest of the pilot and data clusters unchanged:

$$\boldsymbol{\gamma}^{(i)}_k = \boldsymbol{\gamma}^{(i)}_{k-1}\mathbf{U}_k, \quad \tilde{\boldsymbol{\nu}}^{(i)}_k = \boldsymbol{\nu}^{(i)}_{k-1}\mathbf{U}_k.$$

Then, from the two-cluster case, we know that

$$\mathcal{E}\left(\boldsymbol{\gamma}^{(i)}_k,\tilde{\boldsymbol{\nu}}^{(i)}_k\right) \leq \mathcal{E}\left(\boldsymbol{\gamma}^{(i)}_{k-1},\boldsymbol{\nu}^{(i)}_{k-1}\right).$$

ii) By Lemma 3, there exists $\boldsymbol{\nu}^*_{\boldsymbol{\gamma}}$ such that, for all $i = 1,\ldots,n$

$$\mathcal{E}(\boldsymbol{\gamma}^{(i)}_k,\boldsymbol{\nu}^*_{\boldsymbol{\gamma}}) = M_{d_i}(\boldsymbol{\gamma}^{(i)}_k,\boldsymbol{\nu}^*_{\boldsymbol{\gamma}}) = \min_{\boldsymbol{\nu}} \mathcal{E}(\boldsymbol{\gamma}^{(i)}_k,\boldsymbol{\nu}).$$

Let $\boldsymbol{\nu}^{(i)}_k = \boldsymbol{\nu}^*_{\boldsymbol{\gamma}}$. Then, we have

$$\mathcal{E}\left(\boldsymbol{\gamma}^{(i)}_k,\boldsymbol{\nu}^{(i)}_k\right) \leq \mathcal{E}\left(\boldsymbol{\gamma}^{(i)}_{k-1},\boldsymbol{\nu}^{(i)}_{k-1}\right).$$

iii) If $k < n-1$, then let $k := k+1$, and repeat i)–iii).

iv) Let $(\boldsymbol{\gamma}^{(i+1)},\boldsymbol{\nu}^{(i+1)}) = (\boldsymbol{\gamma}^{(i)}_{n-1},\boldsymbol{\nu}^{(i)}_{n-1})$. Then

$$\mathcal{E}(\boldsymbol{\gamma}^{(i+1)},\boldsymbol{\nu}^{(i+1)}) \leq \mathcal{E}(\boldsymbol{\gamma}^{(i)},\boldsymbol{\nu}^{(i)}).$$

3) Let $i := i+1$. Repeat 1)–3).

Repeat the above procedure. As $i \to \infty$, we have

$$\mathcal{E}(\lim_{i\to\infty} \boldsymbol{\gamma}^{(i)}, \lim_{i\to\infty} \boldsymbol{\nu}^{(i)}) \leq \mathcal{E}(\boldsymbol{\gamma},\boldsymbol{\nu}).$$

If we can show that

$$\lim_{i\to\infty} \boldsymbol{\gamma}^{(i)} = \boldsymbol{\gamma}^* = [\gamma,\ldots,\gamma] \qquad (62)$$

$$\lim_{i\to\infty} \boldsymbol{\nu}^{(i)} = \boldsymbol{\nu}^* = [\nu,\ldots,\nu] \qquad (63)$$

then, because $\mathcal{P} = (\boldsymbol{\gamma}, \boldsymbol{\nu})$ is arbitrary, it follows that

$$\mathcal{E}(\mathcal{P}^*) = \min_{\mathcal{P}} \mathcal{E}(\mathcal{P})$$

and we prove Theorem 2.

Now, we show that (62) is true. For a given $\boldsymbol{\gamma}^{(i)}$, define Δ_i as the maximum difference of lengths between the pilot clusters at step i.

$$\Delta_i = \max_{m \neq j, m,j=1,\ldots,n} |\boldsymbol{\gamma}^{(i)}[m] - \boldsymbol{\gamma}^{(i)}[j]|.$$

The procedure we have described in part 2) tries to even the lengths of all pilot clusters. After part 2) is finished, we have

$$\min_{1\le j\le n}\{\gamma^{(i+1)}[j]\} \ge \min_{1\le j\le n}\{\gamma^{(i)}[j]\}$$
$$\max_{1\le j\le n}\{\gamma^{(i+1)}[j]\} \le \max_{1\le j\le n}\{\gamma^{(i)}[j]\}.$$

Thus, we have

$$\Delta_i \ge \Delta_{i+1} \ge 0.$$

Because the sequence $\{\Delta_i\}$ monotonically decreases and is bounded from below, its limit exists. We now bound the decrement of Δ_i. Because there are n pilot clusters, at the ith step, there exists at least two consecutive clusters, say jth and $(j+1)$th clusters, so that

$$|\gamma^{(i)}[j] - \gamma^{(i)}[j+1]| \ge \frac{1}{n}\Delta_i.$$

After $(n-1)$ times of the pilot clusters averaging process in part 2), the decrement of the largest pilot cluster length, or the increment of the smallest pilot cluster length, is lower bounded by

$$\max\left\{\left(\max_{1\le j\le n}\gamma^{(i)}[j] - \max_{1\le j\le n}\gamma^{(i+1)}[j]\right)\left(\min_{1\le j\le n}\gamma^{(i+1)}[j] - \min_{1\le j\le n}\gamma^{(i)}[j]\right)\right\} \ge \frac{1}{2^{n-1}n}\Delta_i.$$

Then, it follows that

$$\Delta_{i+1} \le \underbrace{\left(1 - \frac{1}{2^{n-1}n}\right)}_{b}\Delta_i. \tag{64}$$

Recall that $b < 1$. Thus, $\lim_{i\to\infty}\Delta_i = 0$. The same argument applies to $\boldsymbol{\nu}^{(i)}$. Thus, we have (62) and (63). ■

REFERENCES

[1] J. K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, pp. 686–693, Nov. 1991.

[2] J. H. Lodge and M. L. Moher, "Time diversity for mobile satellite channels using trellis coded modulations," in *Proc. IEEE Global Telecommun. Conf.*, vol. 3, Tokyo, Japan, Nov. 1987.

[3] S. Sampei and T. Sunaga, "Rayleigh fading compensation method for 16 QAM in digital land mobile radio channels," in *Proc. IEEE Veh. Technol. Conf.*, San Francisco, CA, May 1989, pp. 640–646.

[4] J. M. Torrance and L. Hanzo, "Comparative study of pilot symbol assisted modem schemes," in *Proc. Sixth Int. Conf. Radio Receivers Associated Syst.*, Bath, U.K., Sept. 1995, pp. 36–41.

[5] M. K. Tsatsanis and Z. Xu, "Pilot symbol assisted modulation in frequency selective fading wireless channels," *IEEE Trans. Signal Processing*, vol. 7, pp. 2353–2365, Aug. 2000.

[6] M. Medard, I. Abou-Faycal, and U. Madhow, "Adaptive coding with pilot signals," in *Proc. 38th Allerton Conf.*, Oct. 2000.

[7] S. Ohno and G. B. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 712–720, Oct. 2002.

[8] J. A. Gansman, M. P. Fitz, and J. V. Krogmeier, "Optimum and suboptimum frame synchronization for pilot-symbol-assisted modulation," *IEEE Trans. Commun.*, vol. 45, pp. 1327–1337, Oct. 1997.

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

[10] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Processing*, vol. 50, pp. 3055–3069, Dec. 2002.

[11] C. Budianu and L. Tong, "Channel estimation for space-time block coding systems," *IEEE Trans. Signal Processing*, vol. 50, pp. 2515–2528, Oct. 2002.

[12] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of known symbols for frequency-selective block-fading channels," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2338–2353, Aug. 2002.

[13] S. Ohno and G. B. Giannakis, "Capacity maximizing plots and precoders for wireless OFDM over rapidly fading channels," *IEEE Trans. Inform. Theory*, to be published.

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

[15] M. Dong, S. Adireddy, and L. Tong, "Optimal pilot placement for semi-blind channel tracking of packetized transmission over time-varying channels," *IEICE Trans. Fundamentals Electron., Commun., Comput. Sci.*, vol. E86-A, pp. 550–563, Mar. 2003.

[16] A. Gaston, W. Chriss, and E. Walker, "A multipath fading simulator for radio," *IEEE Trans. Veh. Technol.*, vol. VT-22, pp. 241–244, 1973.

[17] R. Iltis, "Joint estimation of PN code delay and multipath using the extended kalman filter," *IEEE Trans. Commun.*, vol. 38, pp. 1677–1685, Oct. 1990.

[18] M. Stojanovic, J. Proakis, and J. Catipovic, "Analysis of the impact of channel estimation errors on the performance of a decision-feedback equalizer in fading multipath channels," *IEEE Trans. Commun.*, vol. 43, pp. 877–886, Feb. 1995.

[19] M. Medard, "The effect upon channel capacity in wireless communication of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, pp. 933–946, May 2000.

[20] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[21] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 2000.

[22] J. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[23] M. Dong, L. Tong, and B. Sadler, Optimal placement of training over time-varying channels, http://acsp.ece.cornell.edu/pubR.html, ACSP TR-01-03-02, Jan. 2003.

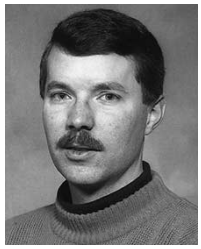
Min Dong (S'00) received the B.Eng. degree from the Department of Automation, Tsinghua University, Beijing, China, in 1998. She is now pursuing the Ph.D. degree at the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY.

Her research interests include statistical signal processing, wireless communications, and communication 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.

He was a Postdoctoral Research Affiliate at the Information Systems Laboratory, Stanford University, Stanford, CA, in 1991. Currently, he is an Associate Professor with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY. His areas of interest include statistical signal processing, adaptive receiver design for communication systems, signal processing for communication networks, and information theory.

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.



Brian M. Sadler (M'90) received the B.S. and M.S. degrees from the University of Maryland, College Park, in 1981 and 1984, respectively, and the Ph.D. degree from the University of Virginia, Charlottesville, in 1993, all in electrical engineering.

He is a senior research scientist at the Army Research Laboratory (ARL), Adelphi, MD. He was a lecturer at the University of Maryland from 1985 to 1987 and has been lecturing at Johns Hopkins University, Baltimore, MD, on statistical signal processing and communications since 1994. His research interests generally include statistical signal processing with applications in communications, radar, and aeroacoustics.

Dr. Sadler was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and is on the editorial board for the *EURASIP Journal on Wireless Communications and Networking*. He is a member of the IEEE Technical Committee on Signal Processing for Communications and co-chaired the Second IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC-99).