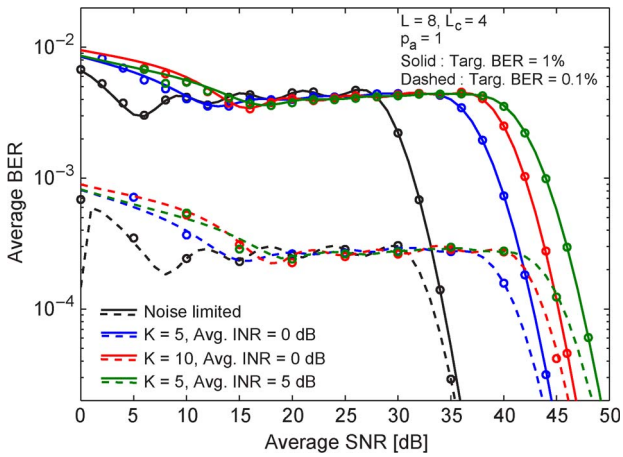


Fig. 5. Average BER against $\bar{\gamma}_D$ for the power-efficient scheme using different combinations of L_I, $\bar{\gamma}_I$, and target BER.

the average BER with $\bar{\gamma}_D$ increases, which explains the shift of these points to the right in the presence of cochannel interference.

VI. Conclusion

This paper has investigated the effect of cochannel interference on the performance of adaptive SNR-based MS-GSC in multipath Rayleigh fading. Specifically, in the absence of perfect knowledge of the instantaneous powers associated with the interfering signals, the adaptation thresholds have been set on the combined SNR to enhance either the spectral efficiency or power efficiency of discrete-time rectangular constellation system in interference-free environments. Analytical formulations for some important performance measures of the two adaptation schemes have been presented with the help of new expressions of the statistics of the combined SINR obtained herein. These formulations are applicable for any diversity order and number of interfering signals and can be used to study the effect of cochannel interference on the performance of various combining schemes, including the conventional SC, MRC, and GSC. The performance degradation of the two adaptation schemes due to cochannel interference has been clarified via several numerical examples.

References

[1] G. L. Stüber, *Principles of Mobile Communications*. Norwood, MA: Kluwer, 1996.
[2] W. C. Lee, *Mobile Communications: Design Fundamentals*, 2nd ed. New York: Wiley, 1993.
[3] M. K. Simon and M.-S. Alouini, *Digital Communications Over Fading Channels*, 2nd ed. Hoboken, NJ: Wiley, 2005.
[4] S. W. Kim, D. S. Ha, and J. H. Reed, "Minimum selection GSC and adaptive low-power RAKE combining scheme," in *Proc. IEEE ISCAS*, 2003, vol. 4, pp. 357–360.
[5] R. K. Mallik, P. Gupta, and Q. T. Zhang, "Minimum selection GSC in independent Rayleigh fading," *IEEE Trans. Veh. Technol.*, vol. 54, no. 3, pp. 1013–1021, May 2005.
[6] H.-C. Yang, "New results on ordered statistics and analysis of the minimum-selection generalized selection combining (GSC)," *IEEE Trans. Wireless Commun.*, vol. 5, no. 7, pp. 1876–1885, Jul. 2006.
[7] V. Hentinen, "Error performance for adaptive transmission on fading channels," *IEEE Trans. Commun.*, vol. COM-22, no. 9, pp. 1331–1337, Sep. 1974.
[8] H.-C. Yang, N. Belhaj, and M.-S. Alouini, "Performance analysis of joint adaptive modulation and diversity combining over fading channels," *IEEE Trans. Commun.*, vol. 55, no. 3, pp. 520–528, Mar. 2007.
[9] M.-S. Alouini and A. J. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *IEEE Trans. Veh. Technol.*, vol. 48, no. 4, pp. 1165–1181, Jul. 1999.
[10] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[11] R. M. Radaydeh, "Performance of non-ideal OT-MRC with co-channel interference," *IEEE Trans. Commun.*, vol. 58, no. 12, pp. 3352–3357, Dec. 2010.
[12] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products*, 5th ed. San Diego, CA: Academic, 1994.
[13] R. M. Radaydeh, "Performance of cellular mobile systems employing SNR-based GSC in the presence of Rayleigh and Nakagami-q cochannel interferers," *IEEE Trans. Veh. Technol.*, vol. 58, no. 6, pp. 3081–3088, Jul. 2009.
[14] R. M. Radaydeh and M.-S. Alouini, "Impact of co-channel interference on the performance of adaptive non-ideal generalized transmit diversity," in *Proc. 7th ISWCS*, York, U.K., 2010, pp. 280–284.

User-Specified Training Symbol Placement for Channel Prediction in TDD MIMO Systems

Shengqian Han, *Student Member, IEEE*, Yafei Tian, *Member, IEEE*, and Chenyang Yang, *Senior Member, IEEE*

***Abstract*—In this paper, we design the training symbol placement for channel prediction in time-division-duplex multiple-antenna systems, where training symbols in a finite-length uplink frame are used for predicting downlink channels at the base station (BS). Aimed at minimizing the normalized sum mean square error of the Wiener predictor, we first prove that, for the first-order Gauss–Markov fading channel model, the optimal positions of training symbols lie at the end of the uplink frame, which is commonly recognized by intuition. For general channel models, we show that the ending placement is no longer optimal, and we propose a low-complexity method for designing training symbol placement based on alternating searching. In practice, the BS finds the optimal positions based on each user's temporal correlation, spatial correlation, and uplink signal-to-noise ratio (SNR) and then transmits the positions to the user, which needs low signaling overhead. Numerical results show significant prediction performance gain of the proposed training symbol placement over typical training placements under various SNR and Doppler frequency.**

***Index Terms*—Channel prediction, ending placement, training symbol placement, uniform placement.**

I. Introduction

Channel state information at transmitter (CSIT) plays a crucial role in various adaptive transmission strategies and multiple-antenna systems, particularly for multiuser multiple-input–multiple-output (MIMO) systems [1], [2]. In time-varying wireless channels, outdated CSIT will severely degrade the performance of adaptive transmission systems and closed-loop MIMO systems. Channel prediction is capable of providing up-to-date channel information, which has been well investigated for both single-input–single-output (SISO) systems and

Manuscript received March 1, 2010; revised February 17, 2011; accepted April 12, 2011. Date of publication May 5, 2011; date of current version July 18, 2011. This work was supported in part by the International S&T Cooperation Program of China under Grant 2008DFA12100 and in part by the National Key Project of Next-Generation Wideband Wireless Communication Networks under Grant 2009ZX03003-005-01. The review of this paper was coordinated by Prof. H. H. Nguyen.

The authors are with the School of Electronics and Information Engineering, Beihang University, Beijing 100191, China (e-mail: sqhan@ee.buaa.edu.cn; ytian@buaa.edu.cn; cyyang@buaa.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.
Digital Object Identifier 10.1109/TVT.2011.2151216

MIMO systems (see, e.g., [1], [3]–[5], and the references therein). It is shown that a substantially longer prediction horizon can be achieved for MIMO channels than for SISO channels [5]. By exploiting the spatial correlation in MIMO channels, a space-time channel predictor is proposed in [6], which reduces the prediction errors led by noises.

The performance of channel prediction depends on the structure of training signals, which is analogous to channel estimation. The training design for optimizing channel estimation in time-varying channels has been extensively studied in the past years [7]–[12]. The training signal structure includes the number, power, and placement of training symbols, whose optimal values are largely dependent on the optimality criterion, *a priori* information, and the channel model. For open-loop systems, it is well known that uniformly inserted training symbols are optimal [7], [8] in a sense of minimizing mean square error (MSE). However, they are no longer optimal for closed-loop training if *a priori* knowledge of channels is available. In [9] and [10], numerical methods have been proposed to search the optimal placement of training symbols. In [11] and [12], the placement with equally spaced clusters, each with multiple training symbols, is optimized, which outperforms the uniform placement. However, an optimal training signal for channel estimation is not necessarily optimal for channel prediction, because they allow different training structures and exploit different statistics of received signals and channels to achieve different goals. In contrast to channel estimation where training symbols are multiplexed into a data frame, the training symbols for channel prediction to assist downlink transmission are separated from data symbols, i.e., they are placed in the uplink and downlink frames, respectively. Except for filtering noise, channel prediction needs to extrapolate the future channels, rather than only reducing the aliasing distortion of channel interpolation and noises as in channel estimation.

Although important, the training design for channel prediction receives relatively little attention. Assuming a first-order Gauss-Markov flat-fading channel model and a periodic training structure, it is shown in [13] that the training symbols consecutively inserted within each period are optimal for minimizing the maximal MSE or bit error rate, and the power allocation between training and data symbols is optimized in [14] for maximizing the lower bound of capacity. These results rely on the assumption that a sufficiently large number of training symbols are available. However, this is often not valid for most future wireless systems, where fast scheduling will prevent each user from keeping active for a long time. In this paper, we consider time-division-duplex (TDD) MIMO systems, where the training symbols inserted in a finite-length uplink frame are employed to predict downlink channels at the base station (BS)[1]. In this context, no prior work exists to optimize the training placement for channel prediction to the best of our knowledge.

The focus of this paper is to design the user-specified training symbol placement in the uplink frame by exploiting each user's spatial correlation, Doppler spectrum (or equivalently temporal correlation), and uplink average signal-to-noise ratio (SNR). Various methods have been proposed to obtain these statistics, e.g., [15]–[18] and the references therein. Intuitively, in TDD systems, the training symbols for channel prediction should be in the end of the uplink frame to minimize the outdating. We will prove that such a placement is optimal in a sense of minimizing the normalized sum MSE of the Wiener predictor if the channels evolve in time, following the first-order Gauss–Markov flat-fading channel model. For general channel models, we show that this is no longer true through asymptotic analysis. We find that the optimal placement is nonuniform, which provides significant performance gain over the ending placement. This motivates us to propose a low-complexity method to select the positions of training symbols.

[1]These training symbols are dedicated to channel prediction for downlink transmission, i.e., for channel sounding, rather than channel estimation for uplink data demodulation. In practical systems, considering the unbalanced uplink and downlink traffic loads, usually, different time-frequency resources are employed for the two kinds of training signals with different purposes.

Notations: $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and the conjugate transpose, respectively. $\mathbb{E}\{\mathbf{A}\}$ denotes the expectation with respect to a random matrix $\mathbf{A}$, $\mathrm{tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$, $\mathrm{vec}(\mathbf{A})$ stacks matrix $\mathbf{A}$ into a vector columnwise, and $\mathbf{A}^{1/2}$ denotes the Hermitian square root of $\mathbf{A}$. $[\mathbf{A}]_{ij}$ represents the element at the ith row and jth column of matrix $\mathbf{A}$, and $[\mathbf{a}]_i$ represents the ith element of the vector $\mathbf{a}$. $\|\mathbf{a}\|$ denotes the 2-norm of vector $\mathbf{a}$, and $\mathbf{I}$ denotes the identity matrix. Finally, $\otimes$ denotes the Kronecker product.

II. System and Channel Models

Consider a TDD system equipped with M transmit antennas at the BS and N receive antennas at the user side. Assume that the uplink training signals for multiple users are separated in an orthogonal frequency-division multiple-access fashion; thus, we only consider the channel of a single user in the following. The frame structure is shown in Fig. 1(a). L training symbols are located in the uplink frame with position set $\mathcal{S} = \{p_1, \ldots, p_L\}$. Define $\mathcal{T} = \{0, \ldots, K_u - 1\}$ as the position pool of training symbols; then, $\mathcal{S}$ is a subset of $\mathcal{T}$.

We model a flat Rayleigh fading MIMO channel as [19]

$$\mathbf{H}(t) = \mathbf{R}_R^{1/2}\mathbf{H}_w(t)\mathbf{R}_T^{1/2} \tag{1}$$

with receive spatial correlation matrix $\mathbf{R}_R$ and transmit spatial correlation matrix $\mathbf{R}_T$. Matrix $\mathbf{H}_w(t)$ consists of independent and identically distributed (*i.i.d.*) complex Gaussian variables with zero mean and unit variance. The temporal correlation function of $\mathbf{H}_w(t)$ is

$$\mathbb{E}\left\{\mathbf{h}_w(t)\mathbf{h}_w^H(t+\tau)\right\} = R_t(\tau)\mathbf{I} \tag{2}$$

where $\mathbf{h}_w = \mathrm{vec}(\mathbf{H}_w)$, and $R_t(\tau) = \mathbb{E}\{[\mathbf{h}_w]_i(t)[\mathbf{h}_w]_i^H(t+\tau)\}$, $i = 1, \ldots, MN$. Then, the spatial-temporal correlation function of $\mathbf{H}(t)$ can be expressed as

$$\mathbb{E}\left\{\mathbf{h}(t)\mathbf{h}^H(t+\tau)\right\} = R_t(\tau)\mathbf{R}_C \tag{3}$$

where $\mathbf{h} = \mathrm{vec}(\mathbf{H})$, and $\mathbf{R}_C = \mathbf{R}_T^T \otimes \mathbf{R}_R$.

Assume that the channel is time invariant within each symbol duration; then, the lth received training symbol at the BS, $l = 1, \ldots, L$, is

$$\mathbf{Y}(t - p_l T_s) = \mathbf{H}^T(t - p_l T_s)\mathbf{X} + \mathbf{Z}_l \tag{4}$$

where $\mathbf{X}$ comprises properly designed training sequences satisfying $\mathbf{X}\mathbf{X}^H = \mathbf{R}_X$ by taking into account the channel correlation, $\mathbf{Z}_l$ is the additive white Gaussian noise consisting of entries with zero mean and variance σ^2, and T_s is the symbol duration. To simplify the notation, the normalized transmit power at the user side is considered, so that the uplink SNR is defined as $1/\sigma^2$. Let $\mathbf{y} = \mathrm{vec}\{\mathbf{Y}^T\}$; then, the collection of the L received training symbols, i.e., L observations, can be rewritten in a columnwise form as

$$\bar{\mathbf{y}} = \left[\mathbf{y}^T(t - p_1 T_s), \ldots, \mathbf{y}^T(t - p_L T_s)\right]^T. \tag{5}$$

The BS can predict downlink channels with uplink observations when channel statistics including spatial correlation matrix, Doppler spectrum, and uplink SNR are available. As commonly assumed for

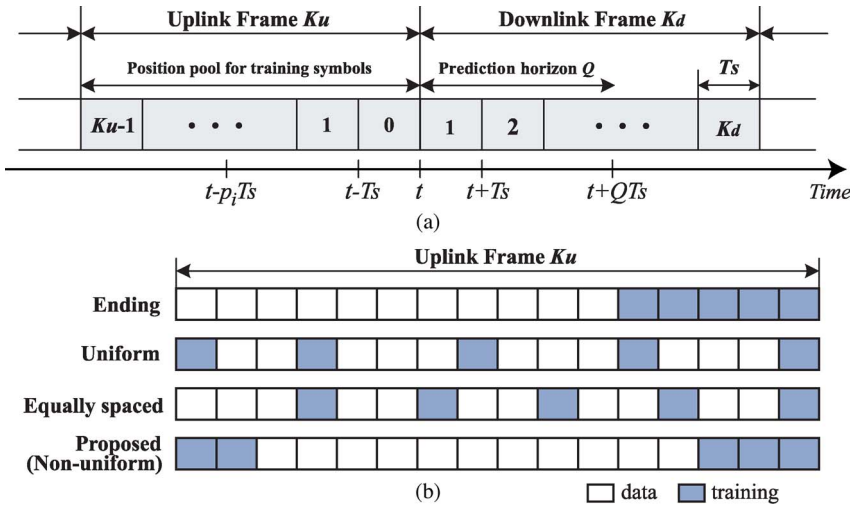


Fig. 1. (a) Frame structure of the considered TDD system. The uplink frame consists of K_u symbols, among which L training symbols with the position set $\mathcal{S} = \{p_1, \ldots, p_L\}$ (relative to the beginning of the downlink frame) are included. The first Q data symbols in the downlink frame are used for closed-loop transmission. The time interval between the ith training symbol and the Qth symbol in the downlink frame is $(p_i + Q)T_s$. (b) Examples of four considered training symbol placements, where five training symbols are marked from 16 uplink symbols. The equally spaced placement and the proposed placement are obtained when Doppler frequency $f_d = 200$ Hz, $Q = 6$, and SNR $= 10$ dB.

channel estimation and prediction, e.g., as in [11]–[14], we assume that *a priori* knowledge of these channel statistics is known. In practice, the statistics can be estimated with the methods proposed in [15]–[18]. We also assume that the designed positions can be conveyed to the user without errors. Based on the minimum MSE (MMSE) criterion, the channels at the Qth symbol in the downlink frame are predicted at the BS with the Wiener predictor as

$$\tilde{\mathbf{h}}(t + QT_s) = \mathbf{R}_{hy}\mathbf{R}_y^{-1}\bar{\mathbf{y}} \tag{6}$$

where $\mathbf{R}_y = (\mathbf{I} \otimes \mathbf{X}^T)(\mathbf{\Gamma}(\mathcal{S}) \otimes \mathbf{R}_C)(\mathbf{I} \otimes \mathbf{X}^T)^H + \sigma^2\mathbf{I}$, $\mathbf{R}_{hy} = (\mathbf{r}^H(\mathcal{S}) \otimes \mathbf{R}_C)(\mathbf{I} \otimes \mathbf{X}^T)^H$, $\mathbf{\Gamma}(\mathcal{S})$ is the temporal autocorrelation matrix of the observations with entries $[\mathbf{\Gamma}(\mathcal{S})]_{ij} = R_t(|p_i - p_j|T_s)$, and $\mathbf{r}(\mathcal{S})$ is the vector of temporal cross correlation between the observations and the channels at the position to be predicted, whose entries are $[\mathbf{r}(\mathcal{S})]_i = R_t((p_i + Q)T_s)$, $1 \le i, j \le L$.

By applying the matrix inversion lemma, the normalized sum MSE of the Wiener predictor can be obtained as

$$\begin{aligned}\epsilon(\mathcal{S}) &= \frac{\mathbb{E}\left\{\left\|\tilde{\mathbf{h}}(t + QT_s) - \mathbf{h}(t + QT_s)\right\|^2\right\}}{\mathbb{E}\left\{\|\mathbf{h}(t + QT_s)\|^2\right\}} \\ &= 1 - \frac{1}{\mathrm{tr}(\mathbf{R}_C)}\mathrm{tr} \\ &\quad \times \Big(\left(\mathbf{r}^H(\mathcal{S}) \otimes \mathbf{R}_C\right)\left(\mathbf{\Gamma}(\mathcal{S}) \otimes \mathbf{R}_C + \sigma^2(\mathbf{I} \otimes \mathbf{R}_X^{-1})\right)^{-1} \\ &\quad \times (\mathbf{r}(\mathcal{S}) \otimes \mathbf{R}_C)\Big).\end{aligned} \tag{7}$$

III. User-Specified Training Design

In this section, we first give an asymptotic analysis of the MSE for two extreme SNR cases to understand the impact of training symbol placement on channel prediction. We then show that, for the first-order Gauss–Markov channel model, the optimal positions of training symbols are located at the end of the uplink frame to minimize the normalized sum MSE. For general time-varying channels, we propose a low-complexity method to find the positions of training symbols.

A. Asymptotic Analysis

For very low SNR, the normalized sum MSE shown in (7) approximately becomes

$$\epsilon(\mathcal{S}) \approx 1 - \frac{\mathrm{tr}\left(\mathbf{R}_C(\mathbf{I} \otimes \mathbf{R}_X)\mathbf{R}_C\right)}{\sigma^2\mathrm{tr}(\mathbf{R}_C)}\mathbf{r}^H(\mathcal{S})\mathbf{r}(\mathcal{S}). \tag{8}$$

To minimize the normalized sum MSE, (8) suggests that the observations should have high cross correlation with the channels at the position to be predicted. Moreover, the spatial correlation is exploited to enhance the SNR, which reduces the variance of noise from σ^2 to $(\mathrm{tr}(\mathbf{R}_C)/\mathrm{tr}(\mathbf{R}_C(\mathbf{I} \otimes \mathbf{R}_X)\mathbf{R}_C))\sigma^2$.

For very high SNR, we have

$$\epsilon(\mathcal{S}) \approx 1 - \mathbf{r}^H(\mathcal{S})\mathbf{\Gamma}^{-1}(\mathcal{S})\mathbf{r}(\mathcal{S}). \tag{9}$$

In this case, the impact of observation noises is negligible. We only need to reduce the channel prediction errors led by channel variation; thereby, the optimal training symbols should provide the maximal amount of information of the channels to be predicted. Again, a high cross correlation $\mathbf{r}(\mathcal{S})$ is beneficial. In addition, a low autocorrelation $\mathbf{\Gamma}(\mathcal{S})$ between the observations can provide more information. Owing to this, the optimal placement of training symbols is neither at the end of the uplink frame nor uniform in general, as will be shown later.

B. Training Symbol Placement Design

Now, we consider the problem of designing optimal training symbol placement aimed at minimizing the normalized sum MSE of channel prediction, which can be formulated as

$$\begin{aligned}\min_{\mathcal{S}} \quad & \epsilon(\mathcal{S}) \\ \text{s.t.} \quad & \mathcal{S} \subset \mathcal{T}, \qquad |\mathcal{S}| = L\end{aligned} \tag{10}$$

where $|\mathcal{S}|$ denotes the size of $\mathcal{S}$.

1) Optimal Training Placement Under Gauss–Markov Channel Model: We can prove the following proposition (see Appendix) for this special channel model:

Proposition 1: When the channels evolve in time, following the first-order Gauss–Markov model [12]–[14], the optimal positions of training symbols for channel prediction lie at the end of the uplink frame. For high SNR, a single training symbol is sufficient to minimize the normalized sum MSE of channel prediction.

Remark 1: For general channel models, the optimal training symbol placement is not at the end of the uplink frame, as is implied by the asymptotic analysis and will be shown later. This is in contrast to the intuition that the training symbols should be placed as close as possible to the channel to be predicted to minimize the outdating.

2) Optimal Training Placement Under General Channel Models: The optimization problem (10) is essentially a binary integer programming problem, and its objective function is not convex on the positions of training symbols. We cannot obtain a closed-form solution to (10) in general cases, but we can find the optimal solution via exhaustive searching over $\binom{K_u}{L}$ possible position sets. This leads to a high computational complexity that might exceed the requirements of online design, particularly for large K_u. Alternating searching is an effective method for providing a near-optimal solution while reducing the searching space by converting a multidimensional searching problem into multiple 1-D searching problems [20]. Bearing this spirit, we propose a low-complexity suboptimal solution, which alternatingly updates the initial training symbol positions until a specific condition is satisfied.

The proposed alternating searching algorithm is given as follows:

1) *Successive initialization:* initialize $\mathcal{S}$ by selecting $p_1, \ldots, p_L$ with the following procedure:

   Let $\mathcal{T}_l$ and $\mathcal{S}_l$ denote the position pool and the selected results at the lth step, $1 \le l \le L$. Set $\mathcal{S}_0 = \phi$ (empty set) and $\mathcal{T}_0 = \{0, \ldots, K_u - 1\}$.

   Then, for $l = 1, \ldots, L$, successively select position p_l from $\mathcal{T}_{l-1}$ based on $\mathcal{S}_{l-1}$ as

$$p_l = \arg\min_{i \in \mathcal{T}_{l-1}} \epsilon\left(\mathcal{S}_{l-1} \cup \{i\}\right). \tag{11}$$

   At each step, update $\mathcal{T}_l$ and $\mathcal{S}_l$, respectively, as $\mathcal{T}_l = \{i \in \mathcal{T}_{l-1}, i \neq p_l\}$ and $\mathcal{S}_l = \mathcal{S}_{l-1} \cup \{p_l\}$.

2) *Alternating update:* iteratively update the already selected positions $p_1, \ldots, p_L$ according to the following alternating searching method:
   a) Let $k = 0$, and calculate the normalized sum MSE ϵ_0 corresponding to position set $\mathcal{S}$ obtained in step 1.
   b) Set $k = k + 1$, and update $\mathcal{S}$ in L steps. At the lth step for $l = 1, \ldots, L$, update $\mathcal{S}$ by replacing p_l as

$$p_l = \arg\min_{i \in \mathcal{T}_{kl}} \epsilon\left(\mathcal{S}_{kl} \cup \{i\}\right) \tag{12}$$

      where $\mathcal{S}_{kl} = \{i \in \mathcal{S}, i \neq p_l\}$, and $\mathcal{T}_{kl} = \{i \in \mathcal{T}_0, i \notin \mathcal{S}_{kl}\}$. After L steps, calculate $\epsilon_k = \epsilon(\mathcal{S})$.
   c) Repeat step b until $\epsilon_{k-1} - \epsilon_k \le \delta$, where δ is a specific threshold.

The algorithm will converge because the normalized sum MSE is monotonically decreasing after each iteration and is lower bounded. The iteration is terminated when the required accuracy or the maximum number of iterations is reached. Although it is not guaranteed that the alternating searching algorithm always returns the global optimum due to the nonconvexity of the normalized sum MSE over the training symbol positions, the sequel numerical analysis shows that it exhibits fast convergence and is able to achieve the same performance as exhaustive searching in the considered scenarios.

TABLE I
COMPARISON OF COMPUTATIONAL COMPLEXITY

Algorithms	Complex Multiplications	Example*
Exhaustive	$\mathcal{O}\left(\binom{K_u}{L}(LMN)^3\right)$	$\mathcal{O}(147292160)$
Alternating	$\mathcal{O}\left((k+1)(K_u-L)L(LMN)^3\right)$	$\mathcal{O}(917504)$

* The parameters in Fig. 7 are used in the example, i.e., $\{K_u, L, M, N, k\} = \{32, 4, 2, 2, 1\}$.

The computational complexity of the proposed algorithm and exhaustive searching is summarized in Table I. The most complex operation of the two algorithms is to compute the inverse of an $LMN \times LMN$ matrix to obtain the normalized sum MSE, which leads to a complexity in an order of $\mathcal{O}((LMN)^3)$ in terms of complex multiplications. In exhaustive searching, the matrix inverse needs to be repeatedly computed for all $\binom{K_u}{L}$ possible position sets. In the proposed algorithm, $(k+1)(K_u - L)L$ times of matrix inverse operation are required, where k denotes the iteration number of the alternating update and the successive initialization is approximately regarded as an additional iteration. Because of the fast convergence, i.e., k is a small number, the proposed algorithm has a much lower complexity than the optimal exhaustive searching, as shown by the example in Table I.

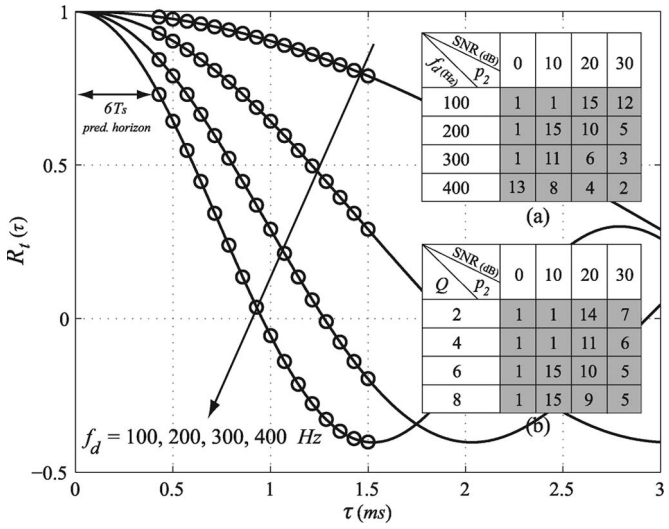
Remark 2: In practice, the proposed method can be implemented as follows to reduce both the computational burden of user and the signaling overhead. The BS finds the positions of training symbols with the proposed method for each user according to its spatial and temporal correlation, and uplink average SNR. The designed position set of training symbols is sent to the user in the downlink frame. Then, the user can transmit training symbols at these positions in the concatenated uplink frame, which are used for channel prediction at the BS. The signaling overhead is associated with the position set sent from the BS to the user, which is selected from $\binom{K_u}{L}$ possible position sets. Hence, each position set conveys $\log_2 \binom{K_u}{L}$ bits of information, which need to be sent to the user in the downlink frame. The overhead is acceptable for small K_u. For the case of very large K_u, the vector quantization method proposed in [9] can be applied to reduce the overhead, which, however, is beyond the scope of this paper. Since the training symbol placement only depends on the channel statistics of the user, which slowly vary, it is unnecessary to be updated for each uplink frame.

Remark 3: The training symbol placement designed here only targets for a specific prediction horizon. In practice, we may need to predict channels over a large span of multiple downlink data symbols. Since the prediction for the last symbol used for closed-form transmission in the downlink frame generally leads to the maximum MSE, we can design the positions of training symbols for the largest prediction horizon. We will evaluate the performance of such a "worst-case" strategy in the next section.

Remark 4: As shown in (3), we assume that the temporal correlation is independent from the spatial correlation. This means that the channel's spatial correlation is only exploited to reduce the noise as in [6]. Nonetheless, the proposed training design method is applicable to more general spatially temporally correlated channel models.

IV. NUMERICAL RESULTS

In this section, we demonstrate the performance of the proposed training symbol placement using alternating searching. For



f_d (Hz) \ SNR (dB), p_2	0	10	20	30
100	1	1	15	12
200	1	15	10	5
300	1	11	6	3
400	13	8	4	2

Q \ SNR (dB), p_2	0	10	20	30
2	1	1	14	7
4	1	1	11	6
6	1	15	10	5
8	1	15	9	5

Fig. 2. Temporal correlation between the observations and the channels to be predicted with different f_d's, and training symbols' positions as a function of SNR, f_d, and prediction horizon Q. $L = 2$ training symbols are inserted in the uplink frame, whose positions are obtained by exhaustive searching. Results show that $p_1 = 0$ (at the end of the uplink frame) in all scenarios. p_2 is presented in tables (a) and (b), where $Q = 6$ and $f_d = 200$ Hz are respectively considered.

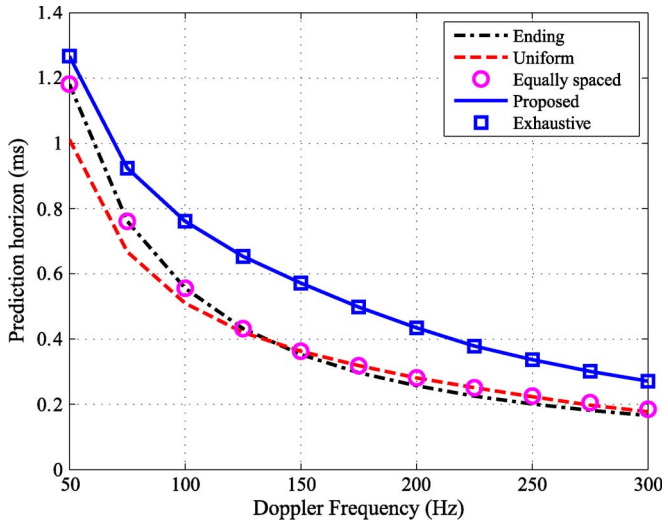


Fig. 3. Prediction horizon of the five training symbol placements versus Doppler frequency. Similar to [5], the prediction horizon is defined as the maximum prediction length with a normalized sum MSE of less than 0.1, i.e., $\epsilon(\mathcal{S}) < 0.1$. SNR = 10 dB and $L = 4$.

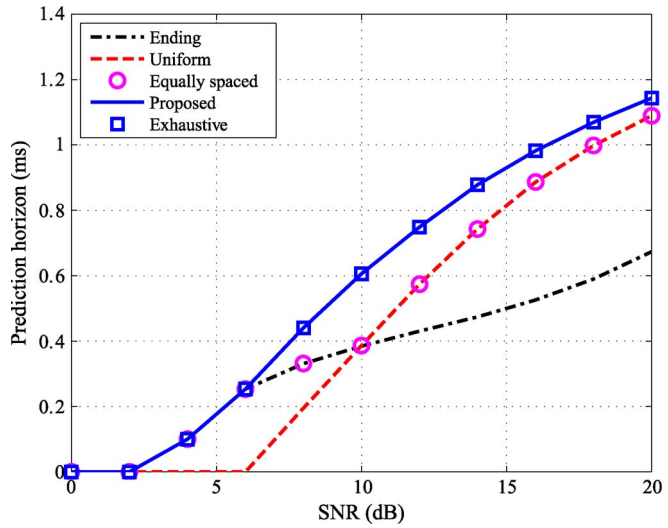


Fig. 4. Prediction horizon of the five training symbol placements versus SNR. The prediction horizon is defined as the maximum prediction length satisfying $\epsilon(\mathcal{S}) < 0.1$. A typical operation frequency of 2 GHz is considered. The velocity of the mobile user is 75 km/h, i.e., $f_d = 139$ Hz, and $L = 4$.

comparison, three relevant placements are also considered: uniform placement that is usually applied for channel estimation [7] and channel prediction [3], ending placement that places training symbols at the end of the uplink frame, and an equally spaced placement with $p_i = (i-1)d$ with numerically optimized interval d, which can balance the autocorrelation and cross correlation between observations and the channel to be predicted. These four considered placements are shown in Fig. 1(b). To see if the proposed suboptimal method performs close to the optimal solution of (10), the performance with optimal placement is also shown, which is obtained via exhaustive searching.

We consider a TDD system with two transmit antennas at the BS and two receive antennas at the user side. The spatial correlation is generated according to an exponential model [21], with the correlation coefficient of $\rho = 0.5$ for both transmit and receive antennas. The antenna-specific orthogonal training sequences are employed within each training symbol, i.e., $\mathbf{R}_X = \mathbf{I}$. We consider the time-varying channels based on Jakes' model [22], with the temporal correlation function $R_t(\tau) = J_0(2\pi f_d \tau)$, where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind and f_d is the Doppler frequency. It has been shown that the widely accepted channel models, including SCM, SCME, and WIM developed by 3GPP/3GPP2 and WINNER, have very similar temporal correlation functions with Jakes' model [23]. Unless otherwise specified, the uplink frame includes 16 symbols, from which four training symbols are selected, and the symbol duration is 71.4 μs [24].

Fig. 2 shows the temporal correlation between the observations and the channels to be predicted, as well as two tables presenting the training symbols' positions as a function of Doppler frequency, SNR, and prediction horizon. Considering $L = 2$ training symbols inserted in the uplink frame, we obtain their optimal positions p_1 and p_2 by exhaustive searching. We find that $p_1 = 0$ in all scenarios and show p_2 in the tables. It can be seen that the optimal positions are at the end of the uplink frame ($p_2 = 1$) when observation noises dominate the prediction errors, e.g., the cases with low SNR, small Doppler frequency, or short prediction horizon. In general scenarios, the ending placement is no longer optimal, even when all the observations are located within the main lobe of the temporal correlation function[2], e.g., when $f_d = 100$ Hz and 200 Hz.

Figs. 3 and 4 show the prediction horizon of the five training symbol placements versus Doppler frequency and SNR. The threshold used in the proposed method is $\delta = 10^{-3}$. The proposed training symbol placement with alternating searching has the same performance as the optimal solution, which provides the longest prediction horizon. Although the ending placement is optimal for the first-order Gauss–Markov model, its performance degrades under Jakes' model. For small Doppler frequency or low SNR, the prediction

[2]The temporal correlation function of Jake's channel model monotonically decreases in the main lobe, similar to that of the first-order Gauss-Markov channel model, but the latter channel model has a unique feature that is essential to ensure the optimality of the ending placement.

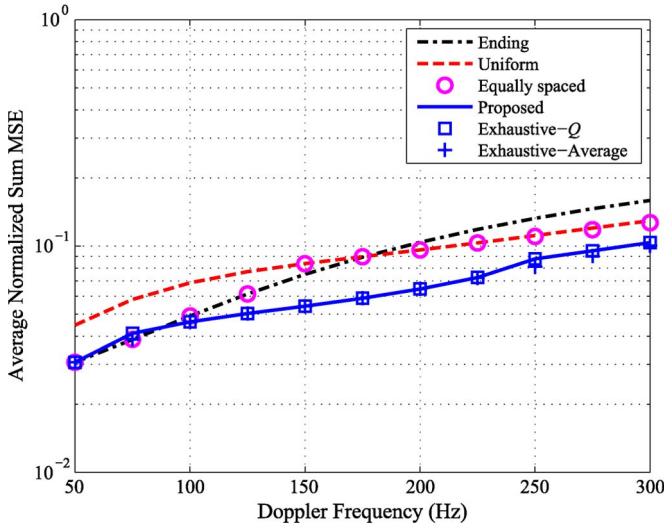


Fig. 5. Average normalized sum MSE over Q data symbols in the downlink frame versus Doppler frequency, $Q = 6$, $L = 4$, and SNR $= 10$ dB.

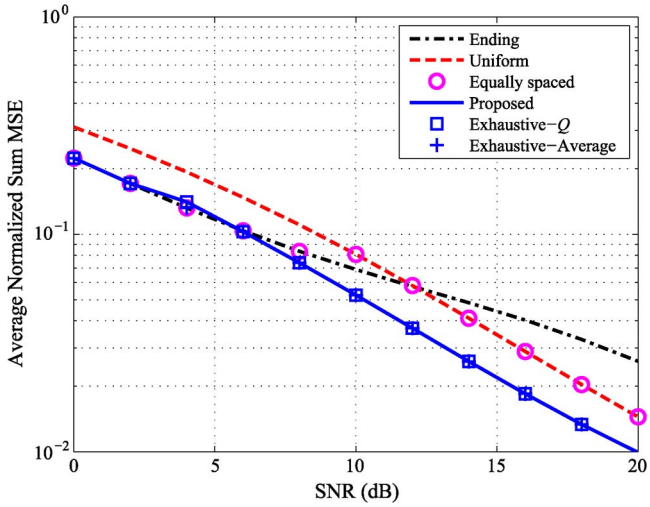


Fig. 6. Average normalized sum MSE over Q data symbols in the downlink frame versus SNR, $Q = 6$, $L = 4$, and $f_d = 139$ Hz.

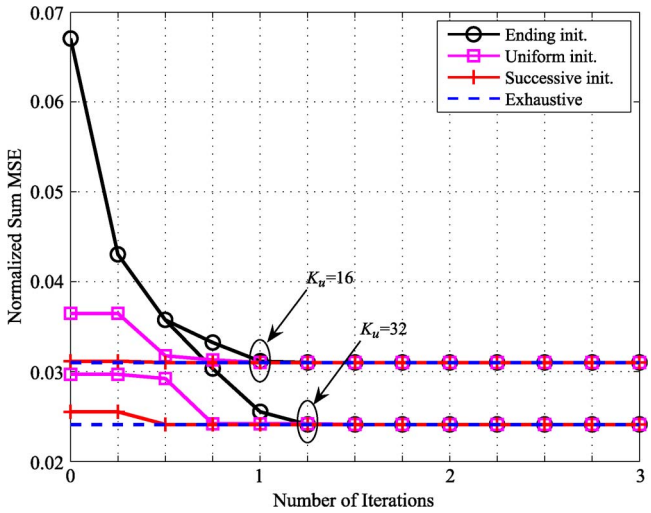


Fig. 7. Normalized sum MSE for the Qth data symbol in the downlink frame versus iteration number of the alternating searching algorithm with different initial values, $K_u = 16$ and 32, $Q = 6$, $L = 4$, SNR $= 20$ dB, and $f_d = 200$ Hz. Note that $L = 4$ steps are required to update the L positions within each iteration.

performance is dominated by noise. In this case, the ending placement can effectively suppress the noise by exploiting the high temporal autocorrelation between adjacent observations and hence outperforms the uniform placement. However, for large Doppler frequency or high SNR, reducing the prediction errors caused by channel variation is more important. Compared with the ending placement, the uniformly inserted training symbols can provide more information about the channels to be predicted, and thus, a longer prediction horizon can be achieved. The equally spaced placement with optimal interval d adaptively adjusts from the ending placement to the uniform placement with the increase in Doppler frequency or SNR. However, there is still an evident performance gap from the proposed placement.

Figs. 5 and 6 show the average normalized sum MSE over the first Q data symbols in the downlink frame[3] versus Doppler frequency and SNR. Two kinds of exhaustive searching are considered, which are aimed at minimizing the normalized sum MSE for the Qth data symbol (denoted by "Exhaustive-Q")[4] and at minimizing the average normalized sum MSE for overall Q data symbols (denoted by "Exhaustive-Average"), respectively. It is shown that the proposed training placement with alternating searching based on the "worst-case" strategy achieves almost the same performance as the two kinds of exhaustive searching and exhibits an evident performance gain over the ending placement, the uniform placement, and the equally spaced placement with the optimized interval d for large Doppler frequency or high SNR. This is because the MSE for predicting the channel of the farmost symbol, i.e., the Qth symbol, dominates the sum MSE or the average MSE for predicting the channels of all Q symbols.

Fig. 7 shows the convergence behavior of the proposed alternating searching algorithm. Different initial values are compared to show their impact on the convergence speed, including those obtained from ending placement, uniform placement, and the proposed successive initialization. It can be observed that the alternating searching algorithm with successive initialization has the fastest convergence speed. The proposed searching algorithm with various initial values almost achieves the same performance as the exhaustive searching after only one iteration.

V. Conclusion

In this paper, we have studied the training symbol placement design for channel prediction in TDD MIMO systems. We have proven that, for the first-order Gauss–Markov channel model, the optimal positions lie at the end of the uplink frame. For general channel models, such an ending placement only performs well in the cases of small Doppler frequency or low SNR. We have then proposed a low-complexity alternating-searching-based training design method. Numerical results have shown that the proposed searching algorithm converges fast and achieves almost the same performance as the optimal training placement found by exhaustive searching. We have shown that the optimal placement is nonuniform, which demonstrates an evident performance gain over the uniform placement, the ending placement, and the equally spaced placement.

[3]In practical systems, downlink closed-loop transmission may be applied for the first several symbols in the downlink frame.

[4]This is in fact a "worst-case" design strategy.

APPENDIX
PROOF OF PROPOSITION 1

According to the first-order Gauss–Markov model, the time-varying channel is

$$\mathbf{H}(t) = \alpha \mathbf{H}(t - T_s) + \mathbf{U}(t) \tag{13}$$

where $\alpha \in (0, 1)$ is the fading correlation coefficient that controls the rate of channel variation, and $\mathbf{U}(t)$ is the driven noise consisting of *i.i.d.* complex Gaussian entries with zero mean and variance $1 - \alpha^2$.

It is well known that this channel model has a special temporal correlation function, which is $R_t((m+n)T_s) = R_t(mT_s)R_t(nT_s) = \alpha^{m+n}$. Then, we have $[\mathbf{\Gamma}(\mathcal{S})]_{ij} = \alpha^{|p_i - p_j|}$ and $[\mathbf{r}(\mathcal{S})]_i = \alpha^{p_i+Q}$, $1 \le i,\ j \le L$, and the normalized sum MSE in (7) can be rewritten as

$$\epsilon(\mathcal{S}) = 1 - \frac{\alpha^{2Q+2p_1}}{\mathrm{tr}(\mathbf{R}_C)} \mathrm{tr}\Big(\left(\tilde{\mathbf{r}}^H(\mathcal{S}) \otimes \mathbf{R}_C\right) \times \left(\mathbf{\Gamma}(\mathcal{S}) \otimes \mathbf{R}_C + \sigma^2 \left(\mathbf{I} \otimes \mathbf{R}_X^{-1}\right)\right)^{-1} \left(\tilde{\mathbf{r}}(\mathcal{S}) \otimes \mathbf{R}_C\right) \Big) \tag{14}$$

where $\tilde{\mathbf{r}}(\mathcal{S}) = [1, \alpha^{p_2 - p_1}, \ldots, \alpha^{p_L - p_1}]^T$, and $p_1 < \cdots < p_L$ is assumed without loss of generality.

When the SNR is very high, it follows that the sum MSE only depends on p_1, noting that $\tilde{\mathbf{r}}(\mathcal{S})$ is exactly the same as the first column vector of $\mathbf{\Gamma}(\mathcal{S})$. That is to say that the MSE will be minimized when a single training symbol is placed at the end of the uplink frame. This implies that the prediction can be equivalently decomposed into the following two steps. We first use the L observations to estimate the channel at the position of p_1 and then use this channel estimate to predict the future channels. This can be more clearly seen as follows.

Using the L observations $\bar{\mathbf{y}}$, the Wiener channel estimate at the position of p_1 can be obtained after some manipulations as

$$\begin{aligned} \hat{\mathbf{h}}(t - p_1 T_s) &= \left(\tilde{\mathbf{r}}^H(\mathcal{S}) \otimes \mathbf{R}_C\right) \\ &\quad \times \left(\mathbf{\Gamma}(\mathcal{S}) \otimes \mathbf{R}_C + \sigma^2 (\mathbf{I} \otimes \mathbf{R}_X^{-1})\right)^{-1} \bar{\mathbf{h}} + \mathbf{z} \\ &\triangleq \mathbf{A}^H \mathbf{B}^{-1} \bar{\mathbf{h}} + \mathbf{z} \end{aligned} \tag{15}$$

where $\bar{\mathbf{h}} = [\mathbf{h}^T(t - p_1 T_s), \ldots, \mathbf{h}^T(t - p_L T_s)]^T$, $\mathbf{z}$ is the complex Gaussian estimation error vector with zero mean and covariance matrix $\mathbf{A}^H \mathbf{B}^{-1} (\mathbf{I} \otimes \mathbf{R}_X^{-1}) \mathbf{B}^{-1} \mathbf{A} \sigma^2$, and $\triangleq$ denotes the definition.

Noting that $\mathbb{E}\{\mathbf{h}(t + QT_s)\hat{\mathbf{h}}^H(t - p_1 T_s)\} = \alpha^{Q+p_1} \mathbf{A}^H \mathbf{B}^{-1} \mathbf{A}$ and $\mathbb{E}\{\hat{\mathbf{h}}(t - p_1 T_s)\hat{\mathbf{h}}^H(t - p_1 T_s)\} = \mathbf{A}^H \mathbf{B}^{-1} \mathbf{A}$, it is not hard to show that the normalized sum MSE of using $\hat{\mathbf{h}}(t - p_1 T_s)$ to predict $\mathbf{h}(t + QT_s)$ is the same as (14). This indicates that the decomposition of prediction to the two steps does not cause any performance loss.

Consequently, the training symbols at the positions of p_i, $i = 2, \ldots, L$ merely contribute to improving the estimation accuracy of the channel at p_1, which only depends on the relative positions of $p_i - p_1$. By applying the matrix inversion lemma of partitioned matrices, we can prove that all training symbols should be consecutively inserted to minimize the MSE of the channel estimation. The details are omitted due to the lack of space. This can be explained by the fact that high correlation between observations can effectively suppress the observation noises. Therefore, in general SNR level, the optimal positions for training symbols are at the end of the uplink frame.

REFERENCES

[1] A. Duel-Hallen, "Fading channel prediction for mobile radio adaptive transmission systems," *Proc. IEEE*, vol. 95, no. 12, pp. 2299–2313, Dec. 2007.

[2] D. Gesbert, M. Kountouris, R. W. Heath, Jr., C. B. Chae, and T. Salzer, "Shifting the MIMO paradigm: From single user to multiuser communications," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Oct. 2007.

[3] I. C. Wong and B. L. Evans, "Sinusoidal modeling and adaptive channel prediction in mobile OFDM systems," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1601–1615, Apr. 2008.

[4] M. D. Larsen, A. L. Swindlehurst, and T. Svantesson, "Performance bounds for MIMO-OFDM channel estimation," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1901–1916, May 2009.

[5] T. Svantesson and A. L. Swindlehurst, "A performance bound for prediction of MIMO channels," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 520–529, Feb. 2006.

[6] I. C. Wong and B. L. Evans, "Exploiting spatio-temporal correlations in MIMO wireless channel prediction," in *Proc. IEEE GLOBECOM*, 2006, pp. 1–5.

[7] X. Ma, G. B. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly selective wireless fading channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351–1366, May 2003.

[8] S. Ohno and G. B. Giannakis, "Capacity maximizing MMSE-optimal pilots for wireless OFDM over frequency-selective block Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2138–2145, Sep. 2004.

[9] A. Y. Panah, R. G. Vaughan, and R. W. Heath, Jr., "Optimizing pilot locations using feedback in OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 58, no. 6, pp. 2803–2814, Jul. 2009.

[10] Z. Tang and G. Leus, "Time-multiplexed training for time-selective channels," *IEEE Signal Process. Lett.*, vol. 14, no. 9, pp. 585–588, Sep. 2007.

[11] C. Shin, J. G. Andrews, and E. J. Powers, "An efficient design of doubly selective channel estimation for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 10, pp. 3790–3802, Oct. 2007.

[12] S. Savazzi and U. Spagnolini, "Optimizing training lengths and training intervals in time-varying fading channels," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1098–1112, Mar. 2009.

[13] M. Dong and L. Tong, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1403–1418, May 2004.

[14] T. A. Lamahewa, P. Sadeghi, R. A. Kennedy, and P. B. Rapajic, "Model-based pilot and data power adaptation in PSAM with periodic delayed feedback," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2247–2252, May 2009.

[15] J. F. Valenzuela-Valdes, A. M. Martinez-Gonzalez, and D. A. Sanchez-Hernandez, "Accurate estimation of correlation and capacity for hybrid spatial-angular MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4036–4045, Oct. 2009.

[16] A. Dogandzic and B. Zhang, "Estimating Jakes' Doppler power spectrum parameters using the Whittle approximation," *IEEE Trans. Signal Process.*, vol. 53, no. 3, pp. 987–1005, Mar. 2005.

[17] H. Zhang and A. Abdi, "Cyclostationarity-based Doppler spread estimation in mobile fading channels," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 1061–1067, Apr. 2009.

[18] H. Zhao, P. Fan, P. T. Mathiopoulos, and S. Papaharalabos, "On SNR estimation techniques for turbo decoding over uncorrelated Rayleigh fading channels with unknown fading parameters," *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 4955–4961, Nov. 2009.

[19] E. A. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1543–1553, Sep. 2004.

[20] H. L. V. Trees, *Optimum Array Processing, Part IV of Detection, Estimation and Modulation Theory*. Hoboken, NJ: Wiley, 2002.

[21] S. L. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sep. 2001.

[22] W. Jakes and D. Cox, *Microwave Mobile Communications*. Piscataway, NJ: Wiley-IEEE Press, 1994.

[23] M. Narandzic, C. Schneider, R. Thoma, T. Jamsa, P. Kyosti, and X. Zhao, "Comparison of SCM, SCME, and WINNER channel models," in *Proc. IEEE VTC*, 2007, pp. 413–417.

[24] 3GPP TS 36.211 V8.2.0, *Evolved Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 8)*, Feb. 2008.