

# Model-Based Channel Estimation Framework for MIMO Multicarrier Communication Systems

Xiaowen Wang and K. J. Ray Liu, *Fellow, IEEE*

**Abstract**—The multicarrier modulation (MCM) using multiple antennas is a promising technique for broadband communications over mobile wireless channels. In this work, we investigate the channel estimation problem for the MCM system with multiple transmitter and receiver antennas. The difficulty of such a problem is that the number of the channel parameters increases proportionally with the number of transmitter antennas while the observations at the receiver do not. A model-based channel estimation framework is proposed to identify the multiple channels simultaneously. Based on this framework, we further discuss the identification condition and the training sequence design by taking into account both the model error and residual noise. Finally, we show the performance of the proposed model-based channel estimation scheme using two types of models, Fourier-transform-based model and polynomial model. We also show the system performance of two space–time codes using the proposed channel estimation algorithm.

**Index Terms**—Channel estimation, multicarrier, multiple-input multiple-output (MIMO).

## I. INTRODUCTION

TO SUPPORT the broadband applications in wireless communications, a system has to combat the serious impairment of the wireless channel, namely the multipath and fading. Multicarrier modulation (MCM) is considered as an effective technique for wireless broadband communications [1] for its great resistance to the intersymbol interference (ISI) caused by the multipath effect. The multiple-input multiple-output (MIMO) system that uses multiple antennas to exploit the diversity in the channel is very effective to combat the other serious impairment, the fading effect. It has been proved that the channel capacity is proportional to the number of the transmitter or receiver antennas [32]. Many spatial-temporal signal processing techniques have been developed [28] to exploit the diversity gain. For the case of multiple receiver antennas, optimal combining is developed to make the best use of the information received by different antennas [27]. For the case of multiple transmitter antennas, the space–time coding and modulation schemes are designed to achieve higher diversity gain [29]–[31]. Since the differential detection is hard

to design for the systems with multiple transmitter antennas, most MIMO MCM systems are designed with the assumption that the channel information is known.

Various channel estimation schemes are proposed for the single antenna MCM communication systems [6]–[11], [13], [14], [16]–[18]. The channel estimation in MCM systems needs to estimate the channel responses of all subchannels which are a large number of parameters. However, those channel responses are not independent but correlated with each other because the physical channel cannot vary randomly. The channel estimation can be greatly improved by exploiting such correlation. One way to exploit the correlation is to model the channel responses by some efficient channel models and estimate a much smaller set of model coefficients instead of the large number of channel responses. The most popular finite impulse response (FIR) filter channel model is used in [7], [8], [10], [11], and [14], while the authors in [15]–[18] use the other type of model, namely the polynomial model.

If there is no cochannel interference, the channel estimation algorithm developed for the single antenna system can be applied to different receiver antennas for the system that only uses multiple receiver antennas. The real challenge is the channel estimation for the system with multiple transmitter antennas. The problem becomes more difficult in this case because we have to estimate multiple sets of channels corresponding to different transmitter antennas. It is impossible to directly estimate these channel responses if the amount of training data does not increase. So either we have to send duplicate training data to train each channel separately as in [23], or we have to use an efficient channel model to reduce the number of parameters required to estimate, for example, the FIR channel model used in [19]–[22] and [24]. Only if we express the channel efficiently enough, i.e., using a small enough number of model coefficients, can we identify the channel. Applying the modeling idea, we develop a more general framework to estimate the channel for the MCM system with transmitter diversity. This framework does not assume any specific underline model and can be used on any window of observations.

Although the efficient channel model is crucial to identify the MIMO channels, it causes model error, which is not considered in [19]–[22] and [24]. Therefore, a tradeoff has to be made in choosing the model basis. Furthermore, we recognize that besides using the appropriate model to capture the major features of the channel, the training sequences sending from different transmitter antennas have to be carefully designed in order to identify the channel. The training sequence design problem is

Manuscript received January 28, 2003; revised September 18, 2003; accepted March 10, 2004. The editor coordinating the review of this paper and approving it for publication is A. F. Molisch.

X. Wang is with the Wireless Systems Research Department, Agere Systems, Allentown, PA 18109 USA (e-mail: xiaowenw@agere.com).

K. J. R. Liu is with the Electrical and Computer Engineering Department, University of Maryland, College Park, MD 20742 USA (e-mail: kjrlu@eng.umd.edu).

Digital Object Identifier 10.1109/TWC.2005.847020

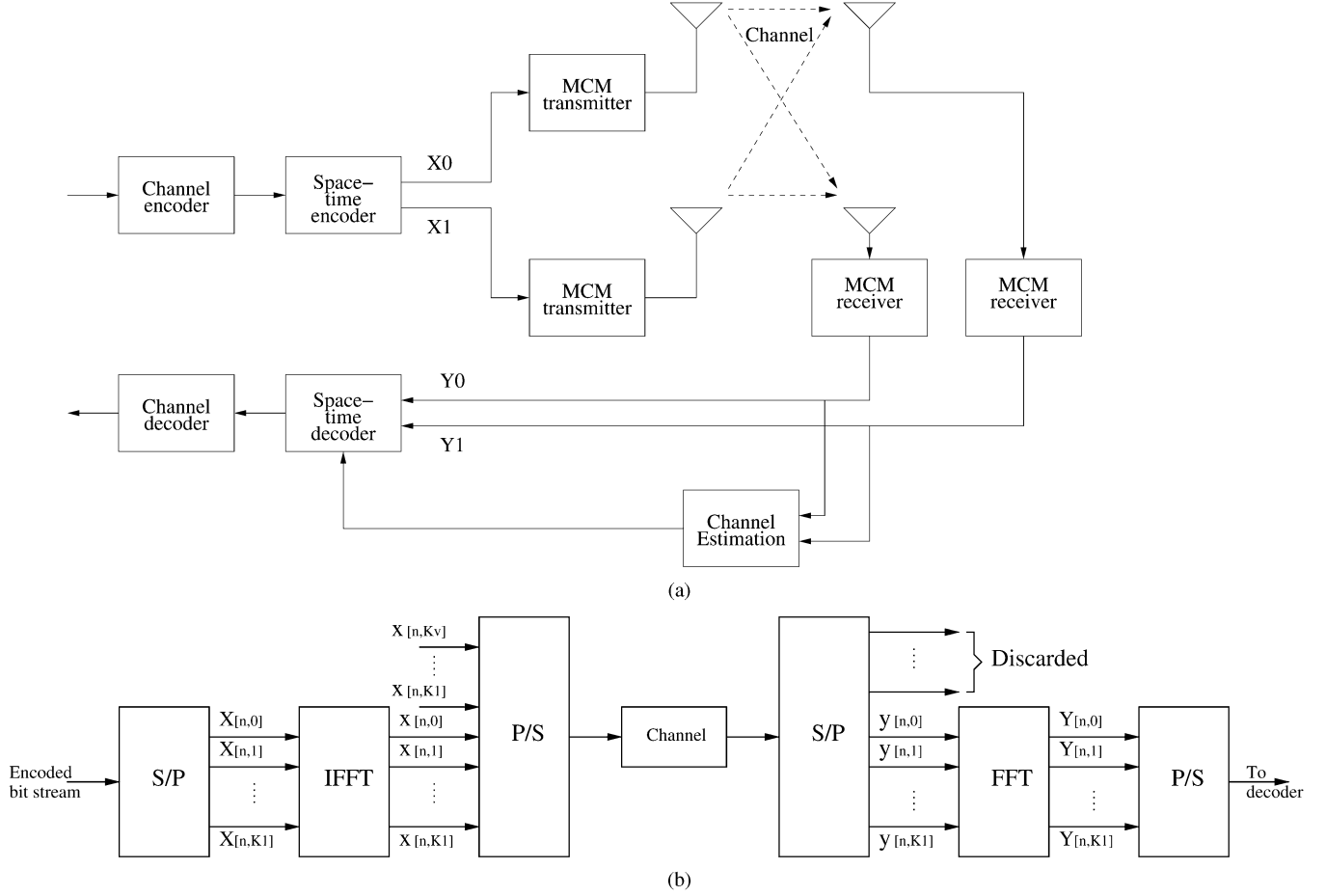


Fig. 1. (a) MCM systems with multiple antennas. (b) MCM transmitter and receiver.

discussed in [19], [21], [22], and [24]. The optimal training sequence design rule based on the FIR channel model is given to minimize the residual estimation noise in [22]. However, because the channel estimation error consists of residual noise as well as model error, we address the training sequence design problem together with the model selection to minimize both model error and residual noise.

In the rest of the paper, we first introduce the MIMO MCM systems with multiple transmitter antennas. Then the model-based channel estimation framework for such a system is derived and analyzed. Based on the analysis, the training sequence design and model selection are discussed. Finally, we use two types of models, the Fourier-transform-based model [6], [12] and the polynomial model in [15] and [17] as examples to demonstrate the performance of the channel estimation schemes in computer simulations.

## II. SYSTEM AND CHANNEL MODELS

### A. MIMO MCM Systems

Fig. 1 shows a schematic diagram of an MIMO MCM system with two transmitter antennas and two receiver antennas. The number of transmitter antennas is denoted as  $m_T$  and the number of receiver antennas as  $m_R$ . As shown in Fig. 1, one

block of MCM signal goes through the space-time encoder to form  $m_T$  blocks of MCM signals. The transmitter of the MCM system at the  $i$ th transmitter antenna and the receiver at the  $m$ th receiver antenna are shown in Fig. 1(b). The whole bandwidth  $B_d$  is divided into  $K$  subchannels. The  $n$ th block of signal generated by the space-time encoder consists of  $K$  subsymbols  $X_i[n, k]$ ,  $k = 0, \dots, K - 1$ . Then the modulation is implemented by  $K$ -point inverse discrete Fourier transform (IDFT), i.e.,  $x_i[n, l] = (1/\sqrt{K}) \sum_{k=0}^{K-1} X_i[n, k] e^{j(2lk\pi)/K}$ ,  $l = 0, \dots, K - 1$ . The modulated data pass through a P/S converter to form the serial data  $x_i[n, k]$ . A cyclic prefix of length  $v$  is inserted before sending out  $x_i[n, k]$ . The cyclic prefix is constructed by  $x_i^f[n, l] = x_i[n, K + l]$ ,  $l = -v, -v + 1, \dots, -1$ . Then this signal goes through a transmitter filter  $g_T(t)$  and transmits to the channel.

The received signal first goes through the front end filter  $g_R(t)$ . Then, the cyclic prefix  $y_m^f[n, l]$  is discarded and the received signal  $y_m[n, l]$  is demodulated by the DFT,  $Y_m[n, k] = (1/\sqrt{K}) \sum_{l=0}^{K-1} y_m[n, l] e^{-j(2lk\pi)/K}$ ,  $k = 0, \dots, K - 1$ .

Suppose the compound filter of the transmitter and receiver antenna  $g_T(t) * g_R(t)$  has a flat spectrum in the band of interest [13], then the effect of the transmitter filter and receiver filter can be ignored. Furthermore, if the cyclic prefix is longer than the channel time delay spread, then there is no ISI between two

MCM symbols. Then the subchannels can be viewed as independent of each other, i.e., the received signal of the  $k$ th subchannel in the  $n$ th block at the  $m$ th receiver antenna is

$$Y_m[n, k] = \sum_{i=1}^{m_T} H_{im}[n, k]X_i[n, k] + \Upsilon_m[n, k] \quad (1)$$

for  $j = 1, \dots, m_R$

where  $H_{im}[n, k]$  is channel response from the  $i$ th transmitter antenna to the  $m$ th receiver antenna, and  $\Upsilon_m[n, k]$  is the noise at the  $m$ th receiver antenna. The noise is assumed to be white Gaussian noise with zero mean and variance of  $\sigma^2$  and independent of different  $n$ ,  $k$  and  $i$ ,  $m$ .

The detection and decoding are done over  $Y_m[n, k]$ . The general detection metrics is

$$\sum_n \sum_k \sum_{m=0}^{m_R} \left\| Y_m[n, k] - \sum_{i=1}^{m_T} H_{im}[n, k]X_i[n, k] \right\|^2 \quad (2)$$

where the  $n$  and  $k$  are summed over one space-time codeword. The detected signal is the transmit sequence that minimize this metric. The knowledge of the channel responses  $H_{im}[n, k]$  is necessary for most of space-time coding schemes. It is also obvious that at each receiver antenna, we need to estimate  $m_T$  sets of channel responses  $H_{im}[n, k]$  for  $i = 1, \dots, m_T$ , which makes the channel estimation problem more challenging.

### B. Wireless Channels

A fading multipath channel caused by both terrain and motion is generally described by [33], [35]

$$h(t, \tau) = \sum_i \gamma_i(t) \delta(\tau - \tau_i) \quad (3)$$

where  $h(t, \tau)$  is the baseband impulse response,  $\tau_i$  denote the different path delays, and  $\gamma_i(t)$  are independent complex Gaussian processes with variance  $p_i$ .  $\gamma_i(t)$  is assumed to have the same normalized correlation function, i.e.,  $r_t(\tau) = (1/\sigma_i^2) E[\gamma_i(t + \tau)\gamma_i^*(t)]$ . Then the channel frequency response is  $H(t, f) = \int h(t, \tau) e^{-j2\pi f \tau} d\tau$ .

Different statistical distributions are used to characterize  $\gamma_i(t)$ . Rayleigh is one of them that is often used to describe a scenario where no line of sight path exists. The time domain correlation function of the Rayleigh fading is  $r_J(t) = J_0(2\pi f_D t)$ , where  $J_0(\cdot)$  is the zeroth-order Bessel function, and  $f_D$  is the Doppler shift describing the variation of the channel response along  $t$ .

The multipath effect is characterized by the delay profile which consists of  $p_i$  and  $\tau_i$ . The maximal delay spread is defined as

$$T_d = \max_i \tau_i. \quad (4)$$

Fig. 2 shows two typical delay profiles, typical urban (TU) and hilly terrain (HT).

The discrete expression in Section II-A is the sampling of the continuous channel response with

$$H_{im}[n, k] = H_{im}(nT_f, k\Delta f) \quad (5)$$

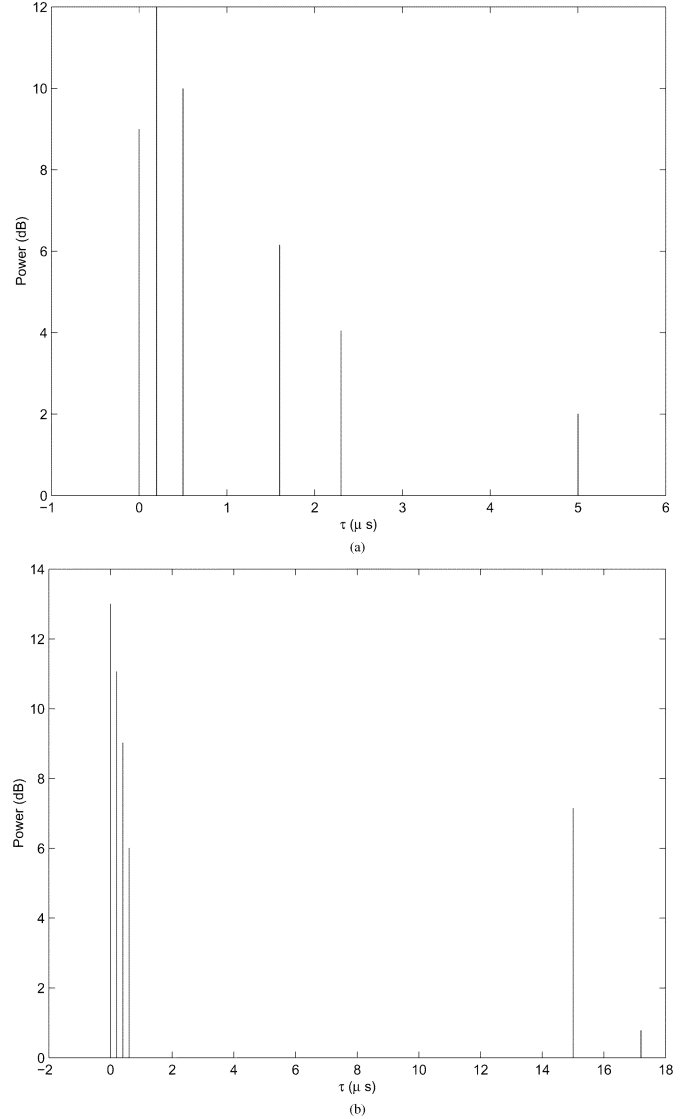


Fig. 2. Delay profiles (a) TU (b) HT.

where  $T_f = ((K + v)/B_d)$  is the block duration and  $\Delta f = (B_d/K)$  is the bandwidth of the subchannel.

We assume that all the channels from different transmitter antennas to different receiver antennas have the same delay and fading property, i.e., the same delay profile [4].

### III. MODEL-BASED CHANNEL ESTIMATION

In this section, we begin to consider the channel estimation problem at  $m$ th receiver antenna. The estimation scheme developed here can be applied to all receiver antennas to obtain all the channel estimations. Hence, we will omit the subscript  $m$ .

We assume that  $m_T$  training sequences  $t_i[k]$  are sent simultaneously from all transmitter antennas. Then considering a window of  $L$  received samples  $\mathbf{Y}[n, k] = [Y[n, k]Y[n, k + 1], \dots, Y[n, k + L - 1]]^T$ , write (1) in the matrix form we have

$$\mathbf{Y}[n, k] = \mathbf{A}[k]\mathbf{H}[n, k] + \mathbf{\Upsilon}[n, k] \quad (6)$$

where  $\mathbf{A}[k] = [\mathbf{A}_1[k], \mathbf{A}_2[k], \dots, \mathbf{A}_{m_T}[k]]$  and  $\mathbf{H}[n, k] = [\mathbf{H}_1^T[n, k], \mathbf{H}_2^T[n, k], \dots, \mathbf{H}_{m_T}^T[n, k]]^T$ , with  $\mathbf{A}_i[k] = \text{diag}(t_i[n, k], t_i[n, k+1], \dots, t_i[n, k+L-1])$  and  $\mathbf{H}_i[n, k] = [H_i[n, k], H_i[n, k+1], \dots, H_i[n, k+L-1]]^T$ , for  $i = 1, 2, \dots, m_T$  and  $\mathbf{Y}[n, k] = [\mathbf{Y}_1^T[n, k], \mathbf{Y}_2^T[n, k], \dots, \mathbf{Y}_{m_T}^T[n, k]]^T$ .

If we only have one transmitter antenna, then  $\mathbf{A}[k]$  and  $\mathbf{H}[n, k]$  degenerate to  $\mathbf{A}_1[k]$  and  $\mathbf{H}_1[n, k]$ . Because  $\mathbf{A}_1[k]$  is a square matrix, we can get a channel estimate directly by matrix inversion as  $\tilde{\mathbf{H}}_m[n, k] = \mathbf{A}_1^{-1}[k] \mathbf{Y}_m[n, k]$ . However, if  $m_T > 1$ , the number of channel responses required to estimate becomes  $m_T L$ , which is  $m_T$  times of that in the single transmitter antenna case, but we still only have  $L$  observations. Our solution is to use an efficient model to reduce the number of the parameters that need to be estimated.

### A. Efficient Channel Modeling

The channel responses can be projected to some function basis because they are correlated. Such a projection can be done in both time and frequency domain just like in [17] and [19]. In this paper, we will project the channel response in frequency domain to exploit the frequency domain correlation in order to keep the derivation simple. However, the framework can be easily extended to exploit the time domain correlation if we choose a time-frequency window instead of just frequency window as discussed in the rest of the paper.

If the channel responses  $H_i[n, k]$  are correlated, which is the case in practice, we can find a model basis  $\{p_j(f) | j = 0, 1, \dots, l-1\}$  that  $H_i[n, k_1]$ , can all be expressed by  $b_i^j[n, k]$ , for  $k \leq k_1 \leq k+L-1$

$$H_i[n, k_1] = \sum_{j=0}^{l-1} b_i^j[n, k] p_j((k_1 - k)\Delta f) \quad (7)$$

with small model error which is often measured by  $\sum_{k_1=k}^{k+L-1} E[\|H_i[n, k_1] - \sum_{j=0}^{l-1} b_i^j[n, k] p_j((k_1 - k)\Delta f)\|^2]$ .

In this case, we can express  $\mathbf{H}_i[n, k]$  as

$$\mathbf{H}_i[n, k] = \mathbf{Q} \mathbf{b}_i[n, k] \quad (8)$$

where  $\mathbf{b}_i[n, k] = [b_i^0[n, k], b_i^1[n, k], \dots, b_i^{l-1}[n, k]]^T$  and  $\mathbf{Q}$  is a  $L \times l$  matrix consisting of the model basis,

$$\mathbf{Q} = \begin{bmatrix} p_0(0) & p_1(0) & \dots & p_{l-1}(0) \\ p_0(\Delta f) & p_1(\Delta f) & \dots & p_{l-1}(\Delta f) \\ \vdots & \vdots & \ddots & \vdots \\ p_0((L-1)\Delta f) & p_1((L-1)\Delta f) & \dots & p_{l-1}((L-1)\Delta f) \end{bmatrix}$$

The problem now is to find the model bases that can express the channel responses with less model coefficients and small model error. In the ideal case, if the correlation of the channel responses, i.e.,

$$\mathbf{R}_H = E[\mathbf{H}_i[n, k] \mathbf{H}_i^H[n, k]] \quad (9)$$

is known. The model basis  $\mathbf{Q}$  ideally should be the matrix consists of the eigenvectors of the correlation matrix corresponding to nonzero eigenvalue of  $\mathbf{R}_H$  [6]. However, the correlation of

the channel responses is usually not available. We need to look for some model basis that can have a fairly small model error for the channels that we are interested in.

For the wireless channel described in Section II-B, one natural choice to model the channel is to use Fourier basis. The energy of the inverse Fourier transform of  $H(t, f)$  regarding  $f$  is actually limited in a finite delay  $T_d$ . Hence, we can use all the “low frequency” within “bandwidth  $T_d$ ” in the transform domain of the inverse Fourier transform to express the channel as in (3). In this case, the basis function is<sup>1</sup>

$$(\mathbf{Q})_{mn} = e^{-j((2k_m n \pi)/L)} \quad (10)$$

for  $m = 0, \dots, L-1$  and  $n = 0, \dots, l-1$ .

Such a model is used in the channel estimation methods of [7], [8], [10], [14], [19], [21], and [22]. In [10], [19], [21], and [22],  $L = K$  and  $k_m = m$ , while in [7] and [8],  $L$  is the length of the embedded pilot tones and  $k_m$  are the indexes of the pilot tones.  $L$  is the number of used subcarriers and  $k_m$  are indexes of the used subcarriers in [14]. The problem with this model is that the system bandwidth is limited. Then there may be a large leakage when the inverse Fourier transform is performed, which leads to a large model approximation error [18].

Now let us look at another model. We know that  $H(t, f)$  is smoothly changing along  $f$ . Based on the approximation theory [3], [36], it was shown that such a smoothly time-varying channel can be closely approximated over a short interval  $L\Delta f$  by a series of polynomial bases, i.e.,

$$H_i[n, k_1] = \sum_{j=0}^{l-1} b_i^j[n, k] \left( \frac{k_1 - k - L}{2} \right)^j + R_l[n, k_1] \quad (11)$$

for  $k \leq k_1 \leq k+L$

where  $b_i^j[n, k] = ((\Delta f)^j / j!) ((\partial^j H(nT_f, f)) / (\partial f^j))|_{f=(k+L/2)\Delta f}$  and  $R_l[n, k_1] = (((k_1 - k - L/2)\Delta f)^l / l!) ((\partial^l H(nT_f, f)) / (\partial f^l))|_{f=f'}$ , with  $|f' - (k+L/2)\Delta f| \leq ((L\Delta f)/2)$ .

When  $\Delta f T_d \ll 1$ ,  $E[\|R_l[n, k_1]\|^2]$  goes to zero for all  $k \leq k_1 \leq k+L-1$ , if  $l$  is large enough. For the practical MCM system, this condition is usually satisfied for a small  $l$ . In this case, the model basis is

$$(\mathbf{Q})_{mn} = \left( \frac{m-L}{2} \right)^n \quad (12)$$

for  $m = 0, \dots, L-1$  and  $n = 0, \dots, l-1$ .

The channel estimation algorithms in [15]–[18] and [25] are examples of using this model in the single antenna system.

### B. Channel Estimation Algorithm

Suppose we find an efficient and accurate model and express the channel as in (8). Substituting it in (6), we have

$$\mathbf{Y}[n, k] = \mathbf{A}[k] \begin{bmatrix} \mathbf{Q} & & \\ & \ddots & \\ & & \mathbf{Q} \end{bmatrix} \begin{bmatrix} \mathbf{b}_1[n, k] \\ \vdots \\ \mathbf{b}_{m_T}[n, k] \end{bmatrix} + \mathbf{Y}[n, k]. \quad (13)$$

<sup>1</sup>In this paper, we use  $(\mathbf{B})_{mn}$  to denote the element at  $m$ th row and  $n$ th column of matrix  $\mathbf{B}$ .

If the model is efficient enough that  $m_T l < L$ , then we can get the least squares (LS) estimate for the model coefficients as

$$\begin{bmatrix} \hat{\mathbf{b}}_1[n, k] \\ \vdots \\ \hat{\mathbf{b}}_{m_T}[n, k] \end{bmatrix} = [\mathbf{A}_1[k]\mathbf{Q} \quad \dots \quad \mathbf{A}_{m_T}[k]\mathbf{Q}]^\dagger \mathbf{Y}[n, k] \quad (14)$$

where  $\dagger$  denotes the pseudoinverse.

Substituting into (8), we get the estimation of the channel as

$$\hat{\mathbf{H}}[n, k] = \mathbf{Q}\mathbf{R}^{-1}[k]\mathbf{Q}^H\mathcal{P}[n, k] \quad (15)$$

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} & \dots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Q} \end{bmatrix},$$

$$\mathcal{R}[k] = \begin{bmatrix} \mathbf{Q}^H\mathbf{C}_{11}[k]\mathbf{Q} & \mathbf{Q}^H\mathbf{C}_{12}[k]\mathbf{Q} & \dots & \mathbf{Q}^H\mathbf{C}_{1m_T}[k]\mathbf{Q} \\ \mathbf{Q}^H\mathbf{C}_{21}[k]\mathbf{Q} & \mathbf{Q}^H\mathbf{C}_{22}[k]\mathbf{Q} & \dots & \mathbf{Q}^H\mathbf{C}_{2m_T}[k]\mathbf{Q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Q}^H\mathbf{C}_{m_T1}[k]\mathbf{Q} & \mathbf{Q}^H\mathbf{C}_{m_T2}[k]\mathbf{Q} & \dots & \mathbf{Q}^H\mathbf{C}_{m_Tm_T}[k]\mathbf{Q} \end{bmatrix}$$

and  $\mathcal{P}[n, k] = [\mathbf{P}_1[n, k] \quad \mathbf{P}_2[n, k] \quad \dots \quad \mathbf{P}_{m_T}[n, k]]^T$  with  $\mathbf{C}_{ij}[n, k] = \mathbf{A}_i^H[k]\mathbf{A}_j[k]$  being the correlation of the transmitted signal and  $\mathbf{P}_i[n, k] = \mathbf{A}_i^H[k]\mathbf{Y}[n, k]$  being the cross-correlation between the transmitted and received signals.

In this estimation algorithm,  $\mathbf{Q}$  and  $\mathcal{R}[k]$  can be calculated off-line once the model and training sequence are determined. The general estimator structure is shown in Fig. 3(a). The computation complexity is low if  $l \ll L$ . The computation can be further reduced if the transformation  $\mathbf{Q}$  can be calculated by some fast algorithm. For example, fast Fourier transform can be used in the case using Fourier transform model [12].

The estimation error generally distributes unevenly among the  $L$  samples. To minimize the estimation error, we can choose only one sample as the final estimate, for example the center point of the window. Then slide the window to get all the estimates. The estimation in this case becomes a filtering process as illustrated in Fig. 3(b). The filter taps are the row of matrix  $\mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}^H$  corresponding to the specific estimation point in the window. The computations needed for the filtering process can be reduced if the filter can be implemented by some iterative algorithms [18].

### C. Mean Squared Estimation Error (MSE)

Let us define the mean squared error matrix as

$$\boldsymbol{\varepsilon} = E[(\mathbf{H}[n, k] - \hat{\mathbf{H}}[n, k])(\mathbf{H}[n, k] - \hat{\mathbf{H}}[n, k])^H]. \quad (16)$$

Then the  $(iL + k)$ th diagonal element of  $\boldsymbol{\varepsilon}$  is the MSE of the  $k$  subcarrier at  $i$ th transmitter antenna, i.e.,  $E[||H_i[n, k] - \hat{H}_i[n, k]||^2] = (\boldsymbol{\varepsilon})_{iL+k, iL+k}$ .

Substituting (15) into (16), we have

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_h + \boldsymbol{\varepsilon}_n \quad (17)$$

where  $\boldsymbol{\varepsilon}_h$  is the mean squared model error caused by inaccurate modeling, which is

$$\boldsymbol{\varepsilon}_h = (\mathbf{Q}\mathbf{R}^{-1}[k]\mathbf{Q}^H\mathbf{A}^H[k]\mathbf{A}[k] - \mathbf{I})E[\mathbf{H}[n, k]\mathbf{H}^H[n, k]] (\mathbf{Q}\mathbf{R}^{-1}[k]\mathbf{Q}^H\mathbf{A}[k]\mathbf{A}^H[k] - \mathbf{I})^H \quad (18)$$

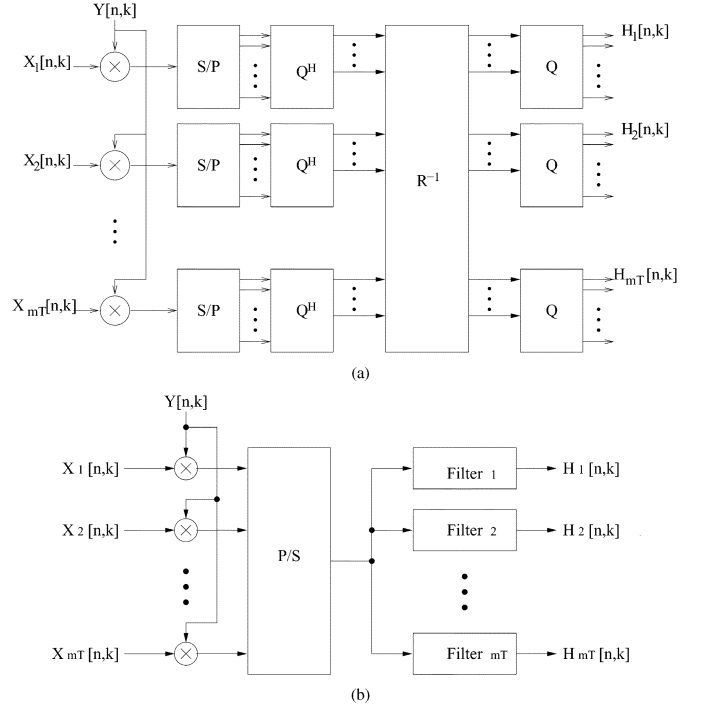


Fig. 3. Estimator structures.

and  $\boldsymbol{\varepsilon}_n$  is the variance of the residual noise, which is

$$\boldsymbol{\varepsilon}_n = \sigma^2\mathbf{Q}\mathbf{R}^{-1}[k]\mathbf{Q}^H. \quad (19)$$

We assume that all the channel pairs  $h_{im}$  have the same correlation function and are independent of  $i$  and  $m$ . Since  $\mathbf{R}_H$  is a hermitian matrix, it can be diagonalized by  $\mathbf{R}_H = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$ , where  $\mathbf{V}$  is a unitary matrix and  $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_L)$  is composed by the eigenvalues of  $\mathbf{R}_H$ .

It is known that  $\mathbf{Q}$  can be decomposed as

$$\mathbf{Q} = \mathbf{U}\mathbf{W} \quad (20)$$

where  $\mathbf{U}_{L \times l}$  satisfies  $\mathbf{U}^H\mathbf{U} = \mathbf{I}$  and  $\mathbf{W}_{l \times l}$  is an invertible matrix. Then denote  $\bar{\mathbf{U}}_{L \times (L-l)}$  as the orthogonal bases of  $\mathbf{U}$ , i.e.,  $\mathbf{U}^H\bar{\mathbf{U}} = \mathbf{0}$ . Then we can express  $\mathbf{V}$  as  $\mathbf{V} = \mathbf{U}\mathbf{T}^H + \bar{\mathbf{U}}\bar{\mathbf{T}}^H$ , where  $\mathbf{T}^H\mathbf{T} = \mathbf{I}_{l \times l}$ ,  $\bar{\mathbf{T}}^H\bar{\mathbf{T}} = \mathbf{I}_{(L-l) \times (L-l)}$ , and  $\mathbf{T}^H\bar{\mathbf{T}} = \mathbf{0}$ . Then, we have

$$\boldsymbol{\varepsilon}_h = (\mathcal{U}\mathcal{R}_u^{-1}[k]\mathcal{U}^H\mathbf{A}^H[k]\mathbf{A}[k]\bar{\mathcal{U}}\bar{\mathcal{T}}^H - \bar{\mathcal{U}}\bar{\mathcal{T}}^H)\mathcal{Z} \cdot (\mathcal{U}\mathcal{R}_u^{-1}[k]\mathcal{U}^H\mathbf{A}^H[k]\mathbf{A}[k]\bar{\mathcal{U}}\bar{\mathcal{T}}^H - \bar{\mathcal{U}}\bar{\mathcal{T}}^H)^H \quad (21)$$

$$\text{where } \mathcal{U} = \begin{bmatrix} \mathbf{U} & & \\ & \ddots & \\ & & \mathbf{U} \end{bmatrix}, \text{ and } \bar{\mathcal{U}} = \begin{bmatrix} \bar{\mathbf{U}} & & \\ & \ddots & \\ & & \bar{\mathbf{U}} \end{bmatrix},$$

$$\mathcal{R}_u[k] = \mathcal{U}^H\mathbf{A}^H[k]\mathbf{A}[k]\mathcal{U}, \quad \bar{\mathcal{T}} = \begin{bmatrix} \bar{\mathbf{T}} & & \\ & \ddots & \\ & & \bar{\mathbf{T}} \end{bmatrix} \text{ and}$$

$$\mathcal{Z} = \begin{bmatrix} \boldsymbol{\Lambda} & & \\ & \ddots & \\ & & \boldsymbol{\Lambda} \end{bmatrix}.$$

Using the above notations, the residual noise can be written as

$$\boldsymbol{\varepsilon}_n = \sigma^2\mathcal{U}\mathbf{W}(\mathcal{W}^H\mathcal{R}_u[k]\mathcal{W})^{-1}\mathcal{W}^H\mathcal{U}^H = \sigma^2\mathcal{U}\mathcal{R}_u^{-1}[k]\mathcal{U}^H \quad (22)$$

where  $\mathcal{W}$  is a block diagonal matrix whose diagonal blocks are  $\mathbf{W}$ .

#### IV. IDENTIFICATION CONDITION AND TRAINING SEQUENCE DESIGN

##### A. Identification Condition

It can be seen that the estimates in (15) can only be obtained when  $\mathcal{R}^{-1}[k]$  exists. This is called *identification condition*. We can derive the necessary condition for the channel to be identifiable in the following lemma.

*Lemma 1:* The necessary condition for  $\mathcal{R}$  to be invertible is  $l \leq (L/m_T)$ .

This necessary condition shows that the channel model has to be sufficiently efficient. The channels are easier to identify if  $l$  is smaller. However, the efficient enough model may not be accurate enough. On the contrary,  $l$  should be as large as possible to make the diagonal elements of  $\boldsymbol{\varepsilon}_h$  small. As we will show later, the model error can be minimized if we know the channel correlation. However, we usually do not know the channel correlation function and hope to use a fixed model to fit to a range of channels. In this case, the efficiency of the channel model depends on the channel characteristics and the model error allowed. It usually requires the model error to be much less than the noise. If the channel is too dispersive to have a model approximation within the allowed model error, then the channel cannot be identified with the simultaneously transmitted training sequences. In such cases, the channels corresponding to different transmitter antennas have to be identified one at a time, i.e., the training sequences from different transmitter antennas have to be sent alternatively, which increases the overhead of the system.

The structure of the model alone cannot guarantee the identification condition. Proper training sequences need to be carefully chosen to make  $\mathcal{R}[k]$  full rank. For example, if the training sequences sent from different antennas are the same, obviously we cannot identify the channel. Any sequences that satisfy  $|\mathcal{R}[k]| = 0$  lead the channel unidentifiable. However, when considering the constraint posed by the signal constellation, there is almost no such sequence except all the antennas transmitting the same sequence. But there are still some sequences that can make the matrix ill conditioned and, hence, produce a large estimation error.

Indeed, both the model basis and the training sequences determine the identifiability of the channel and the performance of the channel estimation algorithm. So we now take a further look of training sequence design together with the model basis selection.

##### B. Training Sequence Design

In this section, we will discuss the optimal training sequence design for the two estimators shown in Fig. 3. Depending on the estimation schemes, we use different quantities to measure the estimation error. We will discuss two such quantities. First is the average mean squared error, i.e.,

$$\bar{\boldsymbol{\varepsilon}} = \frac{1}{m_T L} \text{tr}(\boldsymbol{\varepsilon}). \quad (23)$$

The average mean squared error is suitable for the scheme in Fig. 3(a). However, if we use the scheme in Fig. 3(b), the more appropriate measure would be to minimize the  $(\boldsymbol{\varepsilon})_{mm}$  with  $m$  being the index of the estimation point inside the observation window.

1) *Training Sequence Minimizing the Average Mean Squared Error:* First, we will discuss the training sequence design criteria to minimize the average mean squared error. We have the following theorem about the lower bound on the average model error.

*Theorem 1:* Denote  $\lambda_1, \lambda_2, \dots, \lambda_L$  as the eigenvalues of  $\mathbf{R}_H$  in descending order, then the average model error  $\bar{\boldsymbol{\varepsilon}}_h \triangleq (1/L) \text{tr}(\boldsymbol{\varepsilon}_h)$  is bounded by the following lower bound:

$$\bar{\boldsymbol{\varepsilon}}_h \geq \frac{1}{L} \sum_{i=l+1}^L \lambda_i \sum_{m=1}^{m_T} \mu_{(m-1)L+i} \quad (24)$$

where  $\mu_j$  is the  $j$ th diagonal element of  $\bar{\mathbf{T}}(I + \bar{\mathbf{U}}^H \mathbf{A}^H[k] \mathbf{A}[k] \bar{\mathbf{U}} \mathcal{R}_u^{-2} \bar{\mathbf{U}}^H \mathbf{A}^H[k] \mathbf{A}[k] \bar{\mathbf{U}}) \bar{\mathbf{T}}^H$ . The equality holds if every column of  $\mathbf{Q}$  is a linear combination of the eigenbases corresponding to  $\lambda_1, \dots, \lambda_l$ .

Moreover, if  $\lambda_i = 0$  for  $i = l+1, \dots, L$  or  $\bar{\mathbf{U}}^H \mathbf{A}^H[k] \mathbf{A}[k] \bar{\mathbf{U}} = \mathbf{0}$ , then

$$\bar{\boldsymbol{\varepsilon}}_h \geq \frac{1}{L} \sum_{i=l+1}^L \lambda_i. \quad (25)$$

Based on this theorem, we have the following conclusions regarding the model basis selection and the training sequence design.

- 1) If we can choose the model bases as the eigenbases corresponding to all the nonzero eigenvalues of  $\mathbf{R}_H$ , then no model error exists as long as the training sequences guarantee that  $\mathcal{R}_u^{-1}[k]$  exists.
- 2) If the number of the nonzero eigenvalues of the channel correlation matrix is larger than  $(L/m_T)$ , then  $\mathcal{R}[k]$  is not invertible and the channel cannot be identified without model error. In such cases, we should select the  $(L/m_T)$  eigenvectors corresponding to the largest  $(L/m_T)$  eigenvalues and also select the training sequences such that  $\bar{\mathbf{U}}^H \mathbf{A}^H[k] \mathbf{A}[k] \bar{\mathbf{U}} = 0$  to minimize  $\mu_i$ .
- 3) Even without any knowledge of the channel correlations, the training sequences are still preferred to satisfy

$$\bar{\mathbf{U}}^H \mathbf{A}^H[k] \mathbf{A}[k] \bar{\mathbf{U}} = 0.$$

Then the model error becomes  $m_T \sum_{i=1}^L \lambda_i (\mathbf{T} \mathbf{T}^H)_{ii}$  according to (41) in Appendix I. In this case, the model error only depends on the model basis and the channel correlation. We can then choose a robust model basis according to some reasonable assumptions about the channel, for example, the maximal delay spread, before designing the training sequence.

For the average power of the residual noise, a lower bound can also be derived.

*Theorem 2:* The average power of the residual noise for any training sequences satisfies the following inequality:

$$\bar{\boldsymbol{\varepsilon}}_n \geq \frac{l\sigma^2}{Lc} \quad (26)$$

where  $c$  is the average transmitting power of the training sequences, i.e.,  $c = E[\|t_m[k]\|^2]$ . The equality is valid if and only if  $\mathcal{U}^H \mathbf{A}^H[k] \mathbf{A}[k] \mathcal{U} = c \mathbf{I}$ .

To achieve the minimal variance of the residual noise, we require that

$$\mathbf{U}^H \mathbf{C}_{mn}[k] \mathbf{U} = \begin{cases} c \mathbf{I}, & \text{if } m = n \\ \mathbf{0}, & \text{if } m \neq n. \end{cases} \quad (27)$$

This requirement had been derived in various works in [19], [21], and [22] using the Fourier transform bases.

Notice that

$$\mathbf{U}^H \mathbf{C}_{mn}[k] \bar{\mathbf{U}} = \mathbf{0} \quad (28)$$

is also required to minimize the model error. Depending on the model basis, there may or may not be training sequences that satisfy both conditions in (27) and (28). If there is no training sequence that satisfies these conditions, a tradeoff has to be made between the model error and residual noise based on the specific applications. For some special model bases with some special structure as stated in the following corollary, training sequences can be found that satisfy both conditions.

*Corollary 1:* Denote  $I_m$  as a subset of index,  $1, \dots, L$  and  $\mathbf{U}_{I_m}$  and  $\bar{\mathbf{U}}_{I_m}$  as matrices that consist the  $i$ th rows of  $\mathbf{U}$  and  $\bar{\mathbf{U}}$  with  $i \in I_m$ , respectively. If there exist at least  $m_T$  nonoverlapped subsets such that  $\mathbf{U}_{I_m}^H \mathbf{U}_{I_m} = \mathbf{I}_l$  and  $\mathbf{U}_{I_m}^H \bar{\mathbf{U}}_{I_m} = \mathbf{0}_{l \times (L-l)}$ . Then there exist such training sequences that  $\mathcal{U}^H \mathbf{A}^H[k] \mathbf{A}[k] \mathcal{U} = c \mathbf{I}$  and  $\bar{\mathcal{U}}^H \mathbf{A}^H[k] \mathbf{A}[k] \bar{\mathcal{U}} = \mathbf{0}$ .

*Proof:* Choose  $t_m[k+i] = \sqrt{c} e^{j((2\pi m l)/(m_T))}$ , for  $i \in I_l$ , then  $(\mathbf{C}_{mn}[k])_{ij} = \begin{cases} c e^{-j((2\pi(m-n)l)/(m_T))}, & i = j, i \in I_l \\ 0, & i \neq j. \end{cases}$  Therefore,

$$\begin{aligned} \bar{\mathbf{U}}^H \mathbf{C}_{mn}[k] \mathbf{U} &= c \sum_{l=1}^{m_T} \mathbf{U}_{I_l}^H \mathbf{U}_{I_l} e^{-j((2\pi(m-n)l)/(m_T))} \\ &= c \mathbf{I} \sum_{l=1}^{m_T} e^{-j((2\pi(m-n)l)/(m_T))} \\ &= \begin{cases} c \mathbf{I}, & m = n \\ \mathbf{0}, & m \neq n \end{cases} \\ \bar{\mathbf{U}}^H \mathbf{C}_{mn}[k] \mathbf{U} &= \sum_{l=1}^{m_T} \bar{\mathbf{U}}_{I_l}^H \mathbf{U}_{I_l} e^{-j((2\pi(m-n)l)/(m_T))} = \mathbf{0}. \end{aligned}$$

Hence, we have  $\mathcal{U}^H \mathbf{A}^H[k] \mathbf{A}[k] \mathcal{U} = c \mathbf{I}$  and  $\bar{\mathcal{U}}^H \mathbf{A}^H[k] \mathbf{A}[k] \bar{\mathcal{U}} = \mathbf{0}$ . ■

An example of the model bases satisfies the condition in the corollary is the Fourier transform model. As long as  $L$  is an interger multiple of  $m_T$ , we can choose  $I_m = \{(mL)/m_T + i | i = 0, \dots, (L/m_T) - 1\}$ . The training sequences chosen in this way are exactly the phase shift training in [7], [21], and [22]. We proved here that such training sequences not only minimize

the variance residual noise as shown in [21] and [22], but also minimize the model error. However, in [14],  $L$  is the number of used subcarriers, then the condition in the corollary may not be satisfied and the phase shift training may not be optimal.

On the other hand, the model error also depends on the selection of the model bases and channel correlation. For example, if any delay  $\tau_i$  is not on the sampling paths, the Fourier transform bases used in [19] and [20] are not the eigenbases of the channel correlation matrix. Depending on the specific channel delay profile that determines  $\mathbf{T}$ , the algorithm may end up with different model error. We will show this effect in the simulation.

To avoid the model error, techniques such as in [13] can be used to find the exact eigenbases of the channel correlation matrix. However, these eigenbases no longer satisfy the condition in *Corollary 1*, which means that we may not be able to find the training sequences that satisfy both conditions for model error and residual noise. With the constraints of the system, such as constellation, a limited number of search is needed to find a sub-optimal solution based on these conditions. We will demonstrate such a scheme later.

2) *Training Sequence Minimizing the Mean Squared Error at the Estimation Point:* Now let us take a look of minimizing the mean squared error at the estimation point. In this section, we will focus on the  $2 \times 2$  case, i.e.,  $m_T = m_R = 2$ . In this case, we have

$$\mathcal{R}_u^{-1}[k] = \begin{bmatrix} \mathbf{R}_{11}^{-1}[k] & \mathbf{R}_{12}^{-1}[k] \\ \mathbf{R}_{21}^{-1}[k] & \mathbf{R}_{22}^{-1}[k] \end{bmatrix}$$

where

$$\mathbf{R}_{11}^{-1}[k] = (\mathbf{U}^H \mathbf{C}_{11}[k] \mathbf{U} - \mathbf{U}^H \mathbf{C}_{12}[k] \mathbf{U} (\mathbf{U}^H \mathbf{C}_{22}[k] \mathbf{U})^{-1} \mathbf{U}^H \mathbf{C}_{21}[k] \mathbf{U})^{-1} \quad (29)$$

$$\mathbf{R}_{22}^{-1}[k] = (\mathbf{U}^H \mathbf{C}_{22}[k] \mathbf{U} - \mathbf{U}^H \mathbf{C}_{21}[k] \mathbf{U} (\mathbf{U}^H \mathbf{C}_{11}[k] \mathbf{U})^{-1} \mathbf{U}^H \mathbf{C}_{12}[k] \mathbf{U})^{-1} \quad (30)$$

$$\mathbf{R}_{12}^{-1}[k] = -(\mathbf{U}^H \mathbf{C}_{11}[k] \mathbf{U})^{-1} \mathbf{U}^H \mathbf{C}_{12}[k] \mathbf{U} \mathbf{R}_{22}^{-1}[k] \quad (31)$$

$$\mathbf{R}_{21}^{-1}[k] = -(\mathbf{U}^H \mathbf{C}_{22}[k] \mathbf{U})^{-1} \mathbf{U}^H \mathbf{C}_{21}[k] \mathbf{U} \mathbf{R}_{11}^{-1}[k]. \quad (32)$$

Then based on (21), the model error at the estimation point can be further written as in (33), shown at the bottom of the page, where

$$\mathbf{S}_{11}[k] = (\mathbf{U} \mathbf{R}_{11}^{-1}[k] \mathbf{U}^H \mathbf{C}_{11}[k] + \mathbf{U} \mathbf{R}_{12}^{-1}[k] \mathbf{U}^H \mathbf{C}_{21}[k] - \mathbf{I}) \bar{\mathbf{U}} \bar{\mathbf{T}} \quad (34)$$

$$\mathbf{S}_{12}[k] = (\mathbf{U} \mathbf{R}_{11}^{-1}[k] \mathbf{U}^H \mathbf{C}_{12}[k] + \mathbf{U} \mathbf{R}_{12}^{-1}[k] \mathbf{U}^H \mathbf{C}_{22}[k]) \bar{\mathbf{U}} \bar{\mathbf{T}} \quad (35)$$

$$\mathbf{S}_{22}[k] = (\mathbf{U} \mathbf{R}_{22}^{-1}[k] \mathbf{U}^H \mathbf{C}_{22}[k] + \mathbf{U} \mathbf{R}_{21}^{-1}[k] \mathbf{U}^H \mathbf{C}_{12}[k] - \mathbf{I}) \bar{\mathbf{U}} \bar{\mathbf{T}} \quad (36)$$

$$\mathbf{S}_{21}[k] = (\mathbf{U} \mathbf{R}_{22}^{-1}[k] \mathbf{U}^H \mathbf{C}_{21}[k] + \mathbf{U} \mathbf{R}_{21}^{-1}[k] \mathbf{U}^H \mathbf{C}_{11}[k]) \bar{\mathbf{U}} \bar{\mathbf{T}}. \quad (37)$$

$$(\boldsymbol{\epsilon}_h)_{mm} = \begin{cases} \sum_{j=1}^L \lambda_j (\|(\mathbf{S}_{11}[k])_{mj}\|^2 + \|(\mathbf{S}_{12}[k])_{mj}\|^2), & \text{for } m \leq L \\ \sum_{j=1}^L \lambda_j (\|(\mathbf{S}_{22}[k])_{(m-L)j}\|^2 + \|(\mathbf{S}_{21}[k])_{(m-L)j}\|^2), & \text{for } m > L \end{cases} \quad (33)$$

Similar to the proof of *Theorem 1* in Appendix I, if the model bases are the linear combinations of the eigenbases that correspond to all the nonzero eigenvalues, then the model error is zero. If the other condition in *Theorem 1* such as  $\mathbf{U}^H \mathbf{C}_{mn} \bar{\mathbf{U}} = 0$  for all  $m, n$  is satisfied, then  $\mathbf{S}_{11}[k] = \mathbf{S}_{22}[k] = -\bar{\mathbf{U}}^T$  and  $\mathbf{S}_{12}[k] = \mathbf{S}_{21}[k] = \mathbf{0}$ . In this case, the model error no longer depends on the training sequence, which makes it easy to choose the model basis.

The residual noise at the estimation point  $(\boldsymbol{\varepsilon}_n)_{mm}$  satisfies the following theorem.

*Theorem 3:* The variance of the residual noise at the estimation point satisfies (38), shown at the bottom of the page, with the equality when  $\mathbf{U}^H \mathbf{C}_{12}[k] \mathbf{U} = \mathbf{0}$ .

The theorem shows that the training sent from the two antennas should be orthogonal regarding the model bases. If this requirement is satisfied, then  $\mathbf{R}_{12}^{-1}[k] = \mathbf{R}_{21}^{-1}[k] = \mathbf{0}$  and  $\mathbf{R}_{11}[k]$  and  $\mathbf{R}_{22}[k]$  become

$$\begin{aligned} \mathbf{R}_{11}^{-1}[k] &= (\mathbf{U}^H \mathbf{C}_{11}[k] \mathbf{U})^{-1} \\ \mathbf{R}_{22}^{-1}[k] &= (\mathbf{U}^H \mathbf{C}_{22}[k] \mathbf{U})^{-1}. \end{aligned}$$

To make the channels estimates corresponding to the two transmitter antennas have the same estimation error, obviously we need  $\mathbf{C}_{11}[k] = \mathbf{C}_{22}[k]$ .

3) *Model Basis Selection and Optimal Training Sequence Design:* From the previous discussion, we know that the challenge in selecting a model basis and design corresponding training sequences are first, that the two are dependent on each other, second, that it is very difficult if not impossible to know the channel correlation matrix. Our approach is to first decouple the two problems. As we have shown, if  $\mathbf{U}^H \mathbf{C}_{mn} \bar{\mathbf{U}} = 0$  for all  $m, n$ , then the model error only depends on the model basis. Then, we can select the model basis by deriving certain bound based on some assumption about the channel, for example, the upper bound of the model error is derived in [15] and [17] for the polynomial model. Then we can focus on minimizing the residual noise. Therefore, for a  $2 \times 2$  system, the training sequences should satisfy the following conditions:

- 1)  $\mathbf{A}_i^H[k] \mathbf{A}_i[k] = c \mathbf{I}$  for  $i = 1, 2$  and  $c$  should be as large as possible;
- 2)  $\mathbf{U}^H \mathbf{C}_{mn}[k] [\mathbf{U} \bar{\mathbf{U}}] = \mathbf{0}$ .

The first condition is to probe all the frequency bins [26]. The second condition is to minimize the residual noise and also together with the first one to separate the model basis selection from the training design.

For the first condition, if a multilevel constellation is used, then we should use the points with the largest energy, for example, the corner points in the quadrature amplitude modulator constellation. However, the phase of the training sequence should be carefully arranged to reduce the peak to average power ratio.

Now, let us take a closer look at the second condition. Here we are actually trying to find out  $\mathbf{d} =$

$(1/c)[t_1^*[t_2], \dots, t_1^*[L-1]t_2[L-1]]^T$ . We can see that this is the solution for the two sets of linear equations. One set is obtained by rearranging  $\mathbf{U}^H \mathbf{C}_{12} \mathbf{U} = \mathbf{0}$  to  $\mathbf{G}_1 \mathbf{d} = \mathbf{0}$ . The other set is obtained by rearranging  $\mathbf{U}^H \mathbf{C}_{12} \bar{\mathbf{U}} = \mathbf{0}$  to  $\mathbf{G}_2 \mathbf{d} = \mathbf{0}$ .

Generally, there may not be a solution to satisfy both sets of equations especially when we consider the constellation constraint of the MCM system. For some special cases as we discussed in Section IV-B1, there exists a solution. In the more general cases, if we cannot find the solution, we can look for the least square solution through the following optimization problem:

$$\begin{aligned} \min_{\mathbf{d}} \quad & \|\mathbf{G}_1 \mathbf{d}\|^2 + \alpha \|\mathbf{G}_2 \mathbf{d}\|^2 \\ \text{subject to} \quad & \|d_i\|^2 = 1, \quad \text{for } i = 0, 1, \dots, L-1. \\ \text{and} \quad & \gamma(\mathcal{R}_u) < \epsilon \end{aligned} \quad (39)$$

where  $\gamma(\mathcal{R}_u)$  is the *condition number* of  $\mathcal{R}_u$  defined by the ratio between the largest eigenvalues and smallest eigenvalues of  $\mathcal{R}_u$ . This constraint is to guarantee that the identification condition is met. The other parameter is  $\alpha$ . This is a parameter chosen to balance between the model error and the residual noise. For low signal-to-noise ratio (SNR),  $\alpha$  should be small to minimize the residual noise. It also depends on the channel; if the channel is very dispersive, the model error tends to be larger, then we should choose larger  $\alpha$ .

Based on the constraint, we have  $d_i = e^{-j\theta_i}$ . Then the optimization problem can be solved by

$$\frac{\partial \|\mathbf{G}_1 \mathbf{d}\|^2 + \alpha \|\mathbf{G}_2 \mathbf{d}\|^2}{\partial \theta_i} = 0. \quad (40)$$

In the case of the polynomial model,  $(\mathbf{G})_{mn}$  are all real numbers. Then the solution to the above condition is  $\theta_i - \theta_n = k\pi$  with  $k$  as any integer. Without loss of the generality, we choose  $d_i = 1$  or  $d_i = -1$ . Then we can conduct a search to find the minimum. Considering the symmetry of the cost function, the maximal number of search is  $\sum_{n=1}^{L/2} C_L^n$ , where  $C_L^n$  denotes the number of combinations to choose  $n$  elements from a set of  $L$  elements. Then considering the condition number constraint, the number of search can be reduced even further. Moreover, all these searches are done offline.

Now let us look at an example: suppose  $L = 11$  and  $l = 3$  for a polynomial model. Table I(a) lists four possible training sequences. In Table I(b), we show different measures of these training sequences. In this example, we choose  $\epsilon = 5$ . We also use the MCM system in Section V-A and a two-ray delay profile as an example to show the model error. The sequence *D1* minimizes  $\|\mathbf{G}_1 \mathbf{d}\|^2$  while the sequence *D2* minimizes  $\|\mathbf{G}_2 \mathbf{d}\|^2$ . Sequence *D3* is the one that minimizes the estimation error at SNR of 10 dB by searching all the possible sequences to minimize (16). It also shows that if we choose  $\alpha = 0.5$ , *D3* actually gives the minimum of  $\|\mathbf{G}_1 \mathbf{d}\|^2 + \mathbf{G}_2 \mathbf{d}\|^2$ .

$$(\boldsymbol{\varepsilon}_n)_{m,m} \geq \begin{cases} \sigma^2 (\mathbf{U} (\mathbf{U}^H \mathbf{C}_{11}[k] \mathbf{U})^{-1} \mathbf{U}^H)_{mm}, & \text{for } m \leq L \\ \sigma^2 (\mathbf{U} (\mathbf{U}^H \mathbf{C}_{22}[k] \mathbf{U})^{-1} \mathbf{U}^H)_{(m-L)(m-L)}, & \text{for } m > L \end{cases} \quad (38)$$



TABLE I  
MEAN SQUARED ERROR FOR DIFFERENT TRAINING SEQUENCES OF  
POLYNOMIAL MODEL ( $L = 11$ ,  $l = 3$ , SNR = 10 dB, TWO-RAY DELAY  
PROFILE  $T_d = 5 \mu\text{s}$ )

	$\mathbf{d}$										
$D1$	-1	1	1	-1	-1	1	1	1	-1	-1	1
$D2$	1	-1	1	1	1	-1	1	1	1	-1	1
$D3$	-1	1	-1	1	-1	-1	1	1	1	-1	1
$D4$	1	-1	1	-1	1	-1	1	-1	1	-1	1

(a)

	$\ \mathbf{G}_1 \mathbf{d}\ ^2$	$\ \mathbf{G}_2 \mathbf{d}\ ^2$	$\ \mathbf{G}_1 \mathbf{d}\ ^2 + 0.5\ \mathbf{G}_2 \mathbf{d}\ ^2$
$D1$	0.0925	2.8580	1.5215
$D2$	0.7394	2.2562	1.8675
$D3$	0.1585	2.6903	1.5036
$D4$	0.3152	2.6466	1.6385

	$\varepsilon_h$ (dB)	$\varepsilon_n$ (dB)	$\varepsilon$ (dB)
$D1$	-36.29	-13.55	-13.55
$D2$	-51.86	-11.99	-11.99
$D3$	-36.24	-13.79	-13.77
$D4$	-51.69	-13.56	-13.56

(b)

In practice, the system design may have more constraints that can further reduce the number of searches needed. For example, if we use the estimator structure in Fig. 3(b), we would want the same estimation error as we slide the window, which means that the training sequences should be symmetric every other samples. Then there is only one possible solution,  $D4$ . It turns out that this sequence performs fairly well, which once again shows that (39) is only a suboptimal solution.

From the above design criteria, we also noticed that the optimal training sequence design is actually designing the relative relationship between the training sequences transmitted from different antennas. Once the relationship is decided, i.e., as we found  $\mathbf{d}$ , we can choose any sequence as  $t_1[k]$  for one transmit antenna and then the other sequence is  $t_2[k] = ((t_1^*[k])/d_k)$ . Since we have the freedom to select  $t_1[k]$ , we should choose it in such a way to meet other design constraints, such as the peak-to-average power ratio.

For the case of more than two transmitter antennas, the design criteria for the optimal training sequence become more difficult to derive. However, it is still favorable that  $\mathbf{U}^H \mathbf{C}_{mn}[k] [\mathbf{U} \mathbf{U}]$  should be as close to  $\mathbf{0}$  as possible, for  $m \neq n$ . We can obtain a sequence of multiples  $t_i[k] t_j[k]$  based on this criteria and then choose the possible training sequences for different antennas.

## V. SIMULATION RESULTS

In this section, we show some simulation results for the model-based channel estimation method. First, we introduce the system parameters used in the simulation.

### A. System Parameters

The bandwidth of the system is  $B_d = 800$  kHz. The number of the subchannels is  $K = 128$ . The length of the cyclic prefix

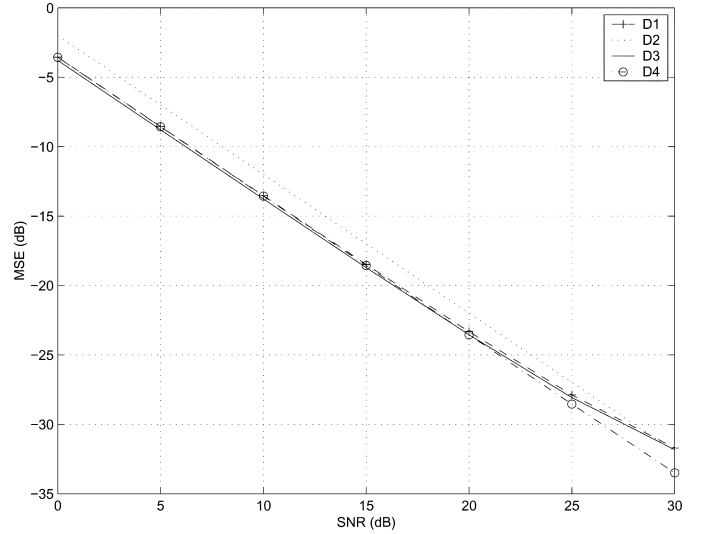


Fig. 4. MSE versus SNR (two-ray,  $T_d = 5 \mu\text{s}$ ,  $L = 11$ , and  $l = 3$ ).

is 32. The four subchannels at each end of one MCM block are used as guarding band. The duration of the MCM block is  $T_f = 200 \mu\text{s}$  and the bandwidth of the subchannel is  $\Delta f = 6.25$  kHz.

The system uses two transmitter antennas and two receiver antennas. Training blocks are sent periodically from both transmitter antennas. After that, the channel estimates are used for decoding the data blocks arriving subsequently. We show the results of different training densities that are defined as the percentage of training blocks of all transmitted MCM blocks. The phase-shift keying (PSK) constellation is used for all the subchannels.

We did simulation for both space-time trellis codes and space-time block codes. A 16-state space-time trellis code using quadrature PSK proposed in [29] is used. One MCM block forms a codeword by forcing the trellis to zero state at the end of block. The Viterbi decoding is then used for decoding. The space-time block code in [31] using 8PSK is also simulated and the decoding scheme in [30] is adopted. In this case, we adopt a Reed-Solomon (RS) code as the outer code to encode each MCM block and then the space time block code is applied across two MCM blocks.

A Rayleigh fading channel with Doppler shift of 40 Hz is used in the simulation. The delay profiles used are TU and HT delay profiles shown in Fig. 2 and two-ray delay profile with two paths separated by  $T_d$ .

### B. Simulation Results

Fig. 4 shows the MSE of the two-ray delay profile of  $T_d = 5 \mu\text{s}$  with different training sequences. The figure further verifies that the design criteria we discussed in Section IV-B. The sequences  $D1$ ,  $D2$ , and  $D3$  have similar performances. While sequence  $D2$  has worse overall performance though it has the smallest model error. From now on, we will use  $D3$  in all the following simulations.

Fig. 5 shows the MSE. Fig. 5(a) shows the estimation error for the TU delay profile and Fig. 5(b) shows the estimation error for the HT delay profile. In Fig. 5(a), we also show the results of the two-ray delay profile with the two paths separated by the

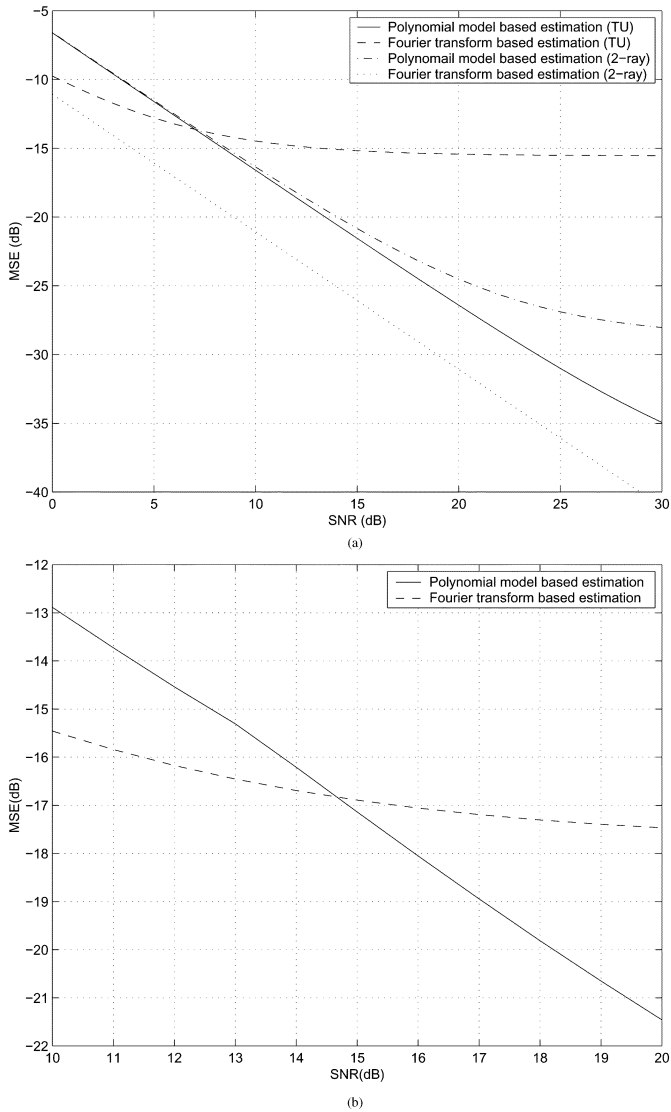


Fig. 5. MSE versus SNR (2 Tx antenna and 2 Rx antenna). (a) TU (b) HT.

$5 \mu\text{s}$  which is the maximal delay of the TU profile. In Fig. 5(a),  $L = 21$  and  $l = 3$  are used for the polynomial model while  $L = 128$  and  $l = 5$  are used for the Fourier-transform-based model. In Fig. 5(b),  $L = 9$  and  $l = 3$  are used for the polynomial model while  $L = 128$  and  $l = 15$  are used for the Fourier-transform-based model. We also use these parameters for the following simulations shown in Figs. 6–8. The model here is selected based the maximal delay spread of the channel just as stated in Section IV-B. The window dimensions for the polynomial model actually can be adapted as suggested in [18]. As shown in Figs. 6–8, the polynomial model has lower estimation error than the Fourier-based model as the SNR goes higher for TU and HT while the Fourier-transform-based model has lower estimation error for two-ray with delay spread of  $5 \mu\text{s}$ . This is because the polynomial model has less model error than the Fourier-transform-based model when there are paths of the channel that are not at the sampling grids of the system. However, for the special case of two-ray with delay spread of  $5 \mu\text{s}$ , both paths are at the sampling grids of the MCM system, the Fourier-transform-based model does not have model error and the minimum mean squared error estimation is achieved.

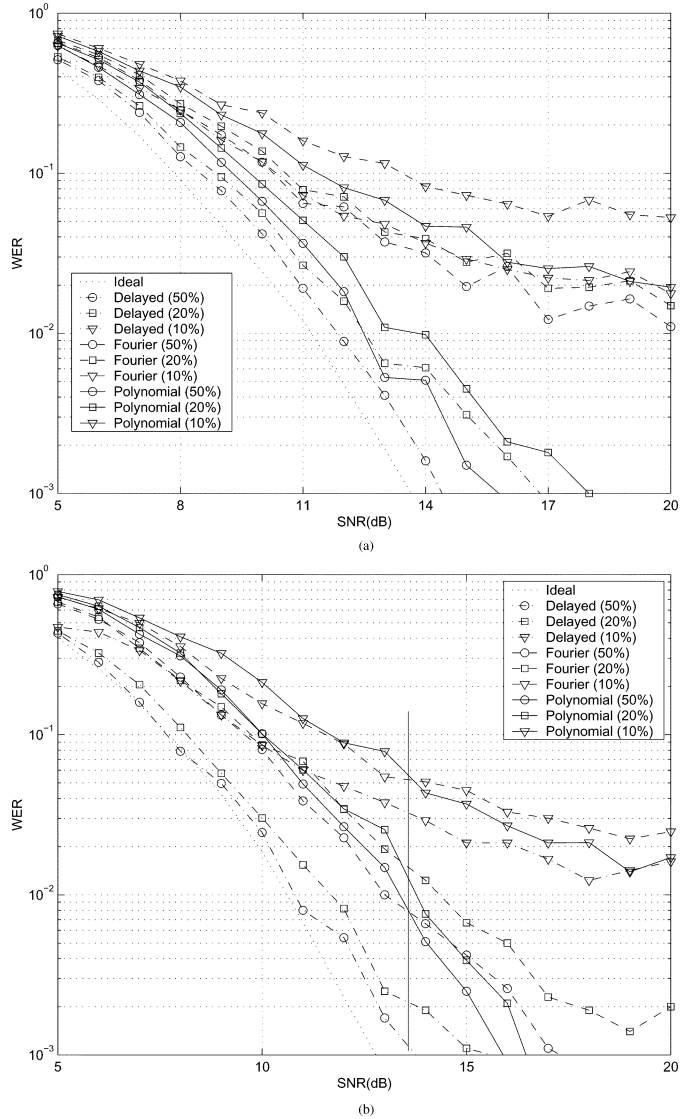


Fig. 6. WER of space-time trellis code versus SNR ( $f_D = 40 \text{ Hz}$ , 2 Tx antenna, and 2 Rx antenna). (a) TU (b) HT.

Fig. 6 shows the word error rate (WER) using different channel information for decoding the 16-state space-time trellis code. The decoding results using the delayed ideal channel information which is the case assuming that the channel estimator can get perfect channel information at the training block are also shown in the figure for comparison. First, we see that for such a Doppler shift, we need a training density more than 20% to avoid the error floor due to the inaccuracy caused by the delay of the channel information. In Fig. 6(a), the polynomial model results are quite close to the results using the delayed ideal channel information which are the best we can achieve. However, the Fourier-transform-based method has an error floor no matter how frequently the channel estimates are updated. In Fig. 6(b), because the HT delay profile has larger dispersion, both estimation schemes have large gap even to the delayed ideal channel information. When SNR exceeds 14 dB, the polynomial model-based method has a better performance than the Fourier transformed-based method, which is consistent with the results of estimation error. There is about 3-dB

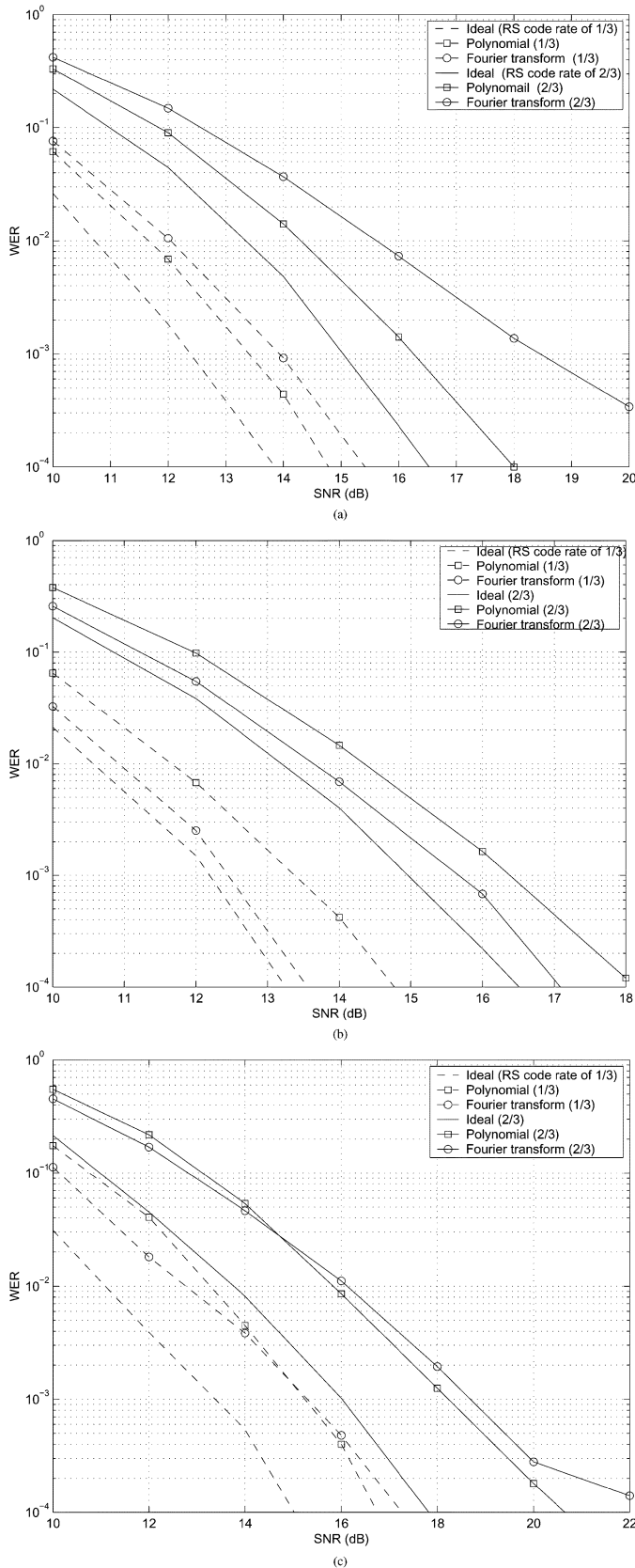


Fig. 7. WER of space-time block code versus SNR (1/3 training  $f_D = 40$  Hz). (a) TU. (b) two-ray,  $T_d = 5 \mu s$ . (c) HT.

difference of using the channel estimation from using the ideal channel information.

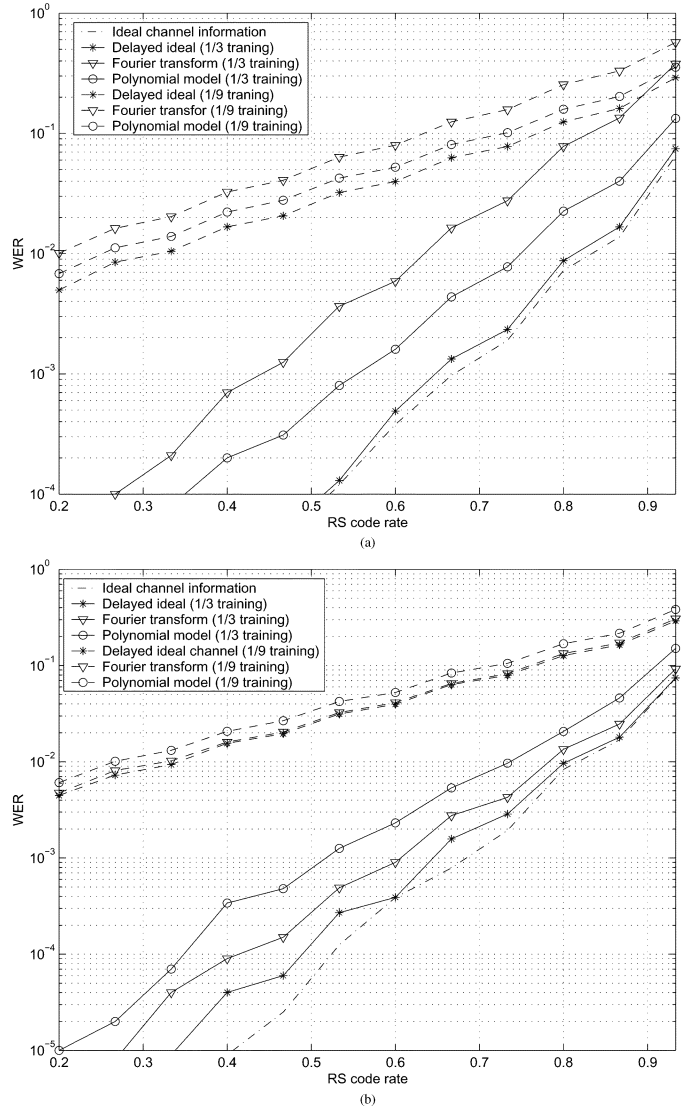


Fig. 8. WER of space-time block code versus RS code rate (SNR = 15 dB,  $f_D = 40$  Hz). (a) TU. (b) two-ray,  $T_d = 5 \mu s$ .

Fig. 7 shows the results of the space-time block code. In this simulation, the information is first encoded by RS code so that one MCM block is one RS codeword. The WERs with different RS code rate for different channel estimation of TU, two-ray, and HT delay profiles are shown in Fig. 7(a)–(c), respectively. The training density in Fig. 7 is 1/3. The results are consistent with Fig. 6. Only in the special case of two-ray delay profile, the Fourier-transform-based model performs better. In the more general cases of TU and HT delay profile, the polynomial model performs better because it has less model error. The difference between using ideal channel information and channel estimation is about 1 dB for TU and two-ray delay profile while 2 dB for HT delay profile.

Fig. 8(a) and (b) shows the WER of RS channel coding rates at SNR of 15 dB with different training densities for both TU delay profile and two-ray delay profile with delay spread of  $5 \mu s$ , respectively. It is shown that the polynomial model performs better than the Fourier transform model in both WER and throughput for TU delay profile while the Fourier transform

model performs better for two-ray delay profile, which is consistent with the results in Fig. 5. The throughput in one data block with 1/3 training is 80% while the throughput of the one with 1/9 training is 30% as the WER is  $10^{-2}$ . Therefore, the overall throughput with higher training density and coding rate is higher than that with lower training density and coding rate.

## VI. CONCLUSION

We have proposed a model-based channel estimation framework for the MIMO MCM systems in this paper. In this framework, the training sequences transmitted simultaneously from all the transmitter antennas, which greatly reduces the overhead of the system. After presenting the challenge of the estimation problem in such a system, we recognized that the model-based approach is not only a method to improve the estimation but also a necessary procedure to identify the channel. The framework of the model-based channel estimation is then derived. Based on this framework, different channel models can be used to design the estimator. Although the knowledge of the statistics of the channel can help us design the estimator, it is not necessary to know the channel statistics. However, the performance of the channel estimation does depend on the channel statistics. Generally speaking, the less dispersive the channel, the better the performance. Moreover, for the very dispersive channel and large number of transmitter antennas, the channel may become unidentifiable. We discussed the identification condition based on the proposed framework. The identifiability depends on both the model and the training sequences and so does the performance of the scheme. We then studied the model selection and training sequence design rules with respect to the estimator structures. The design criteria we proposed try to minimize both the model error and residual estimation noise. The simulation shows that using the proposed scheme we have 0.5- to 3-dB degradation due to the channel estimation error in a two transmitter and two receiver antenna system.

### APPENDIX I PROOF OF THEOREM 1

*Proof:*

$$\begin{aligned}
\bar{\varepsilon}_h &= \frac{1}{m_T L} \text{tr}(\boldsymbol{\varepsilon}_h) \\
&= \frac{1}{m_T L} \text{tr} \left\{ (\mathcal{U} \mathcal{R}_u^{-1}[k] \mathcal{U}^H \mathbf{A}^H[k] \mathbf{A}[k] \bar{\mathcal{U}} \bar{\mathcal{T}} - \bar{\mathcal{U}} \bar{\mathcal{T}})^H \right. \\
&\quad \left. \cdot (\mathcal{U} \mathcal{R}_u^{-1}[k] \mathcal{U}^H \mathbf{A}^H[k] \mathbf{A}[k] \bar{\mathcal{U}} \bar{\mathcal{T}} - \bar{\mathcal{U}} \bar{\mathcal{T}}) \mathcal{Z} \right\} \\
&= \frac{1}{m_T L} \text{tr} \left\{ \bar{\mathcal{T}} (I + \bar{\mathcal{U}} \mathbf{A}^H[k] \mathbf{A}[k] \mathcal{U} \mathcal{R}_u^{-2}[k] \right. \\
&\quad \left. \mathcal{U}^H \mathbf{A}^H[k] \mathbf{A}[k] \bar{\mathcal{U}}) \cdot \bar{\mathcal{T}}^H \mathcal{Z} \right\} \\
&= \sum_{i=1}^L \lambda_i \sum_{m=1}^{m_T} \mu_{(m-1)L+i}. \tag{41}
\end{aligned}$$

Because  $\bar{\mathcal{T}}(I + \bar{\mathcal{U}} \mathbf{A}^H[k] \mathbf{A}[k] \mathcal{U} \mathcal{R}_u^{-2}[k] \mathcal{U}^H \mathbf{A}^H[k] \mathbf{A}[k] \bar{\mathcal{U}}) \bar{\mathcal{T}}^H$  has a form of  $\boldsymbol{\phi}^H \boldsymbol{\phi}$ , then  $\mu_j \geq 0$ , for  $j = 1, 2, \dots, m_T L$ . Furthermore,  $\text{rank}[\bar{\mathcal{T}}(I + \bar{\mathcal{U}} \mathbf{A}^H[k] \mathbf{A}[k] \mathcal{U} \mathcal{R}_u^{-2}[k] \mathcal{U}^H \mathbf{A}^H[k] \mathbf{A}[k] \bar{\mathcal{U}}) \bar{\mathcal{T}}^H] = m_T(L-l)$ . Therefore, at most  $m_T l$

columns in  $\boldsymbol{\phi}$  can be zero vector, which means that at most  $m_T l$   $\mu_j$  equal zero.

If  $\bar{\mathbf{T}}$  has the form of  $[\mathbf{0}_{(L-l) \times l} \tilde{\mathbf{T}}_{(L-l) \times (L-l)}]^T$ , then  $\mathbf{T} = [\tilde{\mathbf{T}}_{l \times l} \mathbf{0}_{l \times (L-l)}]^T$ . Therefore,  $\mathbf{V} = [\mathbf{U} \tilde{\mathbf{T}}^H \bar{\mathbf{U}} \tilde{\mathbf{T}}^H]$ , i.e.,  $\mathbf{U} \tilde{\mathbf{T}}^H$  are the eigenbases corresponding to the first  $l$  eigenvalues of  $\mathbf{R}_H$ . It is obvious then that every column of  $\mathbf{Q}$  is a linear combination of these first  $l$  eigenbases. In this case

$$\begin{aligned}
&\bar{\mathcal{T}} (I + \bar{\mathcal{U}} \mathbf{A}^H[k] \mathbf{A}[k] \mathcal{U} \mathcal{R}_u^{-2}[k] \mathcal{U}^H \mathbf{A}^H[k] \mathbf{A}[k] \bar{\mathcal{U}}) \bar{\mathcal{T}}^H \\
&= \begin{bmatrix} \mathbf{0}_{l \times l} & \mathbf{0} & \cdots & \mathbf{0}_{l \times l} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} + \tilde{\mathbf{T}} \boldsymbol{\Phi}_{11} \tilde{\mathbf{T}} & \cdots & \mathbf{0} & \mathbf{I} + \tilde{\mathbf{T}} \boldsymbol{\Phi}_{1m_T} \tilde{\mathbf{T}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{l \times l} & \mathbf{0} & \cdots & \mathbf{0}_{l \times l} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} + \tilde{\mathbf{T}} \boldsymbol{\Phi}_{m_T 1} \tilde{\mathbf{T}} & \cdots & \mathbf{0} & \mathbf{I} + \tilde{\mathbf{T}} \boldsymbol{\Phi}_{m_T m_T} \tilde{\mathbf{T}} \end{bmatrix}
\end{aligned}$$

where  $\begin{bmatrix} \boldsymbol{\Phi}_{11} & \cdots & \boldsymbol{\Phi}_{1m_T} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Phi}_{m_T 1} & \cdots & \boldsymbol{\Phi}_{m_T m_T} \end{bmatrix}$ . Substituting it to (41), we have

$$\bar{\varepsilon}_h \geq \frac{1}{L} \sum_{i=l+1}^L \lambda_i \sum_{m=1}^{m_T} \mu_{(m-1)L+i}.$$

This lower bound can be further reduced to

$$\bar{\varepsilon}_h \geq \frac{1}{L} \sum_{i=l+1}^L \lambda_i$$

if  $\lambda_i = 0$  for  $i = l+1, \dots, L$  or  $\mu_j = 1$  for  $j = (m-1)L + l+1, \dots, mL$  and  $m = 1, \dots, m_T$ . The first condition means that  $\mathbf{Q}$  is composed by linear combinations of all the eigenbases corresponding to the nonzero eigenvalues. The second condition requires  $\bar{\mathcal{U}} \mathbf{A}^H[k] \mathbf{A}[k] \mathcal{U} \mathcal{R}_u^{-2}[k] \mathcal{U}^H \mathbf{A}^H[k] \mathbf{A}[k] \bar{\mathcal{U}} = \mathbf{0}$ . ■

### APPENDIX II PROOF OF THEOREM 2

*Proof:*

$$\bar{\varepsilon}_n = \frac{\sigma^2}{m_T L} \text{tr} \{ \mathcal{U} \mathcal{R}_u^{-1}[k] \mathcal{U}^H \} = \frac{\sigma^2}{m_T L} \text{tr} \{ \mathcal{R}_u^{-1}[k] \}.$$

Denote the  $\gamma_i$  as the eigenvalues of  $\mathcal{R}_u[k]$ , for  $i = 1, 2, \dots, 2L$ . Then according to *Jensen's inequality*, we have

$$\bar{\varepsilon}_n = \frac{\sigma^2}{m_T L} \sum_{i=1}^{m_T l} \frac{1}{\gamma_i} \geq \frac{m_T l \sigma^2}{m_T L} \frac{1}{\frac{1}{m_T L} \sum_{i=1}^{m_T l} \gamma_i}. \tag{42}$$

The equality holds if and only if  $\gamma_i$  equal to a constant. This means that  $\mathcal{R}_u[k] = c\mathbf{I}$ . Therefore,  $\sum_{i=1}^{m_T L} \gamma_i = m_T L c$ . Then (42) becomes

$$\bar{\varepsilon}_n \geq \frac{l \sigma^2}{L c}. \quad \blacksquare$$

APPENDIX III  
PROOF OF THEOREM 3

*Proof:* First we have

$$\begin{aligned}\mathbf{R}_{11}^{-1}[k] &= (\mathbf{U}^H \mathbf{C}_{11}[k] \mathbf{U})^{-1} + (\mathbf{U}^H \mathbf{C}_{11}[k] \mathbf{U})^{-1} \\ &\quad \mathbf{U}^H \mathbf{C}_{12}[k] \mathbf{U} \mathbf{R}_{22}^{-1} \mathbf{U}^H \mathbf{C}_{21}[k] \mathbf{U} (\mathbf{U}^H \mathbf{C}_{11}[k] \mathbf{U})^{-1} \\ \mathbf{R}_{22}^{-1}[k] &= (\mathbf{U}^H \mathbf{C}_{22}[k] \mathbf{U})^{-1} + (\mathbf{U}^H \mathbf{C}_{22}[k] \mathbf{U})^{-1} \\ &\quad \mathbf{U}^H \mathbf{C}_{21}[k] \mathbf{U} \mathbf{R}_{11}^{-1} \mathbf{U}^H \mathbf{C}_{12}[k] \mathbf{U} (\mathbf{U}^H \mathbf{C}_{22}[k] \mathbf{U})^{-1}.\end{aligned}$$

It is obvious that  $\mathbf{R}_{11}[k]$  and  $\mathbf{R}_{22}[k]$  are both of the form  $\phi^H \phi$ . Therefore, the diagonal elements of these two matrices are all greater than or equal to zero.

Then, we have

$$(\epsilon_n)_{m,m} \geq \sigma^2 (\mathbf{U} (\mathbf{U}^H \mathbf{C}_{11}[k] \mathbf{U})^{-1} \mathbf{U}^H)_{mm},$$

for  $m \leq L$

or

$$(\epsilon_n)_{m,m} \geq \sigma^2 (\mathbf{U} (\mathbf{U}^H \mathbf{C}_{22}[k] \mathbf{U})^{-1} \mathbf{U}^H)_{(m-L)(m-L)},$$

for  $m > L$ .

The equality holds if and only if  $\mathbf{U}^H \mathbf{C}_{12}[k] \mathbf{U} = \mathbf{0}$ . ■

REFERENCES

- [1] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, no. 5, pp. 5–14, May 1990.
- [2] L. J. Cimini Jr, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, vol. 33, no. 7, pp. 665–675, Jul. 1985.
- [3] P. A. Bello, "Characterization of randomly time-Variant linear channels," *IEEE Trans. Commun.*, vol. COM-11, no. 4, pp. 360–393, Dec. 1963.
- [4] P. Soma, D. S. Baum, V. Erceg, R. Krishamoorthy, and A. J. Paulraj, "Analysis and modeling of multiple-input multiple-output (MIMO) radio channel based on outdoor measurements conducted at 2.5 GHz for fixed BWA application," in *ICC 2002*, Apr. 2002, pp. 272–276.
- [5] R. A. Ziegler and J. M. Cioffi, "Estimation of time-varying digital radio channel," *IEEE Trans. Veh. Technol.*, vol. 41, no. 2, pp. 134–151, May 1992.
- [6] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Båörjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931–939, Jul. 1998.
- [7] B. Steiner, "Time domain channel estimation in multicarrier-CDMA mobile radio system concepts," in *Multi-Carrier Spread-Spectrum*. Norwell, MA: Kluwer, 1997, pp. 153–160.
- [8] T. A. Thomas, F. W. Vook, and K. Baum, "Least-squares multiuser frequency-domain channel estimation for broad-band wireless communication systems," presented at the 37th Allerton Conf., Monticello, IL, Sep. 1999.
- [9] V. Mignone and A. Morello, "CD3-OFDM: A Novel demodulation scheme for fixed and mobile receivers," *IEEE Trans. Commun.*, vol. 44, no. 9, pp. 1141–1151, Sep. 1996.
- [10] Y. Li, L. J. Cimini Jr, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902–915, Jul. 1998.
- [11] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1207–1215, Jul. 2000.
- [12] Y. Li and N. R. Sollenberger, "Adaptive antenna arrays for OFDM systems with cochannel interference," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 217–229, Feb. 1999.
- [13] B. Yang, K. B. Letaif, R. S. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 467–479, Mar. 2001.
- [14] L. Deneire, P. Vandenameele, L. van der Perre, B. Gyselinckx, and M. Engels, "A low-complexity ML channel estimator for OFDM," *IEEE Trans. Commun.*, vol. 51, no. 2, pp. 135–140, Feb. 2003.
- [15] D. K. Borah and B. D. Hart, "A robust receiver structure for time-varying, frequency-flat Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 3, pp. 360–364, Mar. 1999.
- [16] —, "Frequency-selective fading channel estimation with a polynomial time-varying channel model," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 862–873, Jun. 1999.
- [17] X. Wang and K. J. R. Liu, "Channel estimation for multicarrier modulation systems using a time-frequency polynomial model," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1045–1049, Jul. 2002.
- [18] —, "An adaptive channel estimation algorithm using time-frequency polynomial model for OFDM with fading multipath channels," *EURASIP J. Appl. Signal Process. (Special Issue on 3G Wireless Communications and Beyond)*, vol. 2002, no. 8, pp. 818–830, Aug. 2002.
- [19] Y. Li and N. Seshadri, "Channel estimation for OFDM systems with transmitter diversity in mobile wireless channels," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 461–471, Mar. 1999.
- [20] Y. Li, J. H. Winters, and N. R. Sollenberger, "MIMO-OFDM for wireless communications: Signal detection with enhanced channel estimation," *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 1471–1477, Sep. 2002.
- [21] T.-L. Tung and K. Yao, "Channel estimation and optimal power allocation for a multiple-antenna OFDM systems," *EURASIP J. Appl. Signal Process.*, pp. 330–339, Mar. 2002.
- [22] I. Barhumi, G. Leus, and M. Moonen, "Optimal training sequences for channel estimation in MIMO OFDM systems in mobile wireless channels," in *Int. Zurich Seminar on Broad-band Communications, Access, Transmission, Networking*, 2002, pp. 44–1–44–6.
- [23] W. Bai, C. He, L. G. Jiang, and X. X. Li, "Robust channel estimation in MIMO OFDM systems," *Electron. Lett.*, vol. 39, no. 2, pp. 242–244, Jan. 2003.
- [24] H. Zhu, B. Farhang-Boroujeny, and C. Schlegel, "Pilot embedding for joint channel estimation and data detection in MIMO communication systems," *IEEE Commun. Lett.*, vol. 7, no. 1, pp. 30–32, Jan. 2003.
- [25] M. Luise, R. Reggiannini, and G. M. Vietta, "Blind equalization/detection for OFDM signals over frequency-selective channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1568–1578, Oct. 1998.
- [26] C. Tellambura, M. G. Parker, Y. J. Guo, S. J. Shepherd, and S. K. Barton, "Optimal sequences for channel estimation using discrete Fourier transform techniques," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 230–237, Feb. 1999.
- [27] J. H. Winters, "Signal acquisition and tracking with adaptive arrays in the digital mobile radio system IS-54 with flat fading," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 377–384, Nov. 1993.
- [28] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Process. Mag.*, vol. 14, no. 6, pp. 49–83, Nov. 1997.
- [29] V. Tarokh and A. Naguib, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.
- [30] S. M. Alamouti, "A simple transmit diversity techniques for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.
- [31] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 451–460, Mar. 1999.
- [32] A. Narula, M. D. Trott, and G. W. Wornell, "Performance limits of coded diversity methods for transmitter antenna arrays," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2418–2433, Nov. 1999.
- [33] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.
- [34] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [35] J. G. Proakis, *Digital Communications*, 2nd ed. New York: McGraw-Hill, 1989.
- [36] E. W. Cheney, *Introduction to Approximation Theory*. New York: McGraw-Hill, 1966.
- [37] H. N. Mhaskar, *Introduction to the Theory of Weighted Polynomial Approximation*, Singapore: World Scientific, 1996.



**Xiaowen Wang** received the B.S. degree in electronics engineering with highest honors from Tsinghua University, Beijing, China, in 1993, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, MD, in 1999 and 2000, respectively.

From 1993 to 1996, she was a Teaching Assistant with Tsinghua University, Beijing, China. From 1996 to 2000, she was a Research Assistant with the University of Maryland, College Park, MD. Since 2000, she has been with the Wireless Systems Research Department,

Agere Systems (formerly Bell Laboratories, Lucent Technologies, Microelectronics). Her research interests include adaptive digital signal processing, wireless communications, and networking.

Dr. Wang served as a Guest Editor for *EURASIP Journal on Applied Signal Processing*, *Special Issue on MIMO Communications and Signal Processing* and Technical Committee Member of ICC'02. She was the recipient of the Graduate School Fellowship from University of Maryland.



**K. J. Ray Liu** (S'87–M'90–SM'93–F'03) received the B.S. degree from the National Taiwan University and the Ph.D. degree from the University of California, Los Angeles, in 1983 and 1990, respectively, both in electrical engineering.

He is a Professor and Director of Communications and Signal Processing Laboratories of the Electrical and Computer Engineering Department and Institute for Systems Research, University of Maryland, College Park. His research contributions encompass broad aspects of wireless communications and

networking; information forensics and security; multimedia communications and signal processing; signal processing algorithms and architectures; and bioinformatics, in which he has published over 350 refereed papers.

Dr. Liu is the recipient of numerous honors and awards including IEEE Signal Processing Society 2004 Distinguished Lecturer, the 1994 National Science Foundation Young Investigator Award, the IEEE Signal Processing Society's 1993 Senior Award (Best Paper Award), the IEEE 50th Vehicular Technology Conference Best Paper Award, Amsterdam, The Netherlands, 1999. He also received the George Corcoran Award in 1994 for outstanding contributions to electrical engineering education and the Outstanding Systems Engineering Faculty Award in 1996 in recognition of outstanding contributions in interdisciplinary research, both from the University of Maryland. He is the Editor-in-Chief of the *IEEE Signal Processing Magazine*, the prime proposer and architect of the new IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, and was the founding Editor-in-Chief of *EURASIP Journal on Applied Signal Processing*. He is a member of the Board of Governors of the IEEE Signal Processing Society.