

Performance Analysis for Adaptive Channel Estimation Exploiting Cyclic Prefix in Multicarrier Modulation Systems

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

Abstract—Multicarrier modulation (MCM) has gained growing interest in high-data-rate communications in both wire and wireless environments. The channel estimation is a crucial aspect in MCM systems. In this paper, we first present a novel adaptive channel estimation algorithm exploiting the channel information contained in the cyclic prefix of the MCM system. In simulation, we show that this algorithm outperforms the existing scheme. Then we theoretically analyze the performance of the adaptive algorithm considering both channel noise and decision error. We prove that the algorithm is guaranteed to converge with proper loading. Computer simulation shows that our analytical results are quite close to the simulation.

Index Terms—Channel estimation, cyclic prefix, multicarrier modulation (MCM), performance analysis.

I. INTRODUCTION

MULTICARRIER modulation (MCM) is now considered an effective technique for both wire and wireless communications [1]. MCM partitions the entire bandwidth into several parallel subchannels by dividing the transmit data into several parallel low-bit-rate data streams to modulate the carriers corresponding to those subchannels. It is a scheme compatible to the famous water-filling theorem [3] and provides an optimal way for channel capacity usage by adjusting the bit rate and transmit power according to the conditions of subchannels. MCM is also a block-oriented modulation scheme, which results in a relative longer symbol duration and produces greater immunity to impulse noise and intersymbol interference (ISI). Because of these advantages, MCM is considered a promising technique in digital subscriber line (xDSL), digital video/audio broadcasting, and wireless communications [1], [4].

The channel information plays an important role in the implementation of MCM systems. It is essential to bit and power allocations and signal detections. Without perfect knowledge of channel parameters, the MCM system either cannot work or may incur significant performance loss. Some techniques, such as differential phase-shift keying (PSK) modulation, can be used to eliminate the need for channel information. However,

it causes 3-4-dB signal-to-noise ratio (SNR) loss compared with coherent demodulation if channel information is known.

In applications such as xDSL, some training processes are performed to estimate the channel before the communication is set up. Then, this channel estimate is used throughout the entire communication [3]. If the channel changes, retraining is required to track the variation. In wireless applications, the channel variation is assumed continuous, then pilot symbols are used to catch the channel variation [8]. However, in order to estimate the channel more efficiently, people are trying to estimate the channel information directly from the transmitted data.

We propose an adaptive channel estimation algorithm by exploiting the cyclic prefix in the MCM system [9], [10]. The cyclic prefix used in MCM systems is originally designed to reduce ISI. However, it is nothing but a repeated part of the transmit data which can be used for channel estimation. Based on this observation, we propose a block recursive least-square (RLS) algorithm to estimate the channel, adaptively exploiting the information in cyclic prefix. The algorithm uses decision directed samples, and hence, no extra training is needed. The simulation shows that by using the proposed adaptive algorithm, the MCM system performs more robustly than the existing system with adaptive equalization [3]. In this paper, we will present the adaptive channel estimation algorithm and analyze the performance of it theoretically.

A lot of research has been done on the performance analysis of the decision-directed estimation and equalization schemes [11], [13]–[19], [21]. In [21], the system identification problem with noisy input is visited. In [17]–[19], the error propagation through the fixed decision feedback equalizer is analyzed. However, in our algorithm, the linear equalizer is used and adapted using the decision-directed samples. Blind equalization that does not need extra training is studied for linear equalizer adaptation in [11], [13]–[15], and for decision feedback equalizer adaptation in [16]. In such blind schemes, the channel inverse is estimated from the channel output and the decision-directed samples are used as the desired output of the channel inverse filter. The understanding of these algorithms is that due to the nonlinearity of the decision-directed scheme, the cost function usually has more than one local minima, and some kinds of smart initialization schemes must be used to force the system to converge to the global minimum.

The goal of the proposed adaptive estimation algorithm is to estimate the channel itself, not the channel inverse. The decision-directed samples are treated as the filter input data, while

Paper approved by H. Liu, the Editor for Synchronization and Equalization of the IEEE Communications Society. Manuscript received January 10, 2000; revised January 16, 2002. This paper was presented in part at Globecom, San Francisco, CA, November 2000.

X. Wang is with Agere Systems, Allentown, PA 18109 USA.

K. J. R. Liu is with the Electrical Engineering Department and Institute for Systems Research, University of Maryland, College Park, MD 20742 USA.

Digital Object Identifier 10.1109/TCOMM.2002.807610

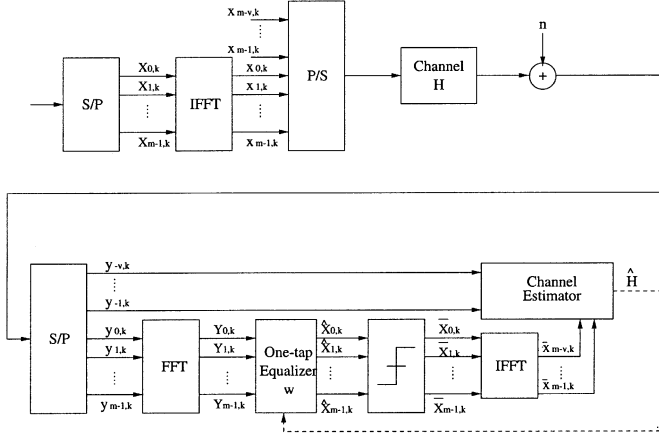


Fig. 1. MCM system with cyclic prefix and adaptive channel estimation.

the channel output is treated as the desired filter output. The decision error in the blind equalization algorithms only appears in the cross-correlation vector of the Wiener–Hopf equation, while in our adaptive estimation algorithm, it appears in both data correlation matrix and cross-correlation vector. In this paper, we are trying to consider both the effect of channel noise and decision error. The problem becomes complicated because the decision error and the channel estimation error affect each other through a closed feedback loop constituted by the signal detection, channel estimation, and equalization. We try to separate the analysis into two parts. First, we analyze the impact of decision error on the channel estimation. Then, we study the impact of estimation error on the signal detection and try to derive the symbol error rate (SER) based on this analysis. Finally, a recursive mapping of SER is constructed using the results of the these two parts. The convergence of this recursive mapping is considered to draw our final conclusions.

The rest of the paper is organized in the following way. First, we will present the MCM system and the adaptive channel estimation algorithm. Then we will do a performance analysis following the outline given above. Because some approximation is applied in the analysis, we will give the computer simulation examples to verify the validity of the theoretical analysis. Finally, we present our conclusions.

II. MCM SYSTEM AND ADAPTIVE CHANNEL ESTIMATION ALGORITHM

In this section, we will first present our adaptive channel estimation scheme using cyclic prefix and then compare it with the existing adaptive equalization scheme in [3]. We will show in simulation that the proposed scheme outperforms the existing scheme.

A. MCM System Using Cyclic Prefix

Fig. 1 shows a MCM system using cyclic prefix with adaptive channel estimation. The system has $m/2$ complex parallel subchannels. The input data can first be coded and interleaved and then are buffered to blocks. Each block of data is then divided into $m/2$ bit streams and mapped to some complex constellation points, $X_{i,k}$, $i = 0, \dots, m/2 - 1$ at block k . The modulation is implemented by m -point inverse discrete Fourier transform

(IDFT) on $\mathbf{X}_k = [X_{0,k} X_{1,k} \dots X_{m-1,k}]$ where the last $m/2$ samples are just the conjugates of the first $m/2$ samples, and therefore, the modulated time domain signal is real, which is

$$x_{i,k} = \frac{1}{\sqrt{m}} \sum_{l=0}^{m-1} X_{l,k} e^{j\frac{2\pi li}{m}}, \quad i = 0, \dots, m-1. \quad (1)$$

The transmitted energy $\Gamma_i \triangleq E[|X_{i,k}|^2]$ and bit rate for different subchannels can be allocated according to the channel condition.

A cyclic prefix $x_k^{(f)} = [x_{-v,k} \dots x_{-1,k}]^T$ is constructed by $x_{-i,k} = x_{m-i,k}$, $i = 1, \dots, v$ and transmitted before \mathbf{x}_k . At the receiver, the prefix part $\mathbf{y}_k^{(f)} = [y_{-v,k} \dots y_{-1,k}]^T$ is discarded. The demodulation is performed only on $\mathbf{y}_k = [y_{0,k} y_{1,k} \dots y_{m-1,k}]^T$ by the DFT operation. The demodulated data is $\mathbf{Y}_k = [Y_{0,k} Y_{1,k} \dots Y_{m-1,k}]^T$ with

$$Y_{i,k} = \frac{1}{\sqrt{m}} \sum_{l=0}^{m-1} y_{l,k} e^{j\frac{2\pi li}{m}}, \quad i = 0, \dots, m-1. \quad (2)$$

The channel is usually modeled as a finite impulse response (FIR) filter with $v+1$ real taps. The impulse response of the channel is $\mathbf{h} = [h_0, h_1, \dots, h_v]^T$. The channel noise $n_{i,k}$, $i = -v, \dots, -1, 0, \dots, m-1$ is assumed to be independent identically distributed (i.i.d.) real Gaussian distribution with zero mean and variance σ^2 . Then the relationship between the channel input and output can be expressed as

$$y_{i,k} = \begin{cases} \sum_{l=0}^v h_l x_{i-l,k} + n_{i,k}, & i = 0, \dots, m-1 \\ \sum_{l=0}^{v+i} h_l x_{i-l,k} \\ \quad + \sum_{l=v+i+1}^v h_l x_{m-l,k-1} \\ \quad + n_{i,k}, & i = -v, \dots, -1. \end{cases} \quad (3)$$

From (3), we can see that there is no interference from the previous blocks in the received signal \mathbf{y}_k . It shows that the cyclic prefix reduces the ISI between \mathbf{X}_k 's and hence, the subchannels can be viewed as independent with each other, i.e.,

$$Y_{i,k} = X_{i,k} H_i + N_{i,k} \quad (4)$$

where $H_i = \sum_{l=0}^v h_l e^{-j((2\pi li)/m)}$ and $N_{i,k} = (1/\sqrt{m}) \sum_{l=0}^{m-1} n_{l,k} e^{-j((2\pi li)/m)}$ is the noise of the i th subchannel, which is also with zero mean and variance σ^2 and independent with that of other subchannels.

For the independent subchannel of (4), only a one-tap equalizer T_i is needed to get the estimation of $X_{i,k}$ from $Y_{i,k}$, i.e.,

$$\hat{X}_{i,k} = Y_{i,k} \cdot W_i \quad (5)$$

where

$$W_i = \frac{1}{H_i}. \quad (6)$$

Then the decision is made upon $\hat{X}_{i,k}$, resulting in $\bar{X}_{i,k} = q(\hat{X}_{i,k})$, where $q(\cdot)$ is some type of quantization function. Then the decoding and deinterleaving are done based on $\bar{X}_{i,k}$, if any of coding and interleaving are used.

B. Adaptive Channel Estimation Algorithm

In the MCM system, usually the received cyclic prefix part $\mathbf{y}_k^{(f)}$ is discarded. However, it is found that if all the prefix parts concatenate together as a pair of sequences $\{x_i^{(f)}\} = \{\dots x_{-v,k-1} \dots x_{-1,k-1} x_{-v,k} \dots x_{-1,k} \dots\}$ and $\{y_i^{(f)}\} = \{\dots y_{-v,k-1} \dots y_{-1,k-1} y_{-v,k} \dots y_{-1,k} \dots\}$, the relationship between these two satisfies [9]

$$y_i^{(f)} = x_i^{(f)} * h_i + n_i. \quad (7)$$

Based on this equation, a block RLS algorithm can be adopted to adaptively estimate the channel by directly solving (7).

First, the estimated transmitted cyclic prefix is obtained by performing IDFT on the decision-directed sample $\bar{X}_{i,k}$,

$$\bar{x}_{i,k} = \frac{1}{\sqrt{m}} \sum_{l=0}^{m-1} \bar{X}_{j,k} e^{j \frac{2\pi l i}{m}}, \quad i = -v, \dots, -1. \quad (8)$$

The estimated transmitted cyclic prefix can also be constructed from the decoding data. However, in this paper, we only use the decision-directed ones.

Then the estimated correlation matrix and the cross-correlation vector are formed as

$$\bar{\Phi}(k) = \mu_1 \bar{\Phi}(k-1) + \sum_{l=1}^v \mu_2^{l-1} \bar{\mathbf{u}}_l(k) \bar{\mathbf{u}}_l^H(k) \quad (9)$$

$$\bar{\mathbf{z}}(k) = \mu_1 \bar{\mathbf{z}}(k-1) + \sum_{l=1}^v \mu_2^{l-1} \bar{\mathbf{u}}_l(k) y_{-l,k} \quad (10)$$

where $\bar{\mathbf{u}}_l(k) = [\bar{x}_{-l,k} \dots \bar{x}_{-v,k} \bar{x}_{-1,k-1} \dots \bar{x}_{-l,k-1}]^T$ is the estimated data vector. μ_1 and μ_2 are forgetting factors across blocks and within blocks, respectively. These two factors should be both equal to or less than one.

The channel estimation then is obtained as

$$\hat{\mathbf{h}}(k) = \bar{\Phi}^{-1}(k) \bar{\mathbf{z}}(k) \quad (11)$$

where $\hat{\mathbf{h}}(k) = [\hat{h}_{0,k} \hat{h}_{1,k} \dots \hat{h}_{v,k}]^T$.

Here we also would like to define the ideal data correlation matrix and the cross-correlation vector with perfect knowledge of the transmitted data for the future discussion. The ideal data correlation matrix is

$$\Phi(k) = \sum_{n=1}^k \mu_1^{k-n} \sum_{l=1}^v \mu_2^{l-1} \mathbf{u}_l(k) \mathbf{u}_l^H(k) \quad (12)$$

and the ideal cross-correlation vector is

$$\mathbf{z}(k) = \sum_{n=1}^k \mu_1^{k-n} \sum_{l=1}^v \mu_2^{l-1} \mathbf{u}_l(k) y_{-l,k} \quad (13)$$

where $\mathbf{u}_l(k) = [x_{-l,k} \dots x_{-v,k} x_{-1,k-1} \dots x_{-l,k-1}]^T$ is the ideal data vector.

Clearly, the above algorithm is a decision-directed scheme. In order to start the algorithm, we need to do some initialization. At initialization, we send an initial training to get the initial channel response $H_{i,0}$. Using a quadrature amplitude modulation (QAM) constellation for all subchannels, the loading

is done according to $H_{i,0}$'s with the following requirement on SER:

$$P_{e,i}(0) = P_e(0) = 4Q\left(\frac{d_i \|H_{i,0}\|}{\sqrt{2}\sigma}\right), \quad i = 0, \dots, m-1 \quad (14)$$

where $P_{e,i}(0)$ is the initial SER and d_i is the minimum distance between the constellation points of the i th subchannel, respectively. $P_e(0)$ is some preset required value of SER to control the decision error. The Q-function is defined as $Q(x) = \int_x^\infty (1/\sqrt{2\pi}) e^{-x^2/2} dx$. Then

$$d_i = \frac{\sqrt{2}\sigma}{\|H_{i,0}\|} Q^{-1}\left(\frac{P_e(0)}{4}\right). \quad (15)$$

This optimization problem is subjected to the energy constraint, i.e.,

$$E_x \triangleq E[\|x_{i,k}\|^2] = \frac{1}{m} \sum_{i \in U} \Gamma_i \leq C \quad (16)$$

where U is the set of all the used subchannels.

For QAM constellation

$$\Gamma_i = \frac{M_i - 1}{6} d_i^2 \quad (17)$$

with M_i as the number of constellation points used in the i th subchannels.

As the loading is done, the data are transmitted according to the bit and energy allocated to each subchannel. The receiver then performs the following adaptive channel estimation algorithm.

Input: $\mathbf{y}_k^{(f)}$ and \mathbf{Y}_k .

Known parameters: Γ_i and σ_i^2 .

Selecting parameters: μ_1 and μ_2 .

Initialization: $k = 0$, an initial training process is used to initialize $\hat{\mathbf{h}}(0)$ and $\bar{\Phi}(0)$.

Computation: $k = 1, 2, 3, \dots$

- 1) $\hat{H}_{i,k-1} = \sum_{l=0}^v \hat{h}_{l,k-1} e^{-j2\pi il/m}$, $\hat{W}_{i,k-1} = 1/\hat{H}_{i,k-1}$.
- 2) $\hat{X}_{i,k} = Y_{i,k} \hat{W}_{i,k-1}$, $i = 0, \dots, m-1$.
- 3) $\bar{x}_{i,k} = (1/\sqrt{m}) \sum_{l=0}^{m-1} q(\hat{X}_{l,k}) e^{j2\pi il/m}$, $i = m-v, \dots, m-1$.
- 4) $\bar{\Phi}(k) = \mu_1 \bar{\Phi}(k-1) + \sum_{l=1}^v \mu_2^{l-1} \bar{\mathbf{u}}_l(k) \bar{\mathbf{u}}_l^H(k)$,
 $\bar{\mathbf{z}}(k) = \mu_1 \bar{\mathbf{z}}(k-1) + \sum_{l=1}^v \mu_2^{l-1} \bar{\mathbf{u}}_{l-1}(k) y_{-l,k}$.
- 5) $\hat{\mathbf{h}}(k) = \bar{\Phi}^{-1}(k) \bar{\mathbf{z}}(k)$.

Here steps 4 and 5 can be replaced by existing fast RLS algorithms [20].

C. Comparison With Existing Adaptive Equalization Scheme

In the MCM system, the subchannels are considered independent. An adaptive equalization scheme for single-channel systems then can be applied for each subchannel. As described in [3], such an adaptive equalization scheme is shown in Fig. 2. The equalizer coefficient is updated by

$$W_{i,k} = W_{i,k-1} + \mu(\bar{X}_{i,k} - \hat{X}_{i,k})^* Y_{i,k}. \quad (18)$$

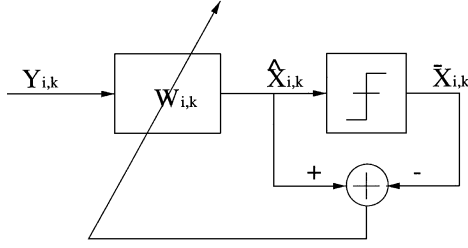
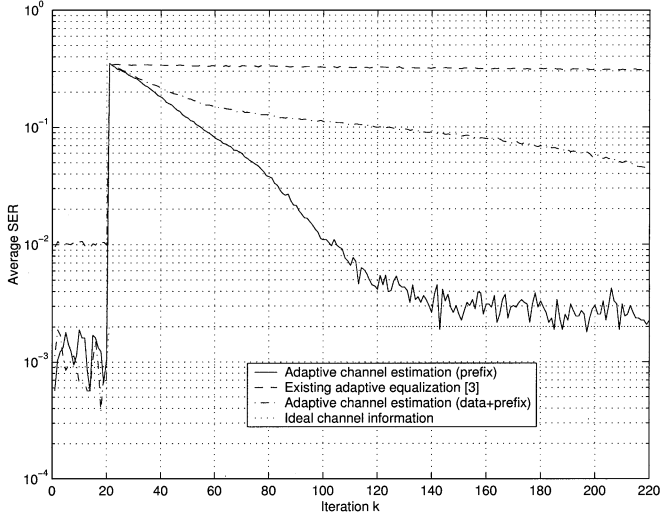


Fig. 2. Adaptive equalization scheme.

Fig. 3. SER iteration compared with existing adaptive equalization ($v = 64$, $\sigma^2 = 0.01$, $P_e(0) = 10^{-3}$, $\mu = 0.01$, $\mu_1 = 0.7$, and $\mu_2 = 1$).

In [11] and [13], the convergence of such an adaptive scheme is proved under the condition that the channel noise is small and no decision error exists. However, in Fig. 3, our simulation shows that such a system fails to follow the channel variation while our adaptive channel estimation works.

The MCM system used in the simulation has 256 complex subchannels. The average transmit energy is one. Initially, the channel transfer function used is $H(D) = (0.02 + 0.16D^2)/(1 - 1.5D + 0.54D^2)$ and the initial loading is done according to it. At the 20th block, the channel is changed to $H(D) = (0.05 + 0.27D^2)/(1 - 1.4D + 0.5D^2)$. The figure shows the averaged SER, which is defined as

$$\tilde{p}(k) \triangleq \frac{1}{|U|} \sum_{i \in U} P_{e,i}(k) \quad (19)$$

where $P_{e,i}(k)$ is the SER of the i th subchannel of the k th block. The step size used in (18) is $\mu = 0.01$ while the forgetting factors used in the proposed adaptive channel estimation algorithm are $\mu_1 = 0.7$ and $\mu_2 = 1$. It is clear that the proposed adaptive channel estimation algorithm can follow such a channel variation, whereas the existing scheme fails to do so.

In the figure, we also show the result of the proposed adaptive algorithm using the data part \mathbf{y}_k together with the cyclic prefix \mathbf{y}_k^f to form the estimated correlation matrix $\hat{\Phi}(k)$ and cross-correlation vector $\hat{\mathbf{z}}(k)$. It is shown in the figure that the algorithm converges much slower than the one that only uses

the cyclic prefix. In the following sections, we will discuss the performance analysis problem of the proposed adaptive channel estimation algorithm and try to explain the results.

III. PERFORMANCE ANALYSIS WITH EXISTENCE OF DECISION ERROR

From this section, we begin the analysis of the proposed channel estimation algorithm. First, we study the impact of decision error on the channel estimation. As described in Section II-B, the decision-directed samples are used as the estimated transmitted cyclic prefix part. If we can get the perfect samples of the cyclic prefix, then we know from the literature [20] that the algorithm will converge to an unbiased estimation linearly, and the convergence rate is determined by the eigenvalue spread of the data correlation matrix. However, the detected signals are used as the estimation of the transmitted cyclic prefix and the decision error would affect the channel estimation, which is studied next. The channel estimation with both noise and decision error is analyzed in this section. Both forgetting factors μ_1 and μ_2 are one in the following analysis.

A. Definitions and Assumptions

The adaptive channel estimation algorithm includes two processes. One is the signal detection process, in which the estimated data samples are obtained. The other is the channel estimation process, in which the channel is estimated using the estimated data. To analyze the impacts between the two processes, we need to define the decision error and estimation error first. The signal detection in MCM is done in the frequency domain, while the channel estimation is done in the time domain. Hence, the decision error and the estimation error are defined in both time and frequency domains.

First, define frequency-domain decision error as

$$\eta_{i,k} \triangleq \bar{X}_{i,k} - X_{i,k}, \quad i = 0, 1, \dots, m-1 \quad (20)$$

where $\eta_{i,k}$'s are independent with different i and k . The energy of $\eta_{i,k}$, $\sigma_{\eta_i}^2(k) \triangleq E[|\eta_{i,k}|^2]$, is bounded by $((2\sqrt{M_i} - 1)^2 d_i^2)/4$ for QAM constellation.

The time-domain decision error is given by

$$\begin{aligned} e_{i,k} &\triangleq \bar{x}_{i,k} - x_{i,k} \\ &= \frac{1}{\sqrt{m}} \sum_{l \in U} \eta_{l,k} e^{-j \frac{2\pi l i}{m}}, \quad i = -v, \dots, -1. \end{aligned} \quad (21)$$

Then, the estimated data vector can be written as

$$\bar{\mathbf{u}}_l(k) = \mathbf{u}_l(k) + \mathbf{e}_l(k) \quad (22)$$

where $\mathbf{e}_l(k) = [e_{-l,k} \cdots e_{-v,k} e_{-1,k-1} \cdots e_{-l,k-1}]^T$.

The time-domain estimation error $\varepsilon(k) = [\varepsilon_{0,k} \varepsilon_{1,k} \cdots \varepsilon_{v,k}]^T$ is given by

$$\varepsilon_{i,k} = \hat{h}_{i,k} - h_i, \quad i = 0, \dots, v. \quad (23)$$

The frequency-domain estimation error is defined as

$$\Delta H_{i,k} = \hat{H}_{i,k} - H_i, \quad i = 0, \dots, m-1. \quad (24)$$

The energy of the frequency-domain estimation error is

$$\sigma_{H_i}^2(k) \triangleq \mathbb{E}[\|\Delta H_{i,k}\|^2]. \quad (25)$$

In order to make the analysis problem tractable, we have to make the following three assumptions.

Assumption 1 (Independence Assumption): The cyclic prefix $x_{i,k}$, channel noise $n_{i,k}$, and the decision error $e_{i,k}$ are independent of each other for $i = -v, \dots, -1$.

Assumption 2 (Near-Stationary Assumption): As the system is near the equilibrium, we made the following assumptions about the decision error.

$$\mathbb{E}[\mathbf{e}_l(k)\mathbf{e}_l^H(k)] = \mathbf{R}_e(k) \quad (26)$$

and assume $\mathbf{R}_e(k)$ does not change much with k , so that

$$\mathbb{E}\left[\sum_{n=1}^k \sum_{l=1}^v \mathbf{e}_l(n)\mathbf{e}_l^H(n)\right] \approx kv\mathbf{R}_e(k). \quad (27)$$

Assumption 3 (Gaussian Distribution Assumption): Define $\mathbb{E}[\mathbf{u}_l(k)\mathbf{u}_l^H(k)] = \mathbf{R}$. We assume that the input data vector $\mathbf{u}_l(k)$ is generated from the $(v+1)$ -dimensional Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{R})$ independently. We also assume that the time-domain decision error $\mathbf{e}_l(k)$ is also Gaussian distributed. It is proved in [20, App. J] that

$$\mathbb{E}[\Phi^{-1}(k)] = \frac{1}{kv - v - 2} \mathbf{R}^{-1} \quad (28)$$

$$\mathbb{E}[\Phi^{-2}(k)] = \frac{1}{(kv - v - 2)(kv - v - 4)} \mathbf{R}^{-2}. \quad (29)$$

We have the following remarks about the above assumptions.

1) Ignoring the difference of decision error of different constellation points, the decision error $e_{i,k}$'s are independent of input data $x_{i,k}$'s and only depend on the noise $n_{i,k}$'s inside the data block part, i.e., $i = 0, 1, \dots, m-1$. $e_{i,k}$'s are independent of noise samples $n_{i,k}$'s in the cyclic prefix part. However, if we also use the data part to do the channel estimation, the decision error is correlated with the noise samples inside the data part, which not only makes the analysis in the following section difficult, but also contributes to the slow convergence in Fig. 3.

2) The Gaussian distribution assumption is generally not true. The correlation of the time-domain data is

$$\mathbb{E}[x_{i_1,k_1}x_{i_2,k_2}^*] = \frac{1}{m} \sum_{l=0}^{m-1} \Gamma_l \delta(k_1 - k_2) e^{j\frac{2\pi l(i_1 - i_2)}{m}}.$$

Only when the transmitted data in all the subchannels have same energy, \mathbf{R} is not dependent on l , and according to the central limit theorem, $\mathbf{u}_l(k)$ can be approximated by the Gaussian distribution. However, if we do the loading, the time-domain data becomes correlated. For example, for the channel used in Section II-C, the transmitted energy is focused in the first 100 subchannels. In this case, the correlation of the time-domain data is very large. If we use the data part to form the estimated correlation matrix $\bar{\Phi}(k)$, the matrix could be very ill-conditioned because the difference between $\bar{\mathbf{u}}_l(k)$'s for different l 's is very small. On the other hand, the data vector

formed by the cyclic prefix part contains the data samples from two consecutive blocks that are independent of each other. For this reason, the estimated correlation matrix $\bar{\Phi}(k)$ formed by the cyclic prefix part is better conditioned than the one formed by the data part. This is the other reason that contributes to the result of Fig. 3.

B. Convergence Analysis With Decision Error

Now we will try to analyze the convergence of the estimation algorithm. To do that, we need some approximations on the estimation error. Substituting (22) into (9) results in

$$\bar{\Phi}(k) = \Phi(k) + \Delta\Phi(k) \quad (30)$$

with

$$\begin{aligned} \Delta\Phi(k) &= \sum_{n=1}^k \sum_{l=1}^v \mathbf{e}_l(n)\mathbf{u}_l^H(n) + \mathbf{u}_l(n)\mathbf{e}_l^H(n) \\ &\quad + \mathbf{e}_l(n)\mathbf{e}_l^H(n). \end{aligned} \quad (31)$$

Then,

$$\begin{aligned} \bar{\mathbf{z}}(k) &= \sum_{n=1}^k \sum_{l=1}^v \mathbf{u}_l(n)\mathbf{u}_l^H(n)\mathbf{h} + \mathbf{u}_l(n)n_{-l,n}^* \\ &\quad + \mathbf{e}_l(n)\mathbf{u}_l^H(n)\mathbf{h} + \mathbf{e}_l(n)n_{-l,n}^* \\ &= \bar{\Phi}(k)\mathbf{h} + \epsilon(k) \end{aligned} \quad (32)$$

with

$$\begin{aligned} \epsilon(k) &= \sum_{n=1}^k \sum_{l=1}^v \mathbf{u}_l(n)n_{-l,n}^* + \mathbf{e}_l(n)n_{-l,n}^* \\ &\quad - \mathbf{u}_l(n)\mathbf{e}_l^H(n)\mathbf{h} - \mathbf{e}_l(n)\mathbf{e}_l^H(n)\mathbf{h}. \end{aligned} \quad (33)$$

Thus, the estimation error can be written as

$$\epsilon(k) = \hat{\mathbf{h}}(k) - \mathbf{h} = \bar{\Phi}^{-1}(k)\epsilon(k). \quad (34)$$

Because $\bar{X}_{i,k}$ still has values in the same constellation as that of $X_{i,k}$, $\bar{\Phi}(k)$ is generally full rank if both $X_{i,k}$ and $e_{i,k}$ are generated randomly. Then the above estimation error is bounded, since the decision error is always bounded. Therefore, even if the system cannot converge to the desired point, it would be bounded inside a certain region. Now we are going to show that if the decision error is small enough, the estimation error is going to converge to some steady point.

Suppose $\bar{\Phi}(k)^{-1}\Delta\Phi(k)$ has very small eigenvalues which are all much less than one, then

$$\bar{\Phi}^{-1} \approx \bar{\Phi}^{-1}. \quad (35)$$

Then, we have the following approximation for the estimation error:

$$\epsilon(k) \approx \bar{\Phi}^{-1}(k)\epsilon(k). \quad (36)$$

Then, we have the following statement about the dynamic behavior of the channel estimation.

Theorem 1: The mean value of the estimation error satisfies

$$\mathbb{E}[\epsilon(k)] = \frac{-kv}{kv - v - 2} \mathbf{R}^{-1} \mathbf{R}_e(k) \mathbf{h} \quad (37)$$

$$\lim_{k \rightarrow \infty} \mathbb{E}[\epsilon(k)] = \mathbf{R}^{-1} \mathbf{R}_e \mathbf{h}. \quad (38)$$

The mean squared channel estimation error $E[\varepsilon^H(k)\varepsilon(k)]$ converges linearly to a nonzero steady point as $k \rightarrow \infty$, i.e.,

$$\lim_{k \rightarrow \infty} E[\varepsilon^H(k)\varepsilon(k)] = \mathbf{h}^H \mathbf{R}_e \mathbf{R}^{-2} \mathbf{R}_e \mathbf{h}. \quad (39)$$

Define the estimation error correlation matrix as

$$\mathbf{K}(k) = E[\varepsilon(k)\varepsilon^H(k)]. \quad (40)$$

Then, the theorem can be proved by calculating

$$E[\varepsilon^H(k)\varepsilon(k)] = \text{tr}[\mathbf{K}(k)].$$

The detail of the proof is shown in Appendix I.

The constant in (39) is just the square of (38). The theorem states that the adaptive algorithm will converge to this stationary bias linearly as $k \rightarrow \infty$.

When the decision error is small enough to be ignored, the mean squared estimation error can be approximated as¹

$$\begin{aligned} \text{tr}[\mathbf{K}(k)] &\approx \frac{\sigma^2 + \mathbf{h}^H \mathbf{R}_e \mathbf{h}}{kv - v - 2} \text{tr}[\mathbf{R}^{-1}] \\ &= \frac{\sigma^2 + \mathbf{h}^H \mathbf{R}_e \mathbf{h}}{kv - v - 2} \sum_{i=0}^v \frac{1}{\lambda_i} \end{aligned} \quad (41)$$

where λ_i 's are the eigenvalues of \mathbf{R} .

From the literature about the convergence of the RLS algorithm [20], we know that the convergence of the estimation algorithm is determined in proportion to the inverse of the smallest eigenvalue of the data correlation matrix. (41) shows this also is the case when decision error exists. Ill-conditioned data input may lead to a slower convergence rate.

IV. PERFORMANCE ANALYSIS WITH EXISTENCE OF ESTIMATION ERROR

In this section, we analyze how the estimation error affects the residual noise at decision point which determines the SER and hence, the decision error.

We define the residual noise at the decision point as

$$\xi_{i,k} \triangleq \hat{X}_{i,k} - X_{i,k}. \quad (42)$$

From Section II-A, $\hat{X}_{i,k}$ is the output of the equalizer. In practical systems, the equalizer is obtained from the channel estimation. The equalizer now is

$$\hat{W}_{i,k} = \frac{1}{\hat{H}_{i,k}}. \quad (43)$$

Then, the residual noise is

$$\xi_{i,k} = \hat{W}_{i,k-1} Y_{i,k} - X_{i,k}. \quad (44)$$

Due to the loading algorithm, we can assume $\|H_i\|^2$'s for $i \in U$ is large enough for the following approximation:

$$\hat{W}_{i,k} \approx \frac{1}{H_i} - \frac{\Delta H_{i,k}}{H_i^2}. \quad (45)$$

¹In this case, the dominate term in (67) is $\mathbf{K}_1(k)$ and $\mathbf{K}_3(k)$.

Using the results of the previous section, $\Delta H_{i,k}$ can be viewed as the linear combination of $\mathbf{e}_l(k)$ and $n_{l,k}$. Then, we can say that $\Delta H_{i,k}$ is Gaussian distributed based on the Gaussian distribution assumption of $\mathbf{e}_l(k)$ and $n_{l,k}$.

Theorem 2: Under the independence assumption and ignoring higher order error terms such as $E[|\Delta H_{i,k}|^4]$ and $\sigma^2 E[|\Delta H_{i,k}|^2]$, the channel estimation error propagates to the decision point as an additional Gaussian noise term conditioning on the knowledge of the transmitted signal. The conditioned probability $p(\xi_{i,k}|X_i)$ of the residual noise at the i th subchannel then follows $\mathcal{N}(\theta_k(X_i), \rho_k(X_i))$, where

$$\theta_k(X_i) \approx -\frac{E[\Delta H_{i,k-1}]}{H_i} X_i \quad (46)$$

$$\rho_k(X_i) \approx \|W_i\|^2 (\sigma_{H_i}^2(k) \|X_i\|^2 + \sigma^2). \quad (47)$$

Using the distribution of the residual noise $\xi_{i,k}$ conditioned on the transmitted signal X_i , we can calculate the symbol error probability when transmitted signal is X_i . The SER then is obtained by taking expectation over the signal constellations.

V. RECURSIVE MAPPING OF SER

In this section, we are going to study the SER at the decision point before decoding. The decision error can be calculated once the SER is known. Then the decision error propagates to the channel estimation, and the estimation error affects the residual noise, which determines SER.

Suppose the SER $P_{e,i}(k)$'s are known. Assume the detection only mistakes the detected signal to the neighbor of the transmitted signal. Then, the decision error $\sigma_{\eta_i}^2(k)$ as

$$\sigma_{\eta_i}^2(k) \approx P_{e,i}(k) d_i^2. \quad (48)$$

As stated in the last section, the residual noise at the decision point is a Gaussian noise with mean of $\theta_k(X_i)$ and variance $\rho_k(X_i)$. Then the SER can be calculated as²

$$\begin{aligned} P_{e,i}(k) &= E_{X_i \in \mathcal{C}_i} [P_{e,i|X_i}(k)] \\ &= E \left[Q \left(\frac{d_i - 2|\Re\{\theta_k(X_i)\}|}{\sqrt{2}\rho_k(X_i)} \right) \right. \\ &\quad + Q \left(\frac{d_i + 2|\Re\{\theta_k(X_i)\}|}{\sqrt{2}\rho_k(X_i)} \right) \\ &\quad + Q \left(\frac{d_i - 2|\Im\{\theta_k(X_i)\}|}{\sqrt{2}\rho_k(X_i)} \right) \\ &\quad \left. + Q \left(\frac{d_i + 2|\Im\{\theta_k(X_i)\}|}{\sqrt{2}\rho_k(X_i)} \right) \right] \end{aligned} \quad (49)$$

where $P_{e,i|X_i}(k)$ is the symbol error probability of when X_i is transmitted. \mathcal{C}_i is the constellation of the i th subchannel.

A. Transient Analysis

In this section, we study the case when the adaptive algorithm just started and the estimation error is relatively large. In this case, we can see from (49) that if the bias of channel estimation is so large that $|\Re\{\theta_k(X_i)\}|$ or $|\Im\{\theta_k(X_i)\}|$ is much larger compared to d_i , the SER is

²Refer to Appendix III for detail derivation.

dominated by $Q((d_i - 2|\Re\{\theta_k(X_i)\})/(\sqrt{2}\rho_k(X_i)))$ or $Q((d_i - 2|\Im\{\theta_k(X_i)\})/(\sqrt{2}\rho_k(X_i)))$. In this case, the SER is bounded by

$$P_{e,i}(k) \leq \mathbb{E} \left[2Q \left(\frac{d_i - 2\|\theta_k(X_i)\|}{\sqrt{2}\rho_k(X_i)} \right) \right].$$

Using the results of *Theorem 1*, we have

$$\|\theta_k(X_i)\| = \frac{(k-1)v\|X_i\|}{(k-1)v-v-2} \frac{\mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i}{m} \times \sum_{l \in U} \sigma_{\eta_l}^2 \|H_l \mathbf{w}_i^H \mathbf{w}_l\| \quad (50)$$

$$\leq \frac{(k-1)v^2\|X_i\|}{(k-1)v-v-2} \frac{\mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i}{m} \times \sum_{l \in U} \sigma_{\eta_l}^2 \|H_l\| \quad (51)$$

where $\mathbf{w}_i \triangleq [1e^{-j(2\pi i)/m} \dots e^{-j(2\pi i v)/m}]^T$.

As stated in Section III-B, the estimation error is bounded. Then, we define $\rho = \max_{X_i, k} \rho_k(X_i)$. The SER bound becomes

$$P_{e,i}(k) \leq B_{e,i}(k) = \mathbb{E} \left[2Q \left(\frac{d_i - 2 \frac{(k-1)v^2\|X_i\|}{(k-1)v-v-2} \frac{\mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i}{m} \sum_{l \in U} P_{e,l}(k-1) d_l^2 \|H_l\|}{\sqrt{2}\rho} \right) \right]. \quad (52)$$

Define $r(k-1) = \sum_{l \in U} B_{e,i}(k-1) d_l^2 \|H_l\|$, then

$$r(k) = \sum_{i \in U} \mathbb{E} \left[2Q \left(\frac{d_i - 2 \frac{(k-1)v^2\|X_i\|}{(k-1)v-v-2} \frac{\mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i}{m} r(k-1)}{\sqrt{2}\rho} \right) \right] \times d_i^2 \|H_i\| = \phi(r(k-1)). \quad (53)$$

We now form a recursive mapping, $r = \phi(r)$. The condition for this recursive mapping to converge is that $|\phi'(r)| < 1$, i.e.,

$$|\phi'(r)| = \left| \sum_{i \in U} \frac{4\sqrt{2}(k-1)v^2 \mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i}{((k-1)v-v-2)m\rho} \times \mathbb{E} \left[e^{-\gamma^2(r)} \gamma(r) \|X_i\| \right] d_i \|H_i\| \right| < 1 \quad (54)$$

where $\gamma(r) = (d_i - 2(((k-1)v^2\|X_i\|)/((k-1)v-v-2))(\mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i)/(m)r) / (\sqrt{2}\rho)$.

As the adaptive algorithm begins, the offset $\|\theta_k(X_i)\|$ caused by the estimation error dominates the SER. In this case, (54) is the sufficient condition to make the algorithm converge to an equilibrium. If this condition is satisfied and the SER is bounded to a small range that $\|\theta_k(X_i)\|$ can be ignored, we can further analyze the convergence property of the system.

B. Local Convergence

In this section, we will assume that the system is near the equilibrium, and the SER is small enough to ignore the offset caused by $\theta_k(X_i)$. Then the SER can be approximated by

$$P_{e,i}(k) \approx \mathbb{E} \left[4Q \left(\frac{d_i}{\sqrt{2}\rho_k(X_i)} \right) \right]. \quad (55)$$

Since the decision error is small, approximate the estimation error as

$$\mathbf{K}(k) \approx \frac{\sigma^2 + \mathbf{h}^H \mathbf{R}_e(k) \mathbf{h}}{kv-v-2} \mathbf{R}^{-1}. \quad (56)$$

For the frequency domain estimation error, we have

$$\Delta H_{i,k} = \sum_{l=0}^v \varepsilon_{l,k} e^{-j \frac{2\pi l i}{m}}.$$

Hence, the frequency domain estimation error can be approximated as

$$\begin{aligned} \sigma_{H_i}^2(k) &\approx \frac{\sigma^2 + \mathbf{h}^H \mathbf{R}_e(k) \mathbf{h}}{kv-v-2} \mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i \\ &= \frac{\sigma^2 + \frac{1}{m} \sum_{l \in U} P_{e,l}(k) d_l^2 \|H_l\|^2}{kv-v-2} \mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i. \end{aligned} \quad (57)$$

The SER then can be derived as

$$P_{e,i}(k) = \mathbb{E} \left[4Q \left(\frac{Q^{-1} \left(\frac{P_e(0)}{4} \right) \frac{\|H_i\|}{\|H_{i,0}\|}}{\sqrt{\frac{\mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i \|X_i\|^2}{kv-v-2} \left[1 + \frac{2Q^{-2} \left(\frac{P_e(0)}{4} \right) \sum_{l \in U} \frac{\|H_l\|^2}{\|H_{l,0}\|^2} P_{e,l}(k-1) \right] + 1}} \right) \right]. \quad (58)$$

Define a weighted average SER as

$$\bar{p}(k) \triangleq \frac{1}{m} \sum_{i \in U} \frac{\|H_i\|^2}{\|H_{i,0}\|^2} P_{e,i}(k). \quad (59)$$

The iteration of $\bar{p}(k)$ is

$$\begin{aligned} \bar{p}(k) &= \frac{1}{m} \sum_{i \in U} \frac{\|H_i\|^2}{\|H_{i,0}\|^2} \\ &\times \mathbb{E} \left[4Q \left(\frac{Q^{-1} \left(\frac{P_e(0)}{4} \right) \frac{\|H_i\|}{\|H_{i,0}\|}}{\sqrt{\frac{\mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i \|X_i\|^2}{kv-v-2} \left[1 + 2Q^{-2} \left(\frac{P_e(0)}{4} \right) \bar{p}(k-1) \right] + 1}} \right) \right] \\ &= f(\bar{p}(k-1)). \end{aligned} \quad (60)$$

The condition for this iteration to converge is $|f'(\bar{p})| < 1$.

It is easy to calculate this derivative, and the condition becomes

$$\begin{aligned} |f'(\bar{p})| &= \frac{4}{(kv-v-2)m\sqrt{2\pi}} \\ &\times \sum_{i \in U} \mathbb{E} \left[\mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i \|X_i\|^2 e^{-\frac{g_i^2(\bar{p})}{2}} g_i^3(\bar{p}) \right] < 1 \end{aligned} \quad (61)$$

with

$$g_i(\bar{p}) = \frac{Q^{-1} \left(\frac{P_e(0)}{4} \right) \frac{\|H_i\|}{\|H_{i,0}\|}}{\sqrt{\frac{\mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i \|X_i\|^2}{kv-v-2} \left[1 + 2Q^{-2} \left(\frac{P_e(0)}{4} \right) \bar{p} \right] + 1}}. \quad (62)$$

Now let us consider the function $e^{-(t^2)/2t^3}$. It is easy to find out that this function has one maxima at $\sqrt{3}$. Then $|f'(\bar{p})|$ is bounded by

$$\begin{aligned} |f'(\bar{p})| &\leq \frac{36\sqrt{3}e^{\frac{3}{2}}}{(kv-v-2)m\sqrt{2\pi}} \sum_{i \in U} \mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i \mathbb{E}[\|X_i\|^2] \\ &= \frac{36\sqrt{3}e^{\frac{3}{2}}}{(kv-v-2)m\sqrt{2\pi}} \sum_{i \in U} \mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i \Gamma_i. \end{aligned} \quad (63)$$

The boundness of $|f'(\bar{p})|$ implies that

$$\lim_{k \rightarrow \infty} |f'(\bar{p})| < 1. \quad (64)$$

This means that the adaptive channel estimation algorithm will converge as the iteration goes on.

C. Discussions

In this section, we will discuss some factors that affect the convergence.

First of all, in both (54) and (61), $\mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i$ plays an important role. It is required that $\mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i$ be small in order to get a faster convergence. If \mathbf{R} is a diagonal matrix, then

$$\mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i = \sum_i \frac{1}{\lambda_i}$$

where λ_i is the eigenvalue of \mathbf{R} . This again indicates the conclusion we have stated in Section III already that the input data need to be well-conditioned to guarantee a fast convergence of the system. Because the loading algorithm used in this paper always tries to load the data compatible to the channel spectrum, the time-domain data becomes correlated, which may lead to a large condition number of the data correlation matrix, especially in the case where the whole block of the data is fed back. Such an ill-conditioned correlation matrix significantly slows down the convergence of the algorithm.

It is also noticed that there is no noise variance term in (63). However, it does not mean the noise does not affect the SER iteration. It has its impact through the loading, since the different noise level results in different constellations. Then the expectation in (63) is taken over different constellations. This may cause different results. Moreover, if we fix the transmit energy and the SER requirement, increasing the channel noise may lead some subchannels to become unused, which will significantly affect the correlation of the time-domain data, and then affect the convergence of the system. The only way to combat this phenomena is to increase the transmit power, which again becomes a tradeoff between performance and cost.

Further analysis of $|f'(\bar{p})|$ indicates that the iteration converges faster as $e^{-(g_i^2(\bar{p}))/2} g_i^3(\bar{p})$ goes smaller. Such a function may go small along two directions, either t is small or large. However, when $g_i(\bar{p}) \rightarrow 0$, it will make the SER go to one, which means the system is collapsed. Hence, $g_i(\bar{p}) \rightarrow 0$ is a trivial solution for the equilibrium. In order to make the system converge faster, we should let $g_i(\bar{p})$ be as large as possible. It is easy to see that $g_i(\bar{p})$ goes larger when \bar{p} is small. According to the definition, \bar{p} is weighted by the ratio of ideal channel response $\|H_i\|^2$ over the initial channel response $H_{i,0}$. Since

$\mathbf{Q}(\cdot)$ is a convex function, applying Jensen's inequality to (60) gives us

$$\begin{aligned} \bar{p}(k) \geq \frac{1}{m} \mathbb{E} \left[4\mathbf{Q} \left(\sum_{i \in U} \frac{\|H_i\|^2}{\|H_{i,0}\|^2} \right. \right. \\ \left. \left. \times \frac{\mathbf{Q}^{-1} \left(\frac{P_e(0)}{4} \right) \frac{\|H_i\|}{\|H_{i,0}\|}}{\sqrt{\frac{\mathbf{w}_i^H \mathbf{R}^{-1} \mathbf{w}_i \|X_i\|^2}{kv-v-2} [1+2\mathbf{Q}^{-2} \left(\frac{P_e(0)}{4} \right) \bar{p}(k-1)]^{+1}}} \right) \right]. \end{aligned} \quad (65)$$

The equality is valid if $(\|H_i\|^2)/(\|H_{i,0}\|^2)$ is constant. This means that the initial channel response, on which loading is based, should have the same shape as that of the ideal channel response in order to make the system converge faster. It is also noticed that the right-hand side of the inequality becomes smaller when $(\|H_i\|^2)/(\|H_{i,0}\|^2)$ becomes larger. This means that we can add more gap in loading to make the system converge faster.

The other factor that contributes to the nice convergence property of the proposed algorithm is that the convergence depends on the overall performance of the system. As shown in (60), the SER iteration for any individual subchannel depends on the performance of the whole system $\bar{p}(k-1)$. In contrast, the existing adaptive equalization scheme in Section II-C treats all the subchannels independently and applies the same scheme for each subchannel. Therefore, the convergence only depends on the channel variation and performance of individual subchannel. In this case, if the performance of a specific subchannel goes bad, it may never recover again. However, in our algorithm, it can be recovered if the overall system still performs well. The key point here is that the independence of subchannels is an advantage for the signal detection but a disadvantage for the channel estimation, because the channel responses of different subchannels are actually correlated.

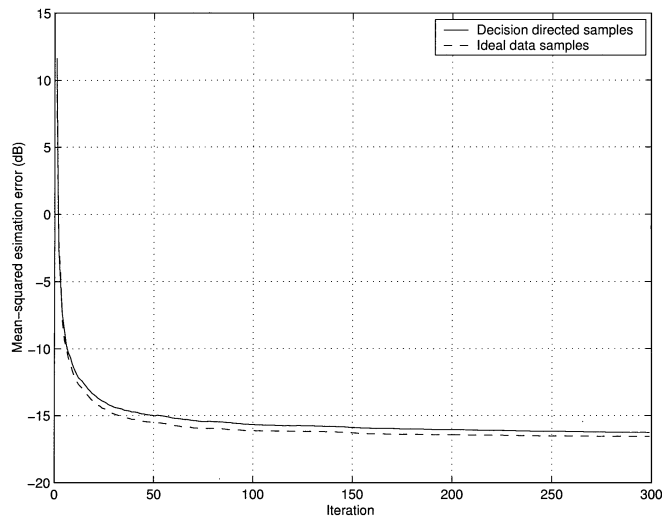
VI. COMPUTER SIMULATION

Since we use some approximations in the previous analysis, computer simulation is done to verify the analysis results. The MCM system adopted in this section has 256 complex subchannels and an average transmit energy of one in all the simulations.

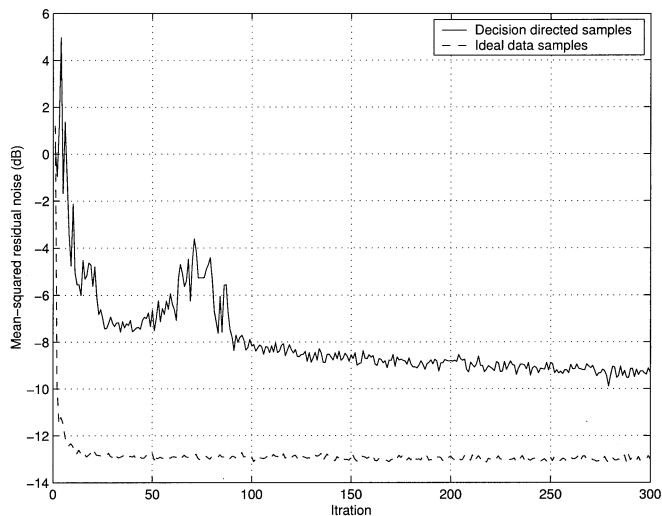
Example 1: By this example, we show how a stationary decision error affects the channel estimation. In order to do that, we use the ideal channel information in the equalizer, i.e., no estimation error propagates to decision point. The transfer function used in this example is $H(D) = (0.02 + 0.16D^2)/(1 - 1.5D + 0.54D^2)$.

Fig. 4(a) and (b) shows the mean-squared channel estimation error $1/v \mathbb{E}[\varepsilon^H(k)\varepsilon(k)]$ and mean-squared residual noise $1/m \sum_{i \in U} \sigma_{\xi_i}^2(k)$ with and without decision error, respectively. It can be seen that there is a constant difference between the two curves of estimation error which corresponds to the bias of the channel estimation caused by the decision error. This bias also propagates to the decision point which is shown in Fig. 4(b).

Example 2: The transfer function of the channel used in this example is $H(D) = 1/(1 - 0.3D)$. We show the SER iteration in (a) and the mean-squared channel estimation error in (b). Two cases of loading are simulated in Fig. 5(a). In one case, the loading is done according to the flat channel response. In the



(a)



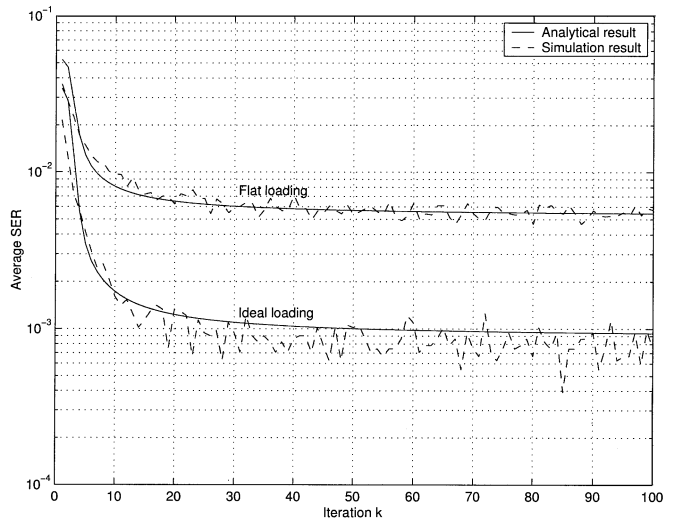
(b)

Fig. 4. (a) Mean-squared estimation error and (b) residual noise at the decision point ($v = 32$, $\mu_1 = \mu_2 = 1$, $\sigma^2 = 0.01$, and $P_e(0) = 10^{-2}$).

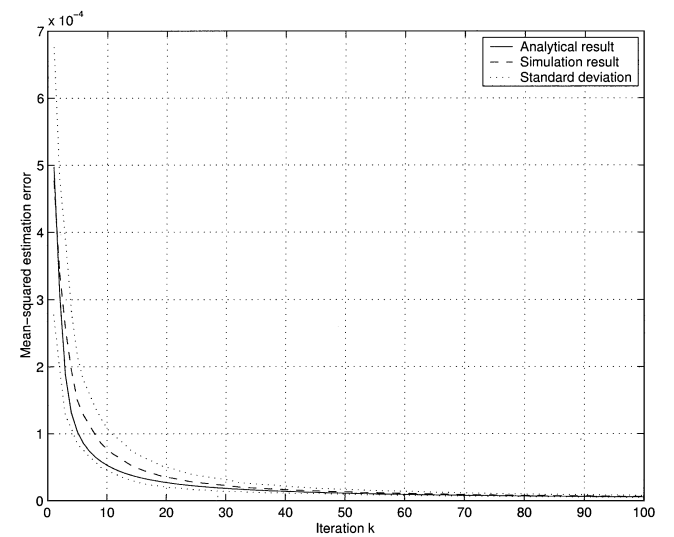
other case, the loading is done according to the ideal channel response $H(D)$. In Fig. 5(b), the case of ideal loading is shown with the standard deviation of the mean-squared error. In both (a) and (b), our analytical results are quite close to the simulations. The figure also verifies that the system converges faster when loading is done according to the ideal channel information. It converges in about 10 iterations when ideal loading is done, and in about 20 iterations with flat loading. Furthermore, the ideal loading has better overall performance than the non-ideal flat loading.

VII. CONCLUSION

In this paper, we first present an adaptive channel estimation algorithm for MCM systems using the cyclic prefix. We observed that the cyclic prefix originally used to reduce ISI is actually a source of channel information. A block RLS algorithm using decision-directed samples then is applied to exploit



(a)



(b)

Fig. 5. (a) SER and (b) mean-squared error iteration ($v = 32$, $\mu_1 = \mu_2 = 1$, $\sigma^2 = 0.01$, and $P_e(0) = 10^{-3}$).

the channel information in such a training sequence. In our simulation, the proposed algorithm performs more robustly than the existing adaptive equalization scheme [3] without sending extra training.

Then we investigate the performance analysis of the adaptive channel estimation algorithm. We first prove that the existence of decision error results in a biased channel estimation and the algorithm converges with the same rate as that without decision error. Then we analyze the effect of the channel estimation error on the system performance and prove that the channel estimation error appears at the decision point as an additional noise. Finally, we derive a recursive mapping of SER using the above conclusions. We first derive the SER bound as the channel estimation error is large. Then we consider local convergence of the recursive mapping and find that the system will be guaranteed to converge as the iteration goes on. The convergence rate is determined by the eigenvalues of the data correlation matrix, which is affected by the channel noise and loading algorithm.

In our analysis, we consider both the channel noise and the decision error in signal detection, which is different from most of the analysis of such a decision-directed algorithm in the existing literature. Thus, the analysis presented in this paper is closer to the practical environment. However, it should be noted that some assumptions are made in the analysis. Careful evaluation of the validity of those assumptions is necessary before the analysis results can be used in the MCM system design.

APPENDIX I

PROOF OF THEOREM 1

Proof: Taking expectation of (36), we have

$$\begin{aligned} \mathbb{E}[\varepsilon(k)] &\approx \mathbb{E}[\Phi^{-1}(k)\varepsilon(k)] \\ &= \mathbb{E}\left[\Phi^{-1}(k)\sum_{n=1}^k\sum_{l=1}^v\mathbf{u}_l(n)n_{-l,n}^* + \mathbf{e}_l(n)n_{-l,n}^* \right. \\ &\quad \left. - \mathbf{u}_l(n)\mathbf{e}_l^H(n)\mathbf{h} - \mathbf{e}_l(n)\mathbf{e}_l^H(n)\mathbf{h}\right]. \end{aligned}$$

According to the independence assumption, only the last term in the expectation is nonzero. Using near-stationary assumption, this term can be derived as

$$\begin{aligned} \mathbb{E}[\varepsilon(k)] &\approx \mathbb{E}\left[-\Phi^{-1}(k)\sum_{n=1}^k\sum_{l=1}^v\mathbf{e}_l(n)\mathbf{e}_l^H(n)\mathbf{h}\right] \\ &\approx \frac{-kv}{kv-v-2}\mathbf{R}^{-1}\mathbf{R}_e(k)\mathbf{h}. \end{aligned} \quad (66)$$

Then using the independence assumption, $\mathbf{K}(k)$ can be simplified as

$$\mathbf{K}(k) = \mathbf{K}_1(k) + \mathbf{K}_2(k) + \mathbf{K}_3(k) + \mathbf{K}_4(k) \quad (67)$$

where

$$\begin{aligned} \mathbf{K}_1(k) &= \mathbb{E}\left[\Phi^{-1}(k)\sum_{n_1=1}^k\sum_{l_1=1}^v\sum_{n_2=1}^k\sum_{l_2=1}^v\right. \\ &\quad \left.\times \mathbf{u}_{l_1}(n_1)n_{-l_1,n_1}^*n_{-l_2,n_2}\mathbf{u}_{l_2}^H(n_2)\Phi^{-1}(k)\right] \\ \mathbf{K}_2(k) &= \mathbb{E}\left[\Phi^{-1}(k)\sum_{n_1=1}^k\sum_{l_1=1}^v\sum_{n_2=1}^k\sum_{l_2=1}^v\right. \\ &\quad \left.\times \mathbf{e}_{l_1}(n_1)n_{-l_1,n_1}^*n_{-l_2,n_2}\mathbf{e}_{l_2}^H(n_2)\Phi^{-1}(k)\right] \\ \mathbf{K}_3(k) &= \mathbb{E}\left[\Phi^{-1}(k)\sum_{n_1=1}^k\sum_{l_1=1}^v\sum_{n_2=1}^k\sum_{l_2=1}^v\right. \\ &\quad \left.\times \mathbf{e}_{l_1}^H(n_1)\mathbf{h}\mathbf{h}^H\mathbf{e}_{l_2}(n_2)\mathbf{u}_{l_2}^H(n_2)\Phi^{-1}(k)\right] \\ \mathbf{K}_4(k) &= \mathbb{E}\left[\Phi^{-1}(k)\sum_{n_1=1}^k\sum_{l_1=1}^v\sum_{n_2=1}^k\sum_{l_2=1}^v\right. \\ &\quad \left.\times \mathbf{h}\mathbf{h}^H\mathbf{e}_{l_2}(n_2)\mathbf{e}_{l_2}^H(n_2)\Phi^{-1}(k)\right]. \end{aligned}$$

Using the near stationary and Gaussian input assumptions, it can be shown that

$$\begin{aligned} \text{tr}[\mathbf{K}_1(k)] &= \text{tr}\{\sigma^2\mathbb{E}[\Phi^{-1}(k)]\} \\ &= \frac{\sigma^2}{kv-v-2}\text{tr}[\mathbf{R}^{-1}] \end{aligned} \quad (68)$$

$$\begin{aligned} \text{tr}[\mathbf{K}_2(k)] &= \sigma^2\mathbb{E}\left[\sum_{n=1}^k\sum_{l=1}^v\mathbf{e}_l^H(n)\Phi^{-1}(k)\Phi^{-1}(k)\mathbf{e}_l(n)\right] \\ &= \frac{\sigma^2}{(kv-v-2)(kv-v-4)} \\ &\quad \times \text{tr}\left\{\mathbb{E}\left[\mathbf{R}^{-1}\sum_{n=1}^k\sum_{l=1}^v\mathbf{e}_l(n)\mathbf{e}_l^H(n)\mathbf{R}^{-1}\right]\right\} \\ &= \frac{\sigma^2kv}{(kv-v-2)(kv-v-4)} \\ &\quad \times \text{tr}[\mathbf{R}^{-1}\mathbf{R}_e(k)\mathbf{R}^{-1}] \end{aligned} \quad (69)$$

$$\begin{aligned} \text{tr}[\mathbf{K}_3(k)] &= \text{tr}\left\{\mathbb{E}\left[\Phi^{-1}(k)\sum_{n=1}^k\sum_{l=1}^v\mathbf{u}_l(n)\right.\right. \\ &\quad \left.\left.\times \mathbf{u}_l^H(n)\mathbf{h}^H\mathbf{R}_e(k)\mathbf{h}\Phi^{-1}(k)\right]\right\} \\ &= \frac{\mathbf{h}^H\mathbf{R}_e(k)\mathbf{h}}{kv-v-2}\text{tr}[\mathbf{R}^{-1}] \end{aligned} \quad (70)$$

$$\begin{aligned} \text{tr}[\mathbf{K}_4(k)] &\approx k^2v^2\text{tr}\left\{\mathbb{E}\left[\Phi^{-1}(k)\mathbf{R}_e(k)\mathbf{h}\mathbf{h}^H\right.\right. \\ &\quad \left.\left.\times \mathbf{R}_e(k)\Phi^{-1}(k)\right]\right\} \\ &= \frac{k^2v^2}{(kv-v-2)(kv-v-4)}\mathbf{h}^H\mathbf{R}_e(k) \\ &\quad \times \mathbf{R}^{-2}\mathbf{R}_e(k)\mathbf{h}. \end{aligned} \quad (71)$$

In the derivation of (71), we use the approximation that with small decision error

$$\begin{aligned} &\mathbb{E}\left[\sum_{n_1=1}^k\sum_{l_1=1}^v\sum_{n_2=1}^k\sum_{l_2=1}^v\mathbf{e}_{l_1}(n_1)\mathbf{e}_{l_1}^H(n_1)\mathbf{h}\mathbf{h}^H\mathbf{e}_{l_2}(n_2)\mathbf{e}_{l_2}^H(n_2)\right] \\ &\approx \mathbb{E}\left[\sum_{n=1}^k\sum_{l=1}^v\mathbf{e}_l(n)\mathbf{e}_l^H(n)\right]\mathbf{h}\mathbf{h}^H\mathbb{E}\left[\sum_{n=1}^k\sum_{l=1}^v\mathbf{e}_l(n)\mathbf{e}_l^H(n)\right] \\ &= k^2v^2\mathbf{R}_e(k)\mathbf{h}\mathbf{h}^H\mathbf{R}_e(k). \end{aligned}$$

From above, we can see that the behavior of the estimation error is related to the behavior of the decision error $\mathbf{R}_e(k)$. If we assume that the decision error is near stationary, then $\text{tr}[\mathbf{K}_1(k)]$, $\text{tr}[\mathbf{K}_2(k)]$ and $\text{tr}[\mathbf{K}_3(k)]$ decay linearly to zero as $k \rightarrow \infty$. As $k \rightarrow \infty$, $\text{tr}[\mathbf{K}_4(k)]$ converges a nonzero constant which is

$$\lim_{k \rightarrow \infty} \text{tr}[\mathbf{K}_4(k)] = \mathbf{h}^H\mathbf{R}_e\mathbf{R}^{-2}\mathbf{R}_e\mathbf{h}. \quad (72)$$

Hence,

$$\lim_{k \rightarrow \infty} \mathbb{E}[\varepsilon^H(k)\varepsilon(k)] = \mathbf{h}^H\mathbf{R}_e\mathbf{R}^{-2}\mathbf{R}_e\mathbf{h}. \quad (73)$$

■

APPENDIX II

PROOF OF THEOREM 2

Proof: The residual noise at the decision point is

$$\xi_{i,k} = -\frac{\Delta H_{i,k-1}}{H_i} X_{i,k} + \hat{W}_{i,k-1} N_{i,k}. \quad (74)$$

It then can be shown that the mean of the residual noise is

$$E[\xi_{i,k}|X_{i,k}] \approx -\frac{E[\Delta H_{i,k-1}]}{H_i} X_{i,k}. \quad (75)$$

The energy of the residual noise is

$$\begin{aligned} E[|\xi_{i,k}|^2|X_{i,k}] &\approx \frac{1}{|H_i|^2} (\sigma_{H_i}^2(k) \|X_{i,k}\|^2 + \sigma^2) \\ &\approx \|W_i\|^2 (\sigma_{H_i}^2(k) \|X_{i,k}\|^2 + \sigma^2). \end{aligned} \quad (76)$$

It shows that the residual noise with the channel estimation error is equivalent to the residual noise of the system with channel noise energy $\sigma_{H_i}^2(k) \|X_{i,k}\|^2 + \sigma^2$ and using perfect channel parameters in equalizers. ■

APPENDIX III

DERIVATION OF SER WITH NONZERO MEAN GAUSSIAN NOISE

Suppose the noise at the decision point is $n \sim \mathcal{N}(\theta, \rho)$. Then following the derivation of SER for QAM constellation in [22], the SER can be derived from the pulse amplitude modulation (PAM) constellation, i.e.,

$$\begin{aligned} P_r &= \frac{\sqrt{M}-1}{\sqrt{M}} \left[\frac{1}{\sqrt{\pi\rho}} \int_{d/2}^{\infty} e^{-\frac{(x-\Re\{\theta\})^2}{\rho}} dx + \frac{1}{\sqrt{\pi\rho}} \int_{-\infty}^{d/2} e^{-\frac{(x-\Re\{\theta\})^2}{\rho}} dx \right] \\ &= \frac{\sqrt{M}-1}{\sqrt{M}} \left[\frac{1}{\sqrt{2\pi}} \int_{\frac{d-2\Re\{\theta\}}{\rho}}^{\infty} e^{-\frac{x^2}{2}} dx + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{d+\Re\{\theta\}}{\rho}} e^{-\frac{x^2}{2}} dx \right] \\ &= \frac{\sqrt{M}-1}{\sqrt{M}} \left[Q\left(\frac{d-2|\Re\{\theta\}|}{\rho}\right) + Q\left(\frac{d+2|\Re\{\theta\}|}{\rho}\right) \right] \end{aligned}$$

where M is number of the constellation points and d is the distance between the adjacent constellation points.

Similarly, for the PAM constellation with noise $\Im\{n\}$, we have

$$P_i = \frac{\sqrt{M}-1}{\sqrt{M}} \left[Q\left(\frac{d-2|\Im\{\theta\}|}{\rho}\right) + Q\left(\frac{d+2|\Im\{\theta\}|}{\rho}\right) \right].$$

Then we can derive the SER for QAM using (77). As M is large and ignores the high-order term, the SER formula can be simplified as

$$P \approx Q\left(\frac{d-2|\Re\{\theta\}|}{\rho}\right) + Q\left(\frac{d+2|\Re\{\theta\}|}{\rho}\right) + Q\left(\frac{d-2|\Im\{\theta\}|}{\rho}\right) + Q\left(\frac{d+2|\Im\{\theta\}|}{\rho}\right). \quad (78)$$

REFERENCES

- [1] J. A. C. Bingham, "Multicarrier modulation for data transmission: an idea whose time has come," *IEEE Commun. Mag.*, pp. 5–14, May 1990.
- [2] J. M. Cioffi, *A Multicarrier Primer* [Online]. Available: <http://isl.stanford.edu/~cioffi/>
- [3] J. S. Chow, J. C. Tu, and J. M. Cioffi, "A discrete multitone transceiver system for HDSL application," *IEEE J. Select. Areas Commun.*, vol. 9, pp. 895–908, Aug. 1991.
- [4] L. J. Cimini, Jr., "Analysis and simulation of a digital mobile channel using orthogonal frequency-division multiplexing," *IEEE Trans. Commun.*, vol. COM-33, pp. 665–675, July 1985.
- [5] P. S. Chow, J. M. Cioffi, and J. A. C. Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Trans. Commun.*, vol. 43, pp. 773–775, Feb.-Apr. 1995.
- [6] H. Zheng and K. J. R. Liu, "Robust image and video transmission over spectrally shaped channels using multicarrier modulation," *IEEE Trans. Multimedia*, vol. 1, pp. 88–103, Mar. 1999.
- [7] R. A. Ziegler and J. M. Cioffi, "Estimation of time-varying digital radio channel," *IEEE Trans. Veh. Technol.*, vol. 41, pp. 134–151, May 1992.
- [8] P. Hoehner, S. Kaiser, and P. Robertson, "Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering," in *Proc. 1997 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Munich, Germany, Apr. 1997, pp. 1845–1848.
- [9] X. Wang and K. J. R. Liu, "Adaptive channel estimation in multicarrier modulation system using cyclic prefix," *IEEE Commun. Lett.*, vol. 3, pp. 291–293, Oct. 1999.
- [10] —, "Joint channel estimation and equalization in multicarrier modulation system using cyclic prefix," in *Proc. ICASSP'99*, vol. 5, Phoenix, AZ, 1999, pp. 2733–2736.
- [11] O. Macchi and E. Eweda, "Convergence analysis of self-adaptive equalizers," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 161–176, Mar. 1984.
- [12] Z. Ding, C. R. Johnson, Jr., and R. A. Kennedy, "On the (non)existence of undesirable equilibria of Godard blind equalizer," *IEEE Trans. Signal Processing*, vol. 40, pp. 2425–2432, Oct. 1992.
- [13] Y. Li and Z. Ding, "Convergence analysis of finite length blind adaptive equalizers," *IEEE Trans. Signal Processing*, vol. 43, pp. 2120–2129, Sept. 1995.
- [14] Y. Li and K. J. R. Liu, "Static and dynamic convergence behavior of adaptive blind equalizers," *IEEE Trans. Signal Processing*, vol. 44, pp. 2736–2744, Nov. 1996.
- [15] D. Kundur and D. Hatzinakos, "On the use of Lyapunov criteria to analyze the convergence of blind deconvolution algorithm," *IEEE Trans. Signal Processing*, vol. 46, pp. 2918–2925, Nov. 1998.
- [16] R. A. Kennedy, "Blind adaptation of decision feedback equalizers: gross convergence properties," *Int. J. Adapt. Control Signal Process.*, vol. 7, pp. 497–523, 1993.
- [17] S. A. Altekar and N. C. Beaulieu, "Upper bounds to the error probability of decision feedback equalization," *IEEE Trans. Inform. Theory*, vol. 39, pp. 145–156, Jan. 1993.
- [18] A. Klein, G. K. Kaleh, and P. W. Baier, "Zero forcing and minimum mean-square-error equalization for multiuser detection in code-division multiple-access channels," *IEEE Trans. Veh. Technol.*, vol. 45, pp. 276–287, May 1996.
- [19] J. E. Smee and N. C. Beaulieu, "Error-rate evaluation of linear equalization and decision feedback equalization with error propagation," *IEEE Trans. Commun.*, vol. 46, pp. 656–665, May 1998.

- [20] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [21] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.
- [22] J. G. Proakis, *Digital Communications*, 2nd ed. New York: McGraw-Hill, 1989.

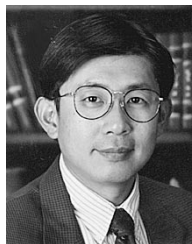


Xiaowen Wang (S'00–M'00) received the B. S. degree (ranked first in class) from Tsinghua University, Beijing, China, in 1993, and the M.S. and Ph.D. degrees from the University of Maryland, College Park, in 1999 and 2000, respectively.

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. Since 2000, she has been with the Wireless Systems Research Department, Agere Systems (formerly Bell Labs, Lu-

cent Technologies, Microelectronics). Her research interests include adaptive digital signal processing, wireless communications, and networking.

Dr. Wang was the recipient of the Graduate School Fellowship from the University of Maryland.



K. J. Ray Liu (S'86–M'86–SM'93–F'03) received the B.S. degree from the National Taiwan University, Taiwan, in 1983, and the Ph.D. degree from the University of California, Los Angeles in 1990, both in electrical engineering.

He is a Professor in the Electrical and Computer Engineering Department and the Institute for Systems Research of the University of Maryland, College Park. His research interests span broad aspects of signal processing algorithms and architectures, multimedia communications, signal

processing, wireless communications, networking, information security, and bioinformatics, in which he has published over 250 refereed papers. He is the Editor-in-Chief of *IEEE Signal Processing Magazine*, and was the Editor-in-Chief of *EURASIP Journal on Applied Signal Processing*. He has served as an Associate Editor of *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, a Guest Editor of special issues on Multimedia Signal Processing of *Proceedings of the IEEE*, a Guest Editor of a special issue on Signal Processing for Wireless Communications of the *IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS*, a Guest Editor of a special issue on Multimedia Communications over Networks of *IEEE Signal Processing Magazine*, a Guest Editor of special issue on Multimedia over IP of the *IEEE TRANSACTIONS ON MULTIMEDIA*, and an Editor of the *Journal of VLSI Signal Processing Systems*.

Dr. Liu is the recipient of numerous awards including the 1994 National Science Foundation Young Investigator Award, the IEEE Signal Processing Society's 1993 Senior Award (Best Paper Award), and the IEEE 50th Vehicular Technology Conference Best Paper Award in 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 has served as Chairman of the Multimedia Signal Processing Technical Committee of the IEEE Signal Processing Society.