A Least-Squares Approach to Blind Channel Identification

Guanghan Xu, *Member, IEEE*, Hui Liu, Lang Tong, *Member, IEEE*, and Thomas Kailath, *Fellow, IEEE*

***Abstract*—Conventional blind channel identification algorithms are based on channel outputs and knowledge of the probabilistic model of channel input. In some practical applications, however, the input statistical model may not be known, or there may not be sufficient data to obtain accurate enough estimates of certain statistics. In this paper, we consider the system input to be an unknown deterministic signal and study the problem of blind identification of multichannel FIR systems without requiring the knowledge of the input statistical model. A new blind identification algorithm based solely on the system outputs is proposed. Necessary and sufficient identifiability conditions in terms of the multichannel systems and the deterministic input signal are also presented.**

I. Introduction

SINCE Sato [20] first proposed the innovative idea of self-recovering (blind) adaptive identification, blind channel identification and equalization, as research problems, have been extensively studied by many researchers [3], [5], [8], [9], [12], [20], [22], [26]. Since it is well known that the second-order *stationary* statistics of *a scalar* system output do not contain sufficient information on a possibly nonminimum phase system, higher order statistics were adopted to identify the system [4], [5], [9], [12], [18], [22], [26]. Although these algorithms provide satisfactory estimation results in certain scenarios, they often require a relative large number of data samples, which may limit their applications in a quickly changing environment. Gardner [8] showed that the second-order statistics of *cyclostationary* signals contain phase information that can be used for nonminimum phase system identification. More recently, a second-order based approach for blind channel identification was proposed using oversampled output data to recover the lost phase information of an FIR channel [14], [24], [25], [27]. Although this algorithm requires less data samples in comparison with the high-order methods, it assumes that all the input symbols are mutually uncorrelated and of the same power.

It is important to point out that the above blind identification methods are *blind* only in the sense that the system (channel) input is not used in the identification; they do require some statistical assumptions on the input. However, the statistical model of the input may not be available, or there may not be enough data samples to find a reasonably accurate statistical estimate. For example, in a fast fading environment, the multipath channels in wireless communications vary rapidly, and we only have a few data samples corresponding to the same channel characteristics. In this kind of scenario, it is not reasonable to assume that the estimates of the signal statistics are close enough to their true (known) values. Hence, it may be easier to solve this problem by treating the input as a *deterministic* signal. This brings up a new challenge in the area of channel identification: Is it possible to identify channel(s) without even having knowledge of statistical models of the input?

In this paper, we show that, under certain conditions, it is possible to identify multichannel FIR systems (as shown in Fig. 1) based *only* on the channel outputs [15]. A novel algorithm for blind channel identification is presented. More importantly, we also present sufficient and necessary conditions under which such blind identification is possible. The basic idea behind this new approach is to exploit different instantiations of the same input signal by multiple FIR channels. Here, the concept of multichannel should not be limited to multiple physical receivers or sensors. As we will show, temporally oversampled digital communication signals can also be modeled as a multichannel system. There are two striking differences between the proposed approach and most of the existing probabilistic methods. First, under some mild conditions that we will elaborate upon later, the new approach applies to an input sequence with arbitrary statistical characteristics, such as nonstationary input or correlated input with unknown correlation functions. Second, if the channel order is known, the identification equation in the proposed approach is linear,[1] which may lead to simple implementations. Computer simulations and RF experiments show promising results in dealing with short data sequences. We also study the necessary and sufficient identifiability conditions in terms of the channel and the input signal.

During the review procedure, many interesting results on the exploitation of the output data structure for blind channel

Manuscript received December 7, 1993; revised July 17, 1995. This work was supported by the National Science Foundation under contract ECS-9310718, Southwestern Bell Technology Resources, Motorola, Inc., the University Research Institute of UT–Austin, the Joint Services Electronics Program under contract F49620-95-C-0045, the Office of Naval Research under contract N00014-95-1-0638, and the National Science Foundation under contract NCR-9321813. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon. The associate editor coordinating the review of this paper and approving it for publication was Dr. R. D. Preuss.

G. Xu and H. Liu are with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA.

L. Tong is with the Department of Electrical and Systems Engineering, University of Connecticut, Storrs, CT 06269-3157 USA.

T. Kailath is with the Information Systems Laboratory, Stanford University, Stanford, CA 94305-4055 USA.

IEEE Log Number 9415284.

[1] Other recently proposed techniques [1], [17], [21], [23] also have this nice property.

1053–587X/95$04.00

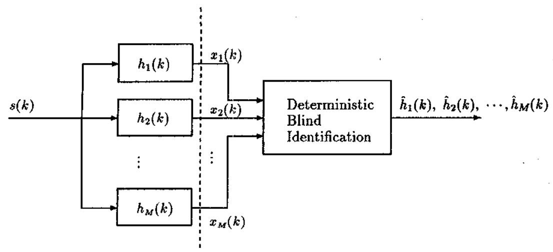


Fig. 1. Multichannel identification.

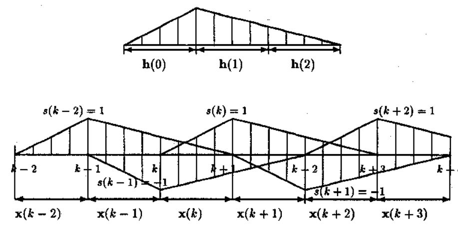


Fig. 2. Special channel with $L = 2$ and $s_k = \pm 1$.

identification emerged [1], [17], [21], [23]. Among them, an almost identical algorithm has been derived independently by Gürelli and Nikias [11]. The contribution of this paper is that we provide a more complete study of the new approach, including the necessary and sufficient identifiability conditions and relationship between the proposed approach and the method introduced in [24].

This paper is organized as follows. In Section II, we review the multichannel blind identification problem and transform a single-receiver communication channel into a multichannel system. The details of the identification derivation are given in Section III. Section IV is dedicated to derivations and proofs of the identifiability conditions. Connections with stochastic blind channel approaches are explored in Section V. Section VI gives a basic algorithm. Results of computer simulations are provided in Section VII.

Notations in this paper are fairly standard. Symbols for matrices (in capital letters) and vectors are usually in boldface. The notations $(\cdot)^*$, $(\cdot)^T$, $\odot$, and $(\cdot)^{(k)}$ stand for Hermitian, transpose, convolution, and the kth derivative operator, respectively. $\lceil x \rceil$ stands for the smallest integer that is greater than or equal to x. $a(z) = a(0) + a(1)z + \cdots + a(p)z^p$ denotes a polynomial whose coefficients are the elements of vector $\mathbf{a}$. A $\hat{\theta}$ denotes the estimate of the parameter θ. The symbol $\mathbf{I}(\mathbf{0})$ stands for the identity (zero) matrix or vector.

II. Problem Statement

As shown in Fig. 1, $x_i(\cdot)$ denotes the output from the ith channel with the FIR channel impulse response $\{h_i(\cdot)\}$, which is driven by the same input $s(\cdot)$. Clearly, for linearly modulated communication signals, the $x_i(\cdot)$, $s(\cdot)$, and $h_i(\cdot)$ are related as follows:

$$x_i(k) = \sum_{j=0}^{L} h_i(j)s(k-j), \quad i = 1 \cdots M. \tag{1}$$

where M is the number of channels, and L is the maximum order of the M channels. The blind identification problem can be stated as follows: Given the observations of channel output $\{x_i(k), i = 1, \cdots, M; k = L, \cdots, N\}$, determine the channels $\{h_i(\cdot)\}_{i=1}^{M}$ and further recover the input signals $\{s(\cdot)\}$.

Here, we will show that the data from the channel of a communication signal corresponding to a single physical receiver can be transformed into a multichannel FIR system if the sampling rate is higher than the baud rate. This is done through an example shown in Fig. 2, where the channel lasts for $L = 2$ neighboring bauds, and the oversampling rate is $M = 4$. Collecting the data samples of each baud period to form the new vectors $\mathbf{x}_k = [x(Mk), x(Mk+1), \cdots, x(Mk+M-1)]^T$, we have

$$\mathbf{x}(k) = \mathbf{h}(2)s(k-2) + \mathbf{h}(1)s(k-1) + \mathbf{h}(0)s(k). \tag{2}$$

Let us look at the ith element $x_i(k)$ of $\mathbf{x}_k$ and express it in terms of the symbols $\{s_k\}$:

$$x_i(k) = h_i(2)s(k-2) + h_i(1)s(k-1) + h_i(0)s(k) \tag{3}$$

where $h_i(k)$ is the ith element of $\mathbf{h}(k)$, $k = 0, 1, 2$. Clearly, (3) is the same multichannel expression as (1) for $L = 2$.

The above transformation of a scalar communication signal to the vector output of a multichannel system is a special case of the well-known fact that scalar cyclostationary signals can be expressed in terms of vector stationary signals [7].

III. Identification Equations

At a first glance, this problem does not seem to have a solution. Indeed, observing each individual channel, we find that it may be impossible to determine $h_i(\cdot)$ and $s(\cdot)$ uniquely without any knowledge of the channel input $s(\cdot)$. However, if we take the multichannel outputs into account, we can find that all the outputs from multiple channels are actually correlated since they are driven by the same input. In the following, it is shown that this correlation may enable us to derive a least-squares approach to blind channel identification without requiring any *a priori* knowledge of the input. For convenience of derivation, we temporarily ignore the noise. The implementation for noisy data will be discussed in Section VII.

A. Exploitation of Cross Relations Between a Channel Output Pair

From Fig. 1 and (1), for any pair of two noise-free outputs $x_i(k)$ and $x_j(k)$

$$\begin{aligned} x_i(k) &= h_i(k) \odot s(k), \\ x_j(k) &= h_j(k) \odot s(k). \end{aligned} \tag{4}$$

Then

$$\begin{aligned} h_j(k) \odot x_i(k) &= h_j(k) \odot [h_i(k) \odot s(k)] \\ &= h_i(k) \odot \underbrace{[h_j(k) \odot s(k)]}_{=x_j(k)} \\ &= h_i(k) \odot x_j(k). \end{aligned} \tag{5}$$

The above equation shows that the outputs of each channel pair are related by their channel responses. Clearly, if we have adequate data samples of the outputs, by (5), we can write out an overdetermined set of linear equations involving $h_i(\cdot)$ and $h_j(\cdot)$. Under certain conditions upon which we will elaborate later, $h_i(\cdot)$ and $h_j(\cdot)$ can be determined uniquely up to a scalar multiple. The use of such a cross relation between each output pair is the basic idea behind the new blind channel identification. Note that this structure is not available in a single channel system.

B. Identification Using an Output Pair

More specifically, for $k = L, \cdots, N$, where N is the last symbol index in the received data $x_i(k)$ and $x_j(k)$, (5) becomes $N-L+1$ linear equations involving $h_j(\cdot)$ and $h_i(\cdot)$:

$$[\mathbf{X}_i(L) \vdots -\mathbf{X}_j(L)] \begin{bmatrix} \mathbf{h}_j \\ \mathbf{h}_i \end{bmatrix} = 0 \tag{6}$$

where $\mathbf{h}_m \triangleq [h_m(L), \cdots, h_m(0)]^T$ and

$$\mathbf{X}_m(L) = \begin{bmatrix} x_m(L) & x_m(L+1) & \cdots & x_m(2L) \\ x_m(L+1) & x_m(L+2) & \cdots & x_m(2L+1) \\ \vdots & \vdots & \ddots & \vdots \\ x_m(N-L) & x_m(N-L+1) & \cdots & x_m(N) \end{bmatrix}. \tag{7}$$

For each pair of (i, j), we can always write out a set of linear equations.[2] Here, we propose to combine all of them and write a larger set of linear equations in terms of $\mathbf{h}_1, \cdots, \mathbf{h}_L$ or simply $\mathbf{h} \triangleq [\mathbf{h}_1^T, \cdots, \mathbf{h}_L^T]^T$ and solve all the channel responses *simultaneously*. Denote

$$\mathbf{X}^i(L) = \left.\begin{bmatrix} \underbrace{\begin{matrix} \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & & \\ \mathbf{0} & \cdots & \mathbf{0} \end{matrix}}_{i-1 \text{ blocks}} & \underbrace{\begin{matrix} \mathbf{X}_{i+1}(L) & -\mathbf{X}_i(L) & \mathbf{0} & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{X}_M(L) & \mathbf{0} & \cdots & -\mathbf{X}_i(L) \end{matrix}}_{M-i+1 \text{ blocks}} \end{bmatrix}\right\} M-i \text{ blocks}$$

where each block, e.g., $\mathbf{0}$ or $\{\mathbf{X}_m(L)\}$, has the size $(N-L+1) \times (L+1)$. In the noise-free case, $\mathbf{h}$ is in the null space of the following large matrix:

$$\mathbf{X}(L) = \left.\underbrace{\begin{bmatrix} \mathbf{X}^1(L) \\ \vdots \\ \mathbf{X}^{M-1}(L) \end{bmatrix}}_{M \text{ blocks}}\right\} \frac{M(M-1)}{2} \text{ blocks} \tag{8}$$

where the actual size of $\mathbf{X}(L)$ is of $(N-L+1)[M(M-1)/2] \times (L+1)M$, or equivalently

$$\mathbf{X}(L)\mathbf{h} = \mathbf{0}. \tag{9}$$

[2] As mentioned earlier, the equation is similar to the algorithm independently developed by Gürelli and Nikias [11]. It is also related to the approach [19] relying on higher order statistics.

In contrast to the statistical approaches to blind channel identification that often involve solving bilinear or even more complicated nonlinear equations, the new proposed approach is based on solving a linear equation (9), which may lead to simpler implementations. In the following, we shall consider under what conditions (9) has a *unique* solution or what types of channels and input signals are identifiable.

IV. Identifiability Conditions

The topic of identifiability conditions is often regarded as difficult, and there is extensive literature in this field (see [2], [16], and the references thereto). As we know, the input signal used can be critical to the identification results. Research has been done on the design of informative input for the best identification. However, the situation is different in the blind identification problem, where the input signal is unknown and to be determined. Our objective here is to find out the conditions when the channel outputs contain information rich enough for blind channel identification. Two issues need to be considered in this regard: 1) the condition of the identifiable channels and 2) the condition of the informative input. These two conditions may not be decoupled for the proposed blind channel identification approach. This will become clear as we proceed.

A. Some Results on Finite Hankel Matrices

Although the proposed algorithm does not require any knowledge of the input, the input characteristics are certainly not negligible. An obvious degenerate case is that the input data are all zeros. Then, no channel information is contained in the outputs, and the channels cannot be identified. Before we study the identifiability conditions, we give three lemmas on the finite Hankel matrix, which characterizes the input. Given finite input data $\{s(n), n = 0, \cdots, N\}$, denote $\mathbf{S}(r)$ the rth-order Hankel matrix of $\{s(\cdot)\}$ as follows:

$$\mathbf{S}(r) = \begin{bmatrix} s(0) & s(1) & \cdots & s(r-1) \\ s(1) & s(2) & \cdots & s(r) \\ \vdots & \vdots & \cdots & \vdots \\ s(N-r-1) & s(N-r) & \cdots & s(N) \end{bmatrix}. \tag{10}$$

Lemma 1: Given a fixed number (N) of data, let p be the highest rank of the Hankel matrices $\mathbf{S}(r)$, $r = 1 \cdots N+1$. The rank of $\mathbf{S}(r)$ is given by

$$\text{Rank}\{\mathbf{S}(r)\} = \begin{cases} r & r \le p \\ p & p < r \le N-p+2 \\ N+2-r & N-p+2 < r \le N+1 \end{cases} \tag{11}$$

Proof: See Appendix A.

Lemma 2: The null space of a rank deficient Hankel matrix $\mathbf{S}(r)$ $(r > p)$ can be uniquely characterized by a pth-order polynomial $a(z) = a(0) + a(1)z + \cdots + a(p)z^p$. The standard form of the null space can be expressed as

$$\underbrace{\begin{bmatrix} a(0) & & & \\ a(1) & a(0) & & \\ \vdots & a(1) & \ddots & \\ a(p) & \vdots & \ddots & a(0) \\ & a(p) & \vdots & a(1) \\ & & \ddots & \vdots \\ & & & a(p) \end{bmatrix}}_{r-p}. \tag{12}$$

Proof: See Appendix B.

We shall define $a(z) = a(0) + a(1)z + \cdots + a(p)z^p$ as a *characteristic polynomial* of the deterministic input $s(\cdot)$. Its roots $\{z_1, \cdots, z_p\}$ are henceforth referred to as the *modes* of $s(\cdot)$. The number of *modes* p, which is often referred to as the *linear complexity*, is a measurement of diversity of a *finite* sequence. It can be analogous to the number of frequency components in an infinite data sequence. In the remainder of this paper, we will use $\mathcal{C}\{s(n)\}$ to denote the linear complexity of $\{s(n)\}$.

$\mathbf{S}(r)$ will be of full rank if $\mathcal{C}\{s(n)\} \geq r$.

Remark: When $a(p)$ equals zero, we modify the characteristic polynomial as $z^{-p}a(z)$ such that $z = \infty$ is also the mode of the input.

Lemma 3: A vector is in the null space of $\mathbf{S}(r)$ iff its corresponding polynomial has roots equal to the modes of $s(\cdot)$: $z_1, \cdots, z_p$.

The proof is trivial and will be left to the reader.

B. Necessary and Sufficient Conditions

We now consider the identifiability conditions of multichannel blind identification associated with the proposed algorithm that identifies the channels $\{\mathbf{h}_i\}$ from the outputs $\{x_i(k), i = 1, \cdots, M, k = L, \cdots, N\}$. The input that affects these outputs is $s(k)$, $k = 0, \cdots, N$. Therefore, all the conditions shall be expressed in terms of these parameters and the deterministic signal. We have the following results.

Observation: The multichannel system can be identified uniquely by solving linear equations $\mathbf{X}(L)\hat{\mathbf{h}} = \mathbf{0}$ iff the data matrix $\mathbf{X}(L)$ is of rank $M(L+1) - 1$.

$\mathbf{X}(L)$ is a matrix of size $M(M-1)(N-L+1)/2 \times M(L+1)$. The above theorem gives the general necessary and sufficient conditions for channels to be uniquely identifiable. The next three theorems give more explicit expressions and provide more insights into the characteristics of the channels and input signal.

Theorem 1 (Sufficient Condition): The blind identification problem has a unique solution if

1) the polynomials $\{h_i(z)\}_{i=1}^M$ are *coprime* or they do *not* share any common roots
2) $\mathcal{C}\{s(n)\} \geq 2L+1$

where L is the maximum order among $\{h_i(z)\}_{i=1}^M$.

Proof: Let us rewrite $\mathbf{X}^i(L)$ as

$$\mathbf{X}^i(L) = \underbrace{\begin{bmatrix} \mathbf{S}(2L+1) & & \\ & \ddots & \\ & & \mathbf{S}(2L+1) \end{bmatrix}}_{\tilde{\mathbf{S}}} \cdot \underbrace{\begin{bmatrix} \overbrace{\mathbf{0} \cdots \mathbf{0}}^{i-1 \text{ blocks}} & \mathbf{H}_{i+1} & -\mathbf{H}_i & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} \cdots \mathbf{0} & \mathbf{H}_M & \mathbf{0} & \cdots & -\mathbf{H}_i \end{bmatrix}}_{\mathbf{H}^i} = \tilde{\mathbf{S}}\mathbf{H}^i, \tag{13}$$

where

$$\mathbf{S}(2L+1) = \begin{bmatrix} s(0) & \cdots & s(2L) \\ s(1) & \cdots & s(2L+1) \\ \vdots & \vdots & \vdots \\ s(N-2L) & \cdots & s(N) \end{bmatrix}$$

$$\mathbf{H}_i = \underbrace{\begin{bmatrix} h_i(L) & 0 & \cdots & 0 \\ h_i(L-1) & h_i(L) & \cdots & \vdots \\ \vdots & h_i(L-1) & \ddots & 0 \\ h_i(0) & \vdots & \ddots & h_i(L) \\ 0 & h_i(0) & \ddots & h_i(L-1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_i(0) \end{bmatrix}}_{L+1}.$$

If condition 2 is satisfied, then $\mathbf{S}(2L+1)$, and $\tilde{\mathbf{S}}$ will be of full column rank (Lemma 2). If $\hat{\mathbf{h}}$ is the solution of $\mathbf{X}(L)\hat{\mathbf{h}} = \mathbf{0}$, it follows that

$$\begin{bmatrix} \mathbf{H}^1 \\ \vdots \\ \mathbf{H}^M \end{bmatrix} \hat{\mathbf{h}} = \mathbf{H}\hat{\mathbf{h}} = \mathbf{0}. \tag{14}$$

By (14), it is not difficult to see

$$h_i(n) \odot \hat{h}_j(n) = h_j(n) \odot \hat{h}_i(n),$$

or

$$h_i(z)\hat{h}_j(z) = h_j(z)\hat{h}_i(z), \quad i, j = 1, \cdots, M. \tag{15}$$

By condition 1, $\{h_i(z)\}_{i=1}^M$ do not share any common roots. Let us assume that $h_k(z)$ has the maximum order L. Then, by (15), we have $h_k(z)\hat{h}_j(z) = h_j(z)\hat{h}_k(z)$ for all j. Clearly

$$\mathcal{Z}\{h_k(z)\} \in \bigcap_j \left(\mathcal{Z}\{h_j(z)\} \bigcup \mathcal{Z}\{\hat{h}_k(z)\}\right) = \mathcal{Z}\{\hat{h}_k(z)\} \bigcup \left(\bigcap_j \mathcal{Z}\{h_j(z)\}\right) \tag{16}$$

where $\mathcal{Z}[h(z)]$ denotes the roots of $h(z)$. Since $\{h_j(z)\}$ have to share some common roots, $\bigcap_j \mathcal{Z}\{h_j(z)\} = 0$. By (16), $\mathcal{Z}\{h_k(z)\} \in \mathcal{Z}\{\hat{h}_k(z)\}$. Since the maximum order of $\hat{h}_k(z)$ is L, $\mathcal{Z}[h_k(z)] = \mathcal{Z}[\hat{h}_k(z)]$, and therefore, $\hat{h}_k(z) = g_k h_k(z)$, where g_k is any nonzero constant. By (15), letting $i = k$, we can easily show that $\hat{h}_j(z) = g_k h_j(z)$, which means that $\hat{\mathbf{h}}(z) = g_k \mathbf{h}(z)$, i.e., the solution is unique up to a scalar multiple. □

Theorem 1 gives the sufficient conditions for blind identification of a multichannel system. Digital communication signals, e.g., white sequences, are generically rich in modes and therefore the sufficient condition can be satisfied by increasing the length of the sample sequences. However, for extreme short sequence, degeneration can easily occur.

It is worth noting that these conditions are the same as those for identifying a rational function with the denominator and numerator both of order L [13], [16]. It is true that our system is only of order L and can be sufficiently identified with known input $s(\cdot)$, which has $L+1$ modes (sometimes referred to as persistent excitation of order $L+1$ [2]). For blind identification, the input is *unknown*, and more than $L+1$ modes are required on the input signal. In fact, conditions 1 and 2 in Theorem 1 are the decoupled conditions regarding system observability and *informative* inputs.

We have just given an upper bound of the number of input modes. Now, we want to study its lower bound. The following theorem gives a necessary condition on blind channel identification.

Theorem 2 (Necessary Conditions): The channels cannot be uniquely identified if 1) there is a common zero shared by all channels, or 2) $\mathcal{C}\{s(n)\} < L+1$.

Proof: First, if $\{h_i(z)\}$ share a common root z_0, then $h_i(z) = h_i'(z)(z - z_0)$. z_0 becomes an unidentifiable parameter of the channels. To see this, let $\hat{h}_i(z) = h_i'(z)(z - z_0')$, and any choice of z_0' can satisfy $\mathbf{H}\hat{\mathbf{h}} = \mathbf{0}$ (14). Clearly, the same $\hat{\mathbf{h}}$ also satisfies $\mathbf{X}(L)\hat{\mathbf{h}} = \mathbf{0}$. Therefore, one necessary condition for the channels to be *uniquely* identifiable is that all the channels share no common zeros.

If $\mathcal{C}\{s(n)\} < L+1$, we will first show that the channels $\{h_i(\cdot)\}$ are not uniquely identifiable even if $s(\cdot)$ is *known*. Since blind channel identification does not have any knowledge of $s(\cdot)$, the necessary condition of standard system identification should also be that of blind channel identification. With the known $s(\cdot)$, the identification of $\hat{h}_i(\cdot)$ or $\hat{\mathbf{h}}_i$ can be found by solving

$$\mathbf{S}(L)\hat{\mathbf{h}}_i = \mathbf{x}_i \tag{17}$$

where $\mathbf{x}_i = [x_i(L), \cdots, x_i(N)]^T$. Since $\mathcal{C}\{s(n)\} < L+1$, by Lemma 1, $\mathbf{S}(L)$ is of rank L, or it is rank deficient. Clearly, (17) does not have a unique solution. Therefore, it requires $L+1$ or more modes to achieve blind channel identification.□

By Theorems 1 and 2, if the coprime condition is satisfied, the channel is identifiable if $\mathcal{C}\{s(n)\} \geq 2L+1$, and it is not if $\mathcal{C}\{s(n)\} < L+1$. What happens when $L+1 \leq \mathcal{C}\{s(n)\} \leq 2L+1$? The following two theorems will address this issue and give necessary and sufficient identifiability conditions. Unlike those given in Theorems 1 and 2, the conditions on input modes and channels cannot be decoupled.

Theorem 3: The blind equalization problem has a unique solution iff

1) $p = \mathcal{C}\{s(n)\} \geq L + 1 + \lceil L/M - 1 \rceil$.[3]
2) The matrix

$$\begin{bmatrix} \mathbf{h}(z_1) & & 0 & \mathbf{Z}(z_1) \\ & \ddots & & \vdots \\ 0 & & \mathbf{h}(z_p) & \mathbf{Z}(z_p) \end{bmatrix} \tag{18}$$

has a 1-D null space, where

$$\mathbf{h}(z_k) = \begin{bmatrix} h_1(z_k) \\ \vdots \\ h_M(z_k) \end{bmatrix}$$

$$\mathbf{Z}(z_k) = \underbrace{\begin{bmatrix} \mathbf{z}(z_k) & & 0 \\ & \ddots & \\ 0 & & \mathbf{z}(z_k) \end{bmatrix}}_{M \text{ blocks}}$$

$$\mathbf{z}(z_k) = [1 \quad z \quad \cdots \quad z^L]_{z=z_k}.$$

Proof: See Appendix C.

The above theorem gives the sufficient and necessary conditions for the blind identification of a multichannel FIR system in terms of the input and channels. As shown in Appendix C

$$[\underbrace{1 \ 1 \ \cdots \ 1}_{p} \quad -\mathbf{h}^T]^T$$

is a basis vector of the null space of the matrix in (18). An interpretation of condition 2 is that the input modes cannot coincide with those values causing the ambiguity of the cross relations among the channel responses $\mathbf{h}$. A more insightful alternative expression is given next.

Theorem 4: The sufficient and necessary conditions for blind multichannel identification can also be stated as follows:

1) $p = \mathcal{C}\{s(n)\} \geq L + 1 + \lceil L/M - 1 \rceil$.
2) There is *no* polynomial $g(z)$ of order between 1 and $p-1$ such that $\{h_i(z)g(z)\}$ can all be reduced to polynomials with an order $\leq L$ by other polynomials containing $a(z)$ as a common factor, i.e.

$$\underbrace{\hat{h}_i(z)}_{L} = \underbrace{h_i(z)}_{L}\underbrace{g(z)}_{<p} + \underbrace{a(z)}_{p}\underbrace{f_i(z)}_{<L}, \quad i = 1, \cdots, M \tag{19}$$

where $a(z)$ is the characteristic polynomial of the input, and $\{f_i(z)\}_{i=1}^M$ can be any polynomial.

Proof: See Appendix D.

If there exists such a $g(z)$ with an order between 1 and $p-1$, which satisfies (19), $\{\hat{\mathbf{h}}_i\}$ become the ambiguous channel responses that are not identifiable based on their outputs. Equation (19) actually combines the identifiability conditions on the input (fully characterized by $a(z)$) and channels $\{\mathbf{h}_i\}$ into one formula. The coprime condition is implied in this formula. To see this, assume that $\{\mathbf{h}_i\}$ are not coprime, i.e., $h_i(z) = h_i'(z)(z - z_0)$. We can always carry out long division of $a(z)/(z - z_0)$ until the step before last; then, we have $a(z) = -(z - z_0)b(z) + (z - z_0')$, where $b(z)$ is of order $p-1$.

[3] Here, we assume that all the modes are distinct and finite. Repeated and infinity modes are treated in Remarks 1 and 2 after the proof of this theorem.

TABLE I
SIMULATION PARAMETERS

	# of Symbols	# of Receivers	Oversamp. Rate	SNR (dB)	Channel Length	# of Trials
Simu. 1 2	$50 \sim 300$	2	2	$10 \sim 40$	$4T$	500
Simu. 3	50	2	2	$10 \sim 60$	$6T$	500

Let us define $g(z) \triangleq -b(z)$ and $f_i(z) \triangleq h_i'(z)$. Then

$$\begin{aligned}\hat{h}_i(z) &= h_i(z)g(z) + a(z)f_i(z)\\ &= h_i(z)g(z) - (z - z_0)g(z)h_i'(z) + h_i'(z)(z - z_0')\\ &= h_i'(z)(z - z_0')\end{aligned}$$

which implies that (19) holds.

Thus far, we have not been able to obtain a satisfactory interpretation of (19). We can only say that if the input signal and channels have such a relation as in (19), the channels are not identifiable. As for condition 1, an important observation can be made regarding the number of channels M. First, if $M = 1$, then the bound goes to infinity, and the channels are not identifiable, which makes sense. As M increases, it requires fewer modes to identify the channels. When M is reasonably large, the number of modes required on the input approximates $L+1$, which is the minimum number of modes required in standard system identification. In other words, in cases in which there is a sufficient number of channels, blind identification (with *unknown* $s(\cdot)$) and conventional system identification (with *known* $s(\cdot)$) are not so different.

V. CONNECTIONS WITH ORIGINAL APPROACHES

In this section, we explore the connections among this, the proposed approach, and the earlier approaches based on second-order statistics [14], [24], [25], [27]. The original methods assume that $\{s(n)\}$ is a white process, i.e., they are mutually uncorrelated and they have the same variance. In addition, we often use a large number (N) of data samples so that we can obtain accurate enough statistics. Under these assumptions, the sufficient and necessary identifiability condition [25], [27] of the earlier stochastic methods is that all the channels are coprime. This can also be seen from Theorem 1. As we pointed out earlier, a long enough data sequence of white inputs is generically rich in modes $(1/N - L + 1)\mathbf{S}^*(2L+1)\mathbf{S}(2L+1) \approx E\{\mathbf{s}^*(n)\mathbf{s}(n)\} = \sigma_n^2\mathbf{I}$, where $\mathbf{s}(n) \triangleq [s(n)s(n+1) \cdots s(n+2L)]$. Since the rank of $\mathbf{S}(2L+1)$ is the same as that of $\mathbf{S}^*(2L+1)\mathbf{S}(2L+1)$, $\mathbf{S}(2L+1)$ is of full rank. Under this condition, following the proof of Theorem 1, we can see that the iff identifiability condition is indeed the coprimeness among $\{h_i(z)\}$.

VI. THE PROPOSED ALGORITHM

The proposed algorithm is based on (9) with practical considerations of noise at the receivers and the unknown channel order L. When the channels are corrupted by noise, we can estimate $\mathbf{h}$ by solving the following least squares problem:

$$\min_{\hat{\mathbf{h}}} ||\mathbf{X}(L)\hat{\mathbf{h}}||^2 \tag{20}$$

where $\hat{\mathbf{h}}$ is subject to certain nontrivial constraints, e.g., $||\hat{\mathbf{h}}|| = 1$ or $\mathbf{c}^*\hat{\mathbf{h}} = 1$ for a constant vector $\mathbf{c}$. Although the treatment of the noise is in (20) may not be statistically optimal, it is perhaps a natural and simple way of formulating this problem. Further investigations may lead to an optimal approach to solving this problem.

Numerical algorithms such as singular value decomposition (SVD) [10] or the more computationally efficient fast subspace decomposition (FSD) [30] can be used to solve this problem. Note that the blind identification problem formulated above is linear, whereas most statistical approaches are nonlinear. Although the size of the matrix $\mathbf{X}(L)$ may be large for large M and L, the computational cost may be reduced by exploiting the sparsity and the block Hankel structure of this matrix. For more details about fast computation, refer to [29] and [31].

It is worth pointing out that the proposed algorithm, along with many other *parametric* estimation algorithms, is generally sensitive to the data model selection. This makes the detection of the channel order L a crucial issue. In this paper, we assume that L is known *a priori*. Interested readers are referred to [6] and [28] for objective criteria such as MDL and AIC.

A. Basic Algorithmic Procedure

1) Overestimate the order as L_e and form $\mathbf{X}(L_e)$ as in (8).
2) Perform an SVD or FSD on $\mathbf{X}(L_e)$ to estimate how many smaller singular values are the *noise* singular values and detect the maximum channel order L.
3) Use the estimated L to form $\mathbf{X}(L)$ as in (8) and find $\hat{\mathbf{h}}$ subject to some nontrivial constraint to minimize $||\mathbf{X}(L)\hat{\mathbf{h}}||$.
4) From $\hat{\mathbf{h}}$, we can have the channel responses for all the channels, i.e., $\hat{\mathbf{h}}_1, \cdots, \hat{\mathbf{h}}_M$.

VII. NUMERICAL RESULTS

A. Computer Simulations

Computer simulations were conducted to evaluate the performance of the proposed algorithm in comparison with that of the existing stochastic algorithm [25]. In all the simulations, two antennas were used, and the received data were sampled at twice the symbol rate. The input signal type is QPSK. For simplicity of comparison, we assumed that the channel order L is known, and the basic algorithm in Section VI-A was used to identify the channels. The key parameters are also summarized in Table I.

We construct the channels to simulate a wireless environment with a long delay multipath. The first set of channels considers a two-ray multipath model with delay at 0 and 1.1 baud periods. The second set is a three-ray multipath channels with delays at 0, 0.5, and 3 baud periods. The channel response values are given in Tables II and III.

The zero distribution of these two sets of channels is plotted in Fig. 3(a) and (b) with different symbols representing zeros of different channels. Although some zeros are clustered together, the channels are clearly identifiable. Root-mean-

TABLE II
CHANNEL RESPONSES #1

i	$h_i(0)$	$h_i(1)$	$h_i(2)$	$h_i(3)$
1	0	−1.280 − 0.301i	1.617 + 2.385i	0.178 + 0.263i
2	−1.023 − 0.501i	0.106 + 1.164i	1.477 + 1.850i	−0.482 − 0.523i
3	0	−0.282 + 0.562i	0.371 − 1.001i	0.041 − 0.110i
4	−0.227 + 0.487i	0.031 − 0.211i	0.336 − 0.866i	−0.110 + 0.271i

TABLE III
CHANNEL RESPONSES #2

i	$h_i(0)$	$h_i(1)$	$h_i(2)$	$h_i(3)$	$h_i(4)$	$h_i(5)$
1	0.0222 − 0.0031i	0.5236 − 1.9483i	−0.0683 + 0.0095i	0.0222 − 0.0031i	−0.0812 − 0.0977i	0.0085 − 0.0012i
2	−0.1065 + 0.0651i	−0.9114 − 0.9867i	0.3268 − 0.1998i	−0.1065 + 0.0651i	0.1887 − 0.1856i	−0.0406 + 0.0249i
3	0.3757 − 1.2429i	0.2682 − 1.2279i	−0.1083 + 0.4256i	0.0267 − 0.2953i	−0.0902 + 0.0914i	0.0472 − 0.0887i
4	−0.7860 − 0.4996i	−0.2713 − 0.8143i	0.2297 + 0.1934i	−0.0658 − 0.1874i	0.1788 − 0.0320i	−0.0955 − 0.0133i

square-error (RMSE) is employed as a performance measure of the channel estimates

$$\text{RMSE} = \frac{1}{||\mathbf{h}||}\sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}||\hat{\mathbf{h}}_{(i)} - \mathbf{h}||^2} \tag{21}$$

where N_t is the number of Monte Carlo trials (500 in our cases), and $\hat{\mathbf{h}}_{(i)}$ is the estimate of the channels from the ith trial.

In the first simulation study, we fixed the SNR to 20 dB and varied the number of symbols from 50–300. Fig. 4 shows the RMSE of the channel estimates from both the original (with $L = 4$, $m = 16$, and $d = 7$ as the operating parameters) and proposed methods. From this figure, we can see that the new method always performs better than the original method, especially when the number of symbols is small. The main reason is that the original method, which is ultimately based on the particular structure of a pencil of two exact (i.e., ensemble averaged) autocovariance matrixes but not checked by their finite-sample estimate (even in absence of noise), is not a data-efficient algorithm. On the other hand, the new method exploits the data structure of the system output and is thus not as sensitive to the number of symbols.

In the second simulation study, we fixed the number of symbols to be 50 and varied the SNR from 10–40 dB. Fig. 5 gives the RMSE's of the channel estimates from these two methods. Despite the fact that the new method always surpasses its counterpart, it is interesting to note that the performance curve of the original method flattens out after the SNR reaches 20 dB, whereas that of the proposed method still declines. This is due to the same finite data effect that cannot be cured by increasing SNR. However, since the additive noise is the only cause of the estimation error in the new proposed method, an increase of SNR certainly leads to its performance improvement.

We repeat the above simulation using the second set of channels, with $L = 6$, $m = 16$, and $d = 9$ as the operating parameters for the original approach. It is seen from Fig. 6 that both curves shift to the right yet follow the pattern similar to that of the second simulation. This is intuitively understandable since for the same amount of data samples, longer channels have more parameters, which could lead to degradation of the performance.

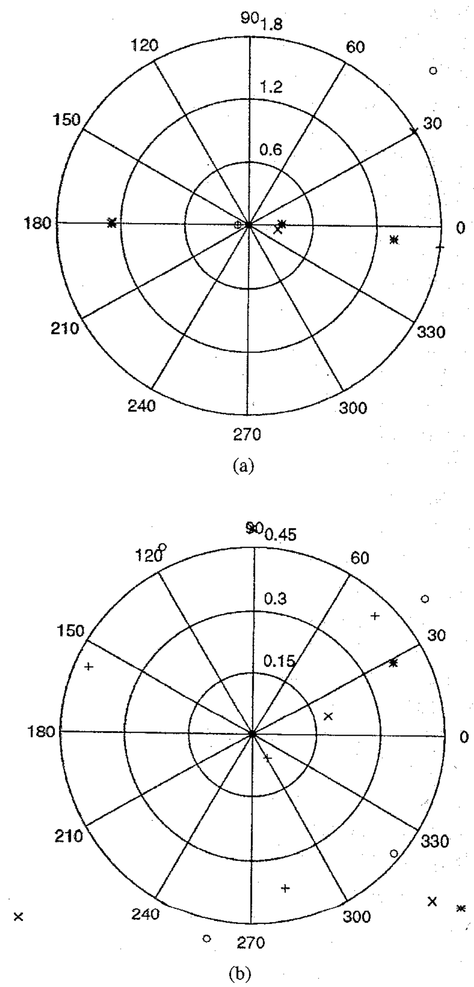


Fig. 3. Zero distribution of the channels: (a) First channel set; (b) second channel set.

B. RF Field Experiments

Finally, we show some implementation results on real data collected from our recent RF experiments. We used an eight-element uniform linear array to *spatially* oversample RF signals at about 900 MHz. The message signal is QPSK with a raised-cosine pulse. The baud rate is 50 KHz.

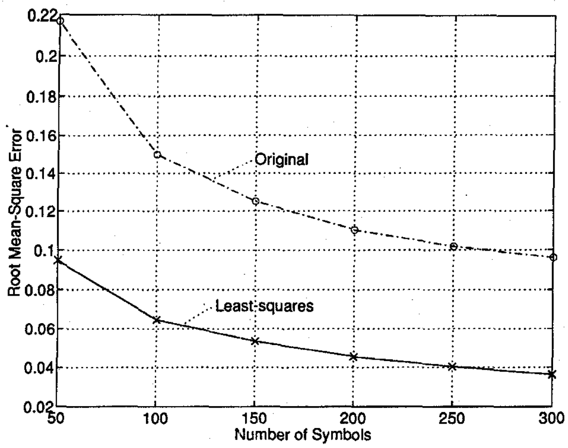


Fig. 4. Comparison of the least-squares and original methods for N = 50–300.

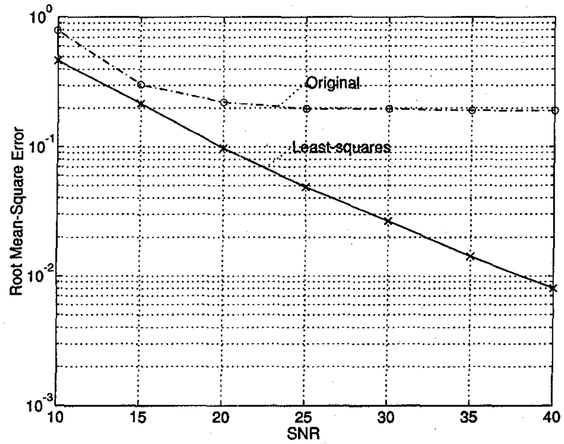


Fig. 5. Comparison of the least-squares and original methods for SNR = 10–40 dB.

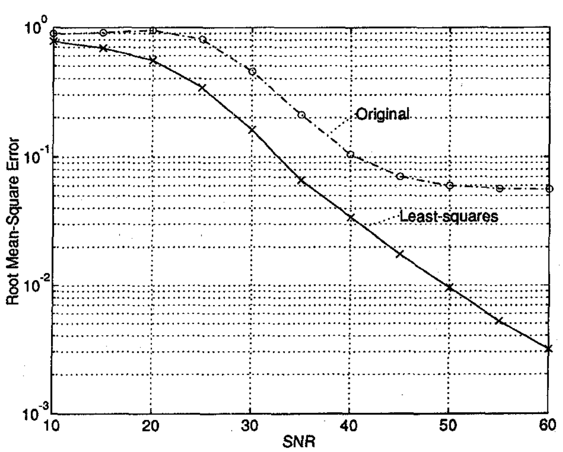


Fig. 6. Comparison of the least-squares and original methods for SNR = 10–60 dB.

To create a reliable long-delay fading environment, we artificially created some long-delay multipaths by transmitting the delayed version of the same signal from different transmitters. The phase pattern of the receiver outputs is plotted in Fig. 7(a).

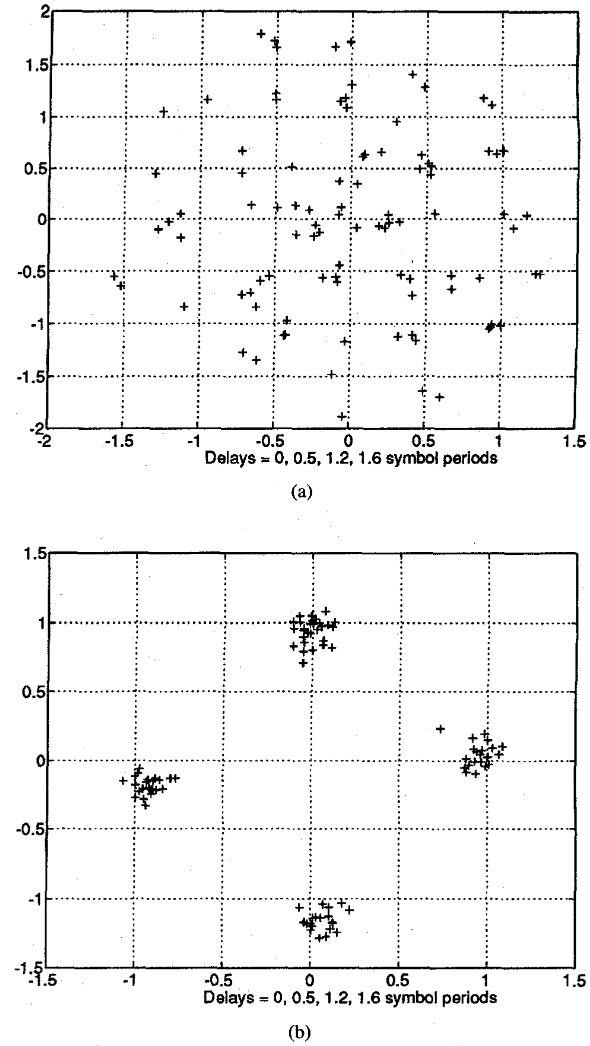


Fig. 7. Comparison of phase patterns: (a) Before equalization; (b) after equalization.

The proposed algorithm was used on 50 snapshots of data for channel identification. The estimated channels were then used for equalization, and the equalized signal phase pattern is given in Fig. 7(b). In comparison with Fig. 7(a), the constellation is clearly much improved.

VIII. Conclusion

In this paper, we present a blind algorithm for identifying multichannel FIR systems with *unknown* deterministic input. With proper channel order selection, the algorithm can accomplish blind identification based solely on the system outputs without any statistical information on the input process. Several useful results that characterize the finite deterministic input are given. Necessary and sufficient identifiability conditions regarding the input signal and channels are derived, which are useful in further algorithm development of deterministic blind identification. Computer simulations and RF experimental results demonstrate the potential of the proposed algorithm.

Appendix A
Rank Condition of a Hankel Matrix

We only need to consider the case for $r \leq N+1/2$, where the matrix $\mathbf{S}(r)$ is a skinny matrix because the rest can be proved by the symmetry property, i.e., $\mathbf{S}(r)^T = \mathbf{S}(N-r+1)$. The rank increase from $\mathbf{S}(1)$ to $\mathbf{S}(p)$ was obvious. The only thing we need to prove is that once $\mathbf{S}(r)$ becomes rank deficient, $\mathbf{S}(r+1)$ will also be rank deficient, and its rank will stay the same until the number of its rows is smaller than p.

Suppose that $\mathbf{S}(r)$ is rank deficient and its rank is $p(<r)$. By (10), we can see that $\mathbf{S}(r+1)$ is constructed by removing the last row of $\mathbf{S}(r)$ and adding a column on the right of $\mathbf{S}(r)$. Since $\mathbf{S}(r)$ is of rank p, then there are $r-p$ independent vectors $\{\mathbf{a}_i\}_{i=1}^{r-p}$ such that $\mathbf{S}(r)\mathbf{a}_i = \mathbf{0}$. Since the first r columns of $\mathbf{S}(r+1)$ are part of $\mathbf{S}(r)$, it is not hard to see that

$$\begin{aligned} \mathbf{S}(r+1)\begin{bmatrix}\mathbf{a}_i \\ 0\end{bmatrix} &= [\mathbf{S}'(r) \,\vdots\, \mathbf{c}]\begin{bmatrix}\mathbf{a}_i \\ 0\end{bmatrix} \\ &= \mathbf{S}'(r)\mathbf{a}_i \\ &= \mathbf{0} \end{aligned}$$

where $\mathbf{S}'(r)$ is $\mathbf{S}(r)$ without the last row, and $\mathbf{c}$ is the new column of $\mathbf{S}(r+1)$. By similar reasoning, we can also show that $\{[\begin{smallmatrix}0\\ \mathbf{a}_i\end{smallmatrix}]\}_{i=1}^{r-p}$ are also possible null vectors of $\mathbf{S}(r+1)$. Now, we want to prove that there are at least $r-p+1$ independent vectors among these $2(r-p)$ vectors, i.e., the rank of

$$\begin{matrix} r\{ \\ 1\{ \end{matrix} \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{r-p} & \vdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \vdots & \mathbf{a}_1 & \cdots & \mathbf{a}_{r-p} \end{bmatrix} \begin{matrix} \}1 \\ \}r \end{matrix} \tag{A.1}$$

is no smaller than $r-p+1$. First of all, it is easy to see that the last $r-p$ columns in (A.1) are independent since $\{\mathbf{a}_i\}_{i=1}^{r-p}$ are independent. Let us assume that the jth row of $[\mathbf{a}_1, \cdots, \mathbf{a}_{r-p}]$ is the first nonzero row, and at least one element, say $\mathbf{a}_m(j)$, $1 \leq m \leq r-p$, is nonzero. Due to the down shift and zero padding, the jth row of the last $r-p$ columns of (A.1) is actually the $(j-1)$th row of the first $r-p$ columns, which is a zero row. Therefore, it is obvious that $[\mathbf{a}_m^T, 0]^T$ cannot be expressed as a linear combination of $[0, \mathbf{a}_1^T]^T, \cdots, [0, \mathbf{a}_{r-p}^T]^T$. In other words, there are at least $r-p+1$ independent vectors in the matrix of (A.1). Let the column null space dimension be n. Then, $n \geq r-p+1$, and rank $\{\mathbf{S}(r+1)\} \triangleq p' = r+1-n \leq r+1-(r-p+1) = p$.

Since $\mathbf{S}(r+1)$ is rank deficient and its rank is p', then there are $(N-r)-p'$ independent row vectors $\{\mathbf{b}_i\}_{i=1}^{N-r-p'}$ such that $\mathbf{b}_i\mathbf{S}(r+1) = \mathbf{0}$. $\mathbf{S}(r)$ is constructed by removing the last column of $\mathbf{S}(r+1)$ and adding a new row to the last of $\mathbf{S}(r+1)$. By similar reasoning, we can show that there are at least $N-r-p'+1$ independent vectors among $\{[\mathbf{b}_i, 0]\}_{i=1}^{N-r-p'}$ and $\{[0, \mathbf{b}_i]\}_{i=1}^{N-r-p'}$. Therefore, the rank $\{\mathbf{S}(r)\} = p \leq N-r-(N-r-p') = p'$. According to another inequality shown above, $p' \leq p$, and hence, $p' = p$.

Now, we want to show that if $\mathbf{S}(p+1)$ is the first matrix becoming rank deficient, it must be of rank p. If rank $\{\mathbf{S}(p+1)\} < p$, then rank $\{\mathbf{S}(p)\} \leq$ rank$\{\mathbf{S}(p+1)\} < p$, which contradicts the fact that rank $\{\mathbf{S}(p)\}$ is of full rank, i.e., of rank p. □

Appendix B
Null Space of $S(L)$

From (1), $\mathbf{S}(p+1)$ is of rank p. There is a unique nontrivial null vector of $\mathbf{S}(p+1)$ (up to a scalar multiple), which can be expressed as

$$\mathbf{a} = [a(0) \quad a(1) \quad \cdots \quad a(p)]^T.$$

Because of the Hankel structure of $\mathbf{S}(r)$, as shown above

$$[\underbrace{0 \ \cdots \ 0}_{i-1} \quad \mathbf{a}^T \quad \underbrace{0 \ \cdots \ 0}_{r-p-i}]^T, \ i = 1, \cdots, r-p$$

must be in the null space of $\mathbf{S}(r), r > p$. In addition, by Lemma 1, $\mathbf{S}(r)$ is of rank p. Following the arguments of the proof of Lemma 1, it is easily seen that the $r-p$ vectors in the matrix

$$\underbrace{\begin{bmatrix} a(0) & & & \\ a(1) & a(0) & & \\ \vdots & a(1) & \ddots & \\ a(p) & \vdots & \ddots & a(0) \\ & a(p) & \vdots & a(1) \\ & & \ddots & \vdots \\ & & & a(p) \end{bmatrix}}_{r-p}. \tag{B.2}$$

are independent. Therefore, the columns of the matrix in (B.2) form the null space of $\mathbf{S}(r)$. □

Appendix C
Proof of Theorem 3

First of all, we want to clarify that condition 2 implies condition 1, i.e., if condition 2 is satisfied, condition 1 naturally holds. Since the matrix in (18) is of $M(L+1)+p$, condition 2 implies that the rank of such a matrix is $M(L+1)+p-1$. Therefore, we must guarantee that the number of its rows must be larger than its rank, i.e., $Mp \geq M(L+1)+p-1$ or $p \geq L+1+L/(M-1)$. In the following, we shall only prove that condition 2 is the necessary and sufficient condition.

By (13), $\mathbf{X}(L)\hat{\mathbf{h}} = \mathbf{0}$ can be rewritten as

$$\mathbf{S}(2L+1)(\mathbf{H}_i\hat{\mathbf{h}}_j - \mathbf{H}_j\hat{\mathbf{h}}_i) = 0; \quad i = 1, \cdots, M, \quad j = i+1, \cdots, M. \tag{C.3}$$

It is clear that $\mathbf{H}_i\hat{\mathbf{h}}_j - \mathbf{H}_j\hat{\mathbf{h}}_i$ must be in the null space of $\mathbf{S}(2L+1)$. When $\mathbf{S}(2L+1)$ is of rank $p < 2L+1$, by Lemma 3, a polynomial with any null vector as coefficients has roots $z_1, \cdots, z_p$. Therefore, (C.3) implies that the polynomials

$$h_i(z)\hat{h}_j(z) - h_j(z)\hat{h}_i(z), \quad i = 1, \cdots, M, \quad j = i+1, \cdots, M, \tag{C.4}$$

have roots $z_1, \cdots, z_p$, or

$$h_i(z_k)\hat{h}_j(z_k) - h_j(z_k)\hat{h}_i(z_k) = 0, \quad i = 1, \cdots, M, \quad j = i+1, \cdots, M, \quad k = 1, \cdots, p. \tag{C.5}$$

Necessary Part: We prove the necessary part by showing that if condition 2 does not hold, the channels are not identifiable. Clearly, $[1, \cdots, 1, -\mathbf{h}^T]^T$ is a solution to (18). If there is another independent solution, say $[g_1, \cdots, g_p, -\hat{\mathbf{h}}^T]^T$, then by plugging this solution into (18), we obtain $\hat{\mathbf{h}}(z_k) = g_k\mathbf{h}(z_k)$, $k = 1, \cdots, p$ or $\hat{h}_i(z_k) = g_k h_i(z_k)$, for $i = 1, \cdots, M$. Clearly, for any pair of (i, j)

$$\begin{aligned} h_i(z_k)\hat{h}_j(z_k) &= h_i(z_k) g_k h_j(z_k) \\ &= h_j(z_k) \underbrace{g_k h_i(z_k)}_{\hat{h}_i(z_k)} \\ &= h_j(z_k)\hat{h}_i(z_k). \end{aligned}$$

Therefore, $\hat{\mathbf{h}}$ is another solution.

Now, we want to make sure that the new channel estimate $\hat{\mathbf{h}}$ is independent of the true channel $\mathbf{h}$. If not, $\hat{\mathbf{h}} = \alpha\mathbf{h}$ for some constant α, and then, $(\alpha - g_k)\mathbf{h}(z_k) = \mathbf{0}$, which means either $g_k = \alpha$ or $\mathbf{h}(z_k) = \mathbf{0}$. If $g_k = \alpha$ for all k's, then it is easy to see that the new solution is dependent on $[1, \cdots, 1, -\mathbf{h}^T]^T$, which is not possible. Therefore, there is at least an l, such that $g_l \neq \alpha$ and $\mathbf{h}(z_l) = \mathbf{0}$, which means that all the channels $\{h_i(z)\}$ share the common zeros z_l. In this case, by Theorem 2, the channel is not uniquely identifiable.

Sufficient Part: If condition 2 holds, we shall first prove that $\{h_i(z)\}$ do not share any common zeros. If they do, i.e., $h_i(z) = (z - z_0)h_i'(z)$, then we can find another set of $\hat{h}_i = h_i'(z)(z - z_0')$, where $z_0' \neq z_0$ and $i = 1, \cdots, M$. In this case,[4]

$$\begin{aligned} \hat{h}_i(z_k) &= h_i'(z_k)(z_k - z_0') \\ &= \underbrace{h_i'(z_k)(z_k - z_0)}_{h_i(z_k)} \underbrace{\frac{z_k - z_0'}{z_k - z_0}}_{\triangleq g_k} \\ &= h_i(z_k) g_k. \end{aligned}$$

Thus, it is easily seen that $[g_1, \cdots, g_p, -\hat{\mathbf{h}}^T]^T$ will be another independent null vector of the matrix in (18).

Now, we are in a position to show that the channel is identifiable if condition 2 is satisfied. We shall show this by contradiction by first assuming that the system is not identifiable, i.e., there is another independent solution $\hat{\mathbf{h}}$ satisfying (C.3) or (C.4).

Let us first write (C.4) into a matrix form

$$\begin{bmatrix} \overline{\mathbf{H}}_1 & & 0 \\ & \ddots & \\ 0 & & \overline{\mathbf{H}}_p \end{bmatrix} \begin{bmatrix} \mathbf{Z}(z_1) \\ \vdots \\ \mathbf{Z}(z_p) \end{bmatrix} \hat{\mathbf{h}} = \mathbf{0}. \qquad \text{(C.6)}$$

Consider all possible pairs for each root z_k, and write down the equations in a matrix form

$$\underbrace{\begin{bmatrix} \cdots & & & & \cdots \\ -h_l(z_k) & & h_1(z_k) & & \\ & \ddots & \vdots & & \\ & -h_l(z_k) & h_{l-1}(z_k) & & \\ & & -h_{l+1}(z_k) & h_l(z_k) & \\ & & \vdots & & \ddots \\ & & -h_M(z_k) & & h_l(z_k) \\ \cdots & & & & \cdots \end{bmatrix}}_{\overline{\mathbf{H}}_k}$$

$$\underbrace{\mathbf{Z}(z_k)\hat{\mathbf{h}}}_{\hat{\mathbf{h}}(z_k)} = \mathbf{0}, \quad l = 1, \cdots, M. \qquad \text{(C.7)}$$

Among $\{h_i(z)\}_{i=1}^M$, we claim that there exists at least one channel, say, the lth channel, such that $h_l(z_k) \neq 0$. Otherwise, z_k will be the common root among all the channels. Therefore, just observing the portion of $\overline{\mathbf{H}}_k$ displayed in (C.7), we can easily see that $\overline{\mathbf{H}}_k$ is of rank $M - 1$ with a single null vector $\hat{\mathbf{h}}(z_k) = \mathbf{h}(z_k)$ (up to a scalar multiple). Therefore, the null space of $[\overline{\mathbf{H}}_1^T, \cdots, \overline{\mathbf{H}}_p^T]^T$ is spanned by the columns of the following matrix:

$$\begin{bmatrix} \mathbf{h}(z_1) & & \\ & \ddots & \\ & & \mathbf{h}(z_p) \end{bmatrix}. \qquad \text{(C.8)}$$

By (C.6), $[\mathbf{Z}^T(z_1), \cdots, \mathbf{Z}^T(z_p)]^T\hat{\mathbf{h}}$ should be in the null space, or it should be a linear combination of the columns of the matrix in (C.8). Therefore

$$\begin{bmatrix} \mathbf{h}(z_1) & & & \mathbf{Z}(z_1) \\ & \ddots & & \vdots \\ & & \mathbf{h}(z_p) & \mathbf{Z}(z_p) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \\ -\hat{\mathbf{h}} \end{bmatrix} = \mathbf{0}. \qquad \text{(C.9)}$$

Since $[\overbrace{1 \ \ 1 \ \cdots \ 1}^{p} \ \ -\mathbf{h}^T]^T$ is another independent solution to (C.9), the matrix in (C.9) or (18) has at least a 2-D null space, which contradicts condition 2. □

Remark 1: We derive the above results by assuming that the modes are *distinct*. If, however, there is a set of repeated roots, some of the above expressions need some modifications. Consider the root z_q of multiplicity $k(>1)$. The equations corresponding to this root are

$$\begin{aligned} &[h_i(z)\hat{h}_j(z) - h_j(z)\hat{h}_i(z)]_{z=z_q}^{(l)} = 0, \\ &l = 0, \cdots, k-1 \end{aligned} \qquad \text{(C.10)}$$

where $(\cdot)^{(l)}$ denotes the lth order derivative with respect to z. Following the same derivation from (C.7)–(C.9), we can obtain the same expression as (C.9), except that the corresponding block of vectors $\mathbf{h}(z_q)$ and matrices $\mathbf{Z}(z_q)$ must be replaced by the following:

$$\mathbf{h}^k(z_q) = \begin{bmatrix} \alpha_{11}\mathbf{h}(z_q) & & & \\ \alpha_{21}\mathbf{h}^{(1)}(z_q) & \alpha_{22}\mathbf{h}(z_q) & & \\ \vdots & \vdots & \ddots & \\ \alpha_{k1}\mathbf{h}^{(k-1)}(z_q) & \alpha_{k2}\mathbf{h}^{(k-2)}(z_q) & \cdots & \alpha_{kk}\mathbf{h}(z_q) \end{bmatrix};$$

$$\mathbf{Z}^k(z_q) = \begin{bmatrix} \mathbf{Z}(z_q) \\ \vdots \\ \mathbf{Z}^{(k-1)}(z_q) \end{bmatrix}$$

where $\alpha_{ij} = \binom{j}{i}$.

Remark 2: If ∞ is one of the input modes, the corresponding $\mathbf{h}(\infty)$ and $\mathbf{Z}(\infty)$ should be evaluated by $z^{-L}\mathbf{h}(z)_{z=\infty}$ and $z^{-L}\mathbf{Z}(z)_{z=\infty}$, respectively. □

[4] Here, we assume that $z_0 \neq z_k$, $k = 1, \cdots, p$. Otherwise, it is obvious that one column of the matrix in (18) will be zero, and condition 2 does not hold.

APPENDIX D
PROOF OF THEOREM 4

The proof of condition 1 has been done previously. Here, we only need to prove that condition 2 of Theorem 4 is the equivalent to its counterpart of Theorem 3.

Sufficient Part: We prove the sufficient part by showing that if condition 2 of Theorem 3 does not hold, neither does its counterpart of Theorem 4.

Suppose that (18) has an independent null vector, $[g_1, \cdots, g_p, -\hat{\mathbf{h}}^T]^T$; then

$$\begin{bmatrix} \mathbf{h}(z_1) & & 0 \\ & \ddots & \\ 0 & & \mathbf{h}(z_p) \end{bmatrix} \begin{bmatrix} g_1 \\ \vdots \\ g_p \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{h}}(z_1) \\ \vdots \\ \hat{\mathbf{h}}(z_p) \end{bmatrix} \quad \text{(D.11)}$$

or

$$h_i(z_k) g_k = \hat{h}_i(z_k), \quad k = 1, \cdots, p, \quad i = 1, \cdots, M. \quad \text{(D.12)}$$

Let us define a polynomial $g(z)$ with an order $\leq p-1$ such that $g(z_k) = g_k$. This can always be done by solving the following equation:

$$\begin{bmatrix} 1 & z_1 & \cdots & z_1^{p-1} \\ 1 & z_2 & \cdots & z_2^{p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_p & \cdots & z_p^{p-1} \end{bmatrix} \begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ g(p-1) \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_p \end{bmatrix} \quad \text{(D.13)}$$

where $g(l)$ is the lth-order coefficient of $g(z)$. For distinct[5] $\{z_k\}$, the Vandermonde matrix in (D.13) is of full rank, and the solution of (D.13) is *unique*. Then, let us define

$$P_i(z) \triangleq \hat{h}_i(z) - h_i(z) g(z). \quad \text{(D.14)}$$

By (D.12), $P_i(z_k) = \hat{h}_i(z_k) - h_i(z_k) g(z_k) = 0$, and $P_i(z)$ has at least p roots, i.e., $z_1, \cdots, z_p$. In other words, we can always factorize $P_i(z)$ into the product of $a(z) \triangleq (z - z_1) \cdots (z - z_p)$ and another polynomial $f_i(z)$, i.e., $P_i(z) = a(z) f_i(z)$. Hence, by (D.14)

$$\underbrace{\hat{h}_i(z)}_{L} = \underbrace{h_i(z)}_{L} \underbrace{g(z)}_{<p} + \underbrace{a(z)}_{p} \underbrace{f_i(z)}_{<L}. \quad \text{(D.15)}$$

Now, we only need to show that $g(z)$ has order larger than 0. If $g(z)$ is of zero order, then $g(z) = g$ and $g_k = g$ for all k's and some constant g. By (D.12), $\hat{h}_i(z_k) = g h_i(z_k)$ since $\hat{h}_i(z) - g h_i(z)$ is of order L and has $p (> > L)$ roots. Hence, $\hat{h}_i(z) \equiv g h_i(z)$ or $[g_1, \cdots, g_p, -\hat{\mathbf{h}}^T]^T = g[1, \cdots, 1, -\mathbf{h}^T]^T$, which means these two solutions are dependent. This clearly contradicts the assumption that both vectors are independent.

[5] If there are repeated modes, we need to slightly modify the Vandermonde matrix in (D.13) in the same fashion as in *Remark 1* after Theorem 3. However, the same conclusion still holds.

Necessary Part: We prove the necessary part by showing that if condition 2 of Theorem 4 does not hold, neither does its counterpart of Theorem 3.

Suppose there exists a $g(z)$ with order between 1 and $p-1$ such that (D.15) holds. Then, by reversing the order of the derivation of the sufficient part, we can easily show that $[g(z_1), \cdots, g(z_p), -\hat{\mathbf{h}}^T]^T$ is another solution to (18). Now, we need to show that it is independent of $[1, \cdots, 1, -\mathbf{h}^T]^T$. If they are dependent, then $[g(z_1), \cdots, g(z_p), -\hat{\mathbf{h}}^T]^T = \alpha[1, \cdots, 1, -\mathbf{h}^T]^T$ or $g(z_k) = \alpha$ and $\hat{\mathbf{h}} = \alpha \mathbf{h}$ for a nonzero constant α. Since $g(z)$ is a polynomial with order $<p$, $g(z) = \alpha$ for p distinct values means that $g(z) \equiv \alpha$ or $g(z)$ is an zero-order polynomial that contradicts the assumption that $g(z)$ has an order between 1 and $p-1$. Hence, $[g(z_1), \cdots, g(z_p), -\hat{\mathbf{h}}^T]^T$ is another independent solution, and condition 2 of Theorem 3 does not hold. □

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for insightful comments and helpful critiques of the manuscript.

REFERENCES

[1] L. A. Baccala and S. Roy, "A new blind time-domain channel identification method based on cyclostationarity," in *Proc. 26th Conf. Inform. Sci. Syst.,* Princeton, NJ, Mar. 1994.

[2] E. W. Bai and S. S. Sastry, "Persistency of excitation, sufficient richness, and parameter convergence in discrete time adaptive control," *Syst. Contr. Lett.,* vol. 6, pp. 153–163, 1985.

[3] A. Benveniste and M. Goursat, "Blind equalizers," *IEEE Trans. Commun.,* vol. COM-32, no. 8, pp. 871–883, Aug. 1984.

[4] A. Benveniste, M. Goursat, and G. Ruget, "Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications," *IEEE Trans. Automat. Contr.,* pp. 385–399, June 1980.

[5] Z. Ding, R. A. Kennedy, B. D. O. Anderson, and C. R. Johnson, "Ill-convergence of Godard blind equalizers in data communication systems," *IEEE Trans. Commun.,* vol. 39, pp. 1313–1327, Sept. 1991.

[6] J.-J Fuchs, "Estimating the number of sinusoids in additive white noise," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 36, no. 12, pp. 1846–1853, 1988.

[7] W. A. Gardner, *Introduction to Random Processes with Application to Signals and Systems.* New York: Macmillan, 1985.

[8] ——, "A new method of channel identification," *IEEE Trans. Commun.,* vol. 39, no. 6, pp. 813–817, June 1991.

[9] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communications systems," *IEEE Trans. Commun.,* vol. COM-28, no. 11, pp. 1867–1875, Nov. 1980.

[10] G. H. Golub and C. F. Van Loan, *Matrix Computations.* Baltimore, MD: Johns Hopkins Univ. Press, 1990.

[11] M. İ. Gürelli and C. L. Nikias, "EVAM: An eigenvector-based algorithm for multichannel blind deconvolution of input colored signals," *IEEE Trans. Signal Processing,* to be published.

[12] D. Hatzinakos and C. Nikias, "Estimation of multipath channel response in frequency selective channels," *IEEE J. Select. Areas Commun.,* vol. S7, no. 1, pp. 12–19, Jan. 1989.

[13] T. Kailath, *Linear Systems.* Englewood Cliffs, NJ: Prentice-Hall, 1980.

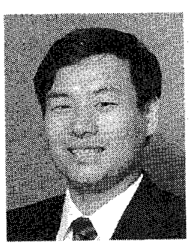
[14] Y. Li and Z. Ding, "Blind channel identification based on second-order cyclostationary statistics," in *Proc. ICASSP'93,* Minneapolis, MN Apr. 1993, pp. 81–84.

[15] H. Liu, G. Xu, and L. Tong, "A deterministic approach to blind equalization," in *Proc. Asilomar Conf. Signals, Syst., Comput.,* Pacific Grove, CA, Nov. 1993.

[16] L. Ljung, *System Identification.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[17] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," in *Proc. IEEE ICASSP'94,* Apr. 1994, pp. IV573–IV576.

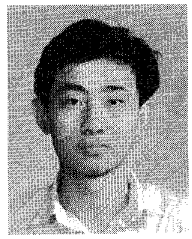
[18] A. P. Petropulu and C. L. Nikias, "Blind deconvolution using signal reconstruction from partial higher order cepstral information," *IEEE Trans. Signal Processing,* vol. 41, no. 6, pp. 2088–2095, June 1993.
[19] A. P. Petropulu, "Identifiability of blind deconvolution of nonwhite signals," in *Proc. 26th Asilomar Conf. Signals, Syst., Comput.,* Pacific Grove, CA, Nov. 1993.
[20] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation," *IEEE Trans. Commun.,* vol. COM-23, no. 6, pp. 679–682, June 1975.
[21] S. V. Schell, D. L. Smith, and S. Roy, "Blind channel identification using subchannel response matching," in *Proc. 1994 Conf. Inform. Sci. Syst.,* Princeton, NJ.
[22] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Inform. Theory,* vol. 36, no. 2, pp. 312–320, Mar. 1990.
[23] D. T. M. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction," in *Proc. IEEE ICASSP'94,* Apr. 1994, pp. IV585–IV588.
[24] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channel," in *Proc. 25th Asilomar Conf. Signals, Syst., Comput.,* Pacific Grove, CA, Nov. 1991.
[25] ——, "Blind identification and equalization using spectral measures, Part II: A time domain approach," in *Cyclostationarity in Communications and Signal Processing,* W. A. Gardner, Ed. New York: IEEE, 1993.
[26] J. K. Tugnait, "Identification of linear stochastic system via second and fourth-order cumulant matching," *IEEE Trans. Inform. Theory,* vol. IT-33, pp. 393–407, May 1987.
[27] ——, "On blind equalization of multipath channels using fractional sampling and second-order cyclostationary statistics," in *Proc. Globecom'93,* Houston, TX, Dec. 1993.
[28] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Sppech, Signal Processing,* vol. ASSP-33, no. 2, pp. 387–392, Apr. 1985.
[29] G. Xu, Y. Cho, and T. Kailath, "Applications of fast signal subspace decomposition to signal processing and communications problems," *IEEE Trans. Signal Processing,* vol. 42, no. 6, pp. 1453–1461, June 1994.
[30] G. Xu and T. Kailath, "Fast subspace decomposition," *IEEE Trans. Signal Processing,* vol. 42, no. 3, pp. 539–551, Mar. 1994.
[31] G. Xu, H. Zha, G. Golub, and T. Kailath, "Fast and robust algorithms for updating signal subspaces," *IEEE Trans. Circuits Syst.,* vol. 41, no. 6, pp. 537–549, June 1994.



Guanghan Xu (S'86–M'92) received the B.S. degree with honors in biomedical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1985, the M.S. degree in electrical engineering from Arizona State University, Tempe, USA, in 1988, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1991.

During the summer of 1989, he was a Research Fellow at the Institute of Robotics, Swiss Federal Institute of Technology, Zurich, Switzerland. From 1990 to 1991, he was a General Electric Fellow in the Fellow-Mentor-Advisor Program at the Center of Integrated Systems, Stanford University. From 1991 to 1992, he was a Research Associate in the Department of Electrical Engineering, Stanford University, and a short-term visiting scientist at the Laboratory of Information and Decision Systems at the Massachusetts Institute of Technology, Cambridge, USA. In 1992, he joined the faculty of the Department of Electrical and Computer Engineering at The University of Texas at Austin. He has worked in several areas including signal processing, communications, numerical linear algebra, multivariate statistics, and semiconductor manufacturing. His current research interest is focused on smart antenna systems for wireless communications.

Dr. Xu is a member of Phi Kappa Phi and is recipient of the 1995 NSF CAREER Award.



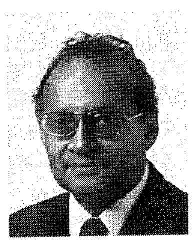
Hui Liu received the B.S. degree in 1988 from Fudan University, Shanghai, China, the M.S. degree in 1992 from Portland State University, Portland, OR, USA, and the Ph.D degree in 1995 from The University of Texas at Austin, all in electrical engineering.

From September 1992 to December 1992, he was a software engineer at Quantitative Technology Corp., Beaverton, OR, USA. During the summer of 1995, he was a consultant for Bell Northern Research, Richardson, TX. He joined the faculty of the Department of Electrical Engineering at the University of Virginia, Charlottesville, USA, in September of 1995.

Lang Tong (S'87–M'91) received the B.E. degree in control engineering from Tsinghua University, Beijing, China, in 1985, and the M.S. and Ph.D. degrees in electrical engineering in 1987 and 1990, respectively, from the University of Notre Dame, Notre Dame, IN, USA.

After being a Postdoctoral Research Affiliate at the Information Systems Laboratory, Stanford University, he joined the Department of Electrical and Computer Engineering, West Virginia University, Mogantown, USA. Since 1993, he has been with the Department of Electrical and Systems Engineering, the University of Connecticut, Storrs, USA. He also held a Visiting Assistant Professor position at Stanford University in the summer of 1992. His research interests include wireless communication, signal processing, and system theory.

Dr. Tong has received the Best Graduating Student Award from the University of Notre Dame and the Outstanding Young Author Award from the IEEE Circuits and Systems Society.



Thomas Kailath (F'70) was educated in Poona, India, and at the Massachusetts Institute of Technology, Cambridge, USA, where he received the S.M. degree in 1959 and the Sc.D. degree in 1961.

From October 1961 to December 1962, he worked at the Jet Propulsion Laboratories, Pasadena, CA, USA, where he also taught part time at the California Institute of Technology. He then came to Stanford, where he served as Director of the Information Systems Laboratory from 1971 through 1980, as Associate Chairman from 1981 to 1987, and currently holds the Hitachi America Professorship in Engineering. He has also held short-term appointments at several institutions around the world. He has worked in a number of areas including information theory, communications, computation, control, signal processing, VLSI design, statistics, linear algebra, and operator theory. His recent interests include applications of signal processing, computation, and control to problems in semiconductor manufacturing and wireless communications. He is the author of *Linear Systems* (Englewood Cliffs, NJ: Prentice-Hall, 1980) and *Lectures on Wiener and Kalman Filtering* (New York: Springer-Verlag, 1981).

Dr. Kailath has held Guggenheim, Churchill, and Royal Society Fellowships, among others, and received awards from the IEEE Information Theory Society and the American Control Council, in addition to the Technical Achievement and Society Awards of the IEEE Signal Processing Society. He served as President of the IEEE Information Theory Society and has been awarded honorary doctorates by Linkoping University, Sweden, and by Strathclyde University, Scotland. He is a Fellow of the Institute of Mathematical Statistics and is a member of the National Academy of Engineering.