Blind Identification and Equalization Based on Second-Order Statistics: A Time Domain Approach

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

***Abstract*— A new *blind* channel identification and equalization method is proposed that exploits the cyclostationarity of oversampled communication signals to achieve identification and equalization of possibly nonminimum phase (multipath) channels without using training signals. Unlike most adaptive blind equalization methods for which the convergence properties are often problematic, the channel estimation algorithm proposed here is asymptotically exact. Moreover, since it is based on second-order statistics, the new approach may achieve equalization with fewer symbols than most techniques based only on higher-order statistics. Simulations have demonstrated promising performance of the proposed algorithm for the blind equalization of a three-ray multipath channel.**

***Index Terms*—Intersymbol interference, equalization, channel identification, blind identification, cyclostationary processes.**

I. Introduction

Intersymbol interference (ISI) is a limiting factor in many communication environments. Intersymbol interference can arise from time-varying multipath fading, which can be severe in, for example, a mobile communication environment. Other channel impairments that contribute to ISI include symbol clock residual jitter, carrier phase jitter, etc. To achieve high-speed reliable communication, channel identification and equalization are necessary to overcome the effects of ISI. Traditionally, channel identification and equalization are achieved either by sending training sequences, or by designing the equalizer based on *a priori* knowledge of the channel. The latter approach is often not suitable for a radio communication environment since little knowledge about such a channel can be assumed *a priori.* The standard adaptive approach, though attractive in handling time-variant channels, has to waste a fraction of the transmission time for a training sequence. Indeed even the so-called decision feedback equalization (DFE), which does not explicitly use a training sequence, requires sending known training sequences periodically to avoid catastrophic error propagation [4].

Manuscript received September 27, 1991; revised April 30, 1993. This work was supported in part by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Office of Scientific Research under Contract F49620-90-C-0014 and by the Army Research Office under Contract DAAL03-90-G-0108, and by the Joint Service Program at Stanford University under Contract DALL03-88-c-0011. This paper was presented in part at the 25th Asilomar Conference on Signals, Systems and Computers, November 4–6, 1991.

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

G. Xu is with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA.

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

IEEE Log Number 9400228.

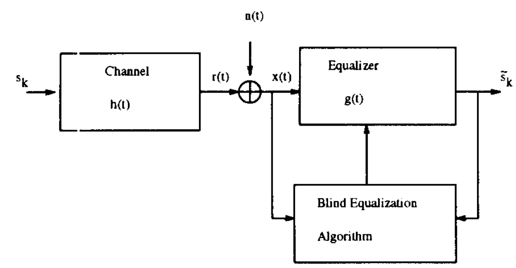

Fig. 1. A schematic of blind equalization.

In contrast to the standard adaptive equalization methods, the so-called blind equalization methods do not require a training sequence. Instead, the statistical properties of the transmitted signals are exploited to carry out the equalization at the receiver *without* access to the symbols being transmitted. A blind equalization scheme is shown in Fig. 1. Instead of choosing the equalizer so that the equalized output sequence $\{\tilde{s}_k\}$ is close to the source symbol sequence $\{s_k\}$, as in the standard equalization formulation, in blind equalization one chooses the equalizer so that the *statistics* of the equalized output sequence $\{\tilde{s}_k\}$ are close to the *statistics* of the source symbol sequence $\{s_k\}$.

A. Existing Blind Channel Identification and Equalization Methods

The innovative idea of self-recovering (blind) adaptive equalization was first proposed by Sato [20], then further developed by Godard [12], Treichler and Agee [24], Benveniste and Goursat [1], Picci and Prati [19], Foschini [6] and more recently Shalvi and Weinstein [21]. Although the blind equalization schemes proposed so far are technically different, they are all derived via some optimization criteria involving certain higher-order statistics (cumulants) of the observation; various gradient-based algorithms are then employed to achieve the optimization. The major problem with the adaptive blind equalization techniques is their slow convergence and many local extrema attractors, as shown in [3] and [15]. As specially pointed out in [15], the global convergence may be jeopardized when the channel has finite impulse response. However, it should be noted that the optimization criterion proposed by Shalvi and Weinstein [21] ensures global optimization whenever ideal equalizers exist.

A different "blind" approach was proposed by Hatzinakos and Nikias [14]. The received signal sampled at the baud rate is modeled by a moving-average process. The multipath channel is then identified from the trispectrum of the received signals. The advantage of this method over the adaptive blind equalization methods is that the algorithm will provide *exact* identification of a possibly nonminimum phase channel, whenever the higher order cumulants and the trispectrum of the observation can be estimated accurately. However, the proposed algorithm is computationally intensive and suffers from the fact that the estimates of higher order statistics usually converge slower than those for second-order statistics. Moreover, since the received signal is sampled at the baud rate, it may also be sensitive to uncertainties associated with timing recovery, unknown phase jitter, and frequency offset. Finally, most current techniques ignore the presence of noise, or assume it to be Gaussian. The effect of non-Gaussian noise may affect the convergence and the performance of the cumulant-based approach.

In contrast to the case when a channel is driven by a stationary process, the second-order statistics of the channel output do contain some phase information of the channel when the input process is nonstationary. For applications in communications, many types of signals exhibit a particular type of nonstationarity called cyclostationarity [8], [9]. The exploitation of cyclostationarity has shown promising results in various applications such as detection and filtering of communication signals, parameter estimation, direction finding, identification of nonlinear systems, etc. (see the recent review [8] and the references therein). In the context of channel identification and equalization, Gardner was perhaps the first to recognize that the phase information is available in the periodically time-varying data correlation function [7]; however the proposed method still relies on a slow-data-rate training period. In an earlier work by Cerrato and Eisenstein [2], the problem of blind deconvolution of cyclostationary signals was approached in the frequency domain with the assumption that the energy of the channel impulse response is confined to an interval of size T; unfortunately, such an assumption is too restrictive for a multipath communication channel. Kormylo and Mendel [17] treated the case when the channel input is *nonstationary* and *Gaussian*.

B. A New Blind Channel Identification and Equalization Method

In this paper, we propose a new blind channel identification and equalization method. The main features of our approach can be summarized as follows.

1) The algorithm provides *exact* identification of possibly *nonminimum phase* channels, if the correlation function of the received signal is known exactly. More realistically, when the estimated signal correlation is used, asymptotic exactness can be established.
2) The algorithm is insensitive to the uncertainties associated with timing recovery. This is achieved by exploiting the so-called signal subspace structure and by sampling the received signal at a rate higher than the baud rate, as is done in fractionally spaced equalization [25]. It was shown in [11] that the oversampling also provides better immunity to noise, interference, and frequency selective fading than that provided by symbol-spaced equalization.
3) The algorithm can be used to initialize various adaptive schemes and the equalized output can be used to facilitate decision feedback adaptation. The identified channel can be used to implement maximum likelihood sequence estimation [5] to further reduce ISI.
4) The algorithm relies only on the second-order statistics of the received signal. Therefore it usually requires fewer symbols than most of the schemes suggested to date. In our simulation at a signal-to-noise ratio (SNR) of 30 dB, 100 symbols were sufficient to give a good estimate of the channel, a much smaller number of symbols than that required by most alternatives. This fact also implies that the algorithm can be used to estimate channels with relatively rapid channel variation.
5) Unlike other equalization methods, the new method provides *exact* reconstruction of the source symbols via a *causal* operation (FIR filter) when the channel (possibly nonminimum phase) can be identified correctly and when there is no noise. Such exact reconstruction cannot be achieved by inverting a nonminimum phase channel.
6) There is no restriction imposed on the probability distribution of the source symbols. The random source may be real or complex, continuous or discrete, or even Gaussian, in contrast to the assumptions made in [21] and most current techniques using higher-order statistics.

It may seem unexpected at first glance that a nonminimum phase system can be identified using only the second-order statistics of the system output. However, it is less surprising when the system is driven by a *nonstationary* process. This is indeed the case for most communication channels where the input signals are cyclostationary (or periodically correlated) rather than stationary.

C. Problem Statement

When the channel is time invariant, the received complex baseband signal $x(\cdot)$ can be expressed as

$$r(t) = \sum_{k=-\infty}^{\infty} s_k h(t - kT), \tag{1}$$

$$x(t) = r(t) + n(t) \tag{2}$$

where

s_k: an information symbol in a signal constellation $\mathcal{S}$,

$h(\cdot)$: the discrete-time "composite" channel impulse response that includes the pulse shaping filter, the channel, and receiving filters,

T: the symbol interval,

$n(\cdot)$: the additive noise.

The objective of *blind* channel identification is to estimate $h(\cdot)$ given only the received signal $x(\cdot)$. Once such identification is achieved, the estimation of the information symbols s_k's becomes more or less a solved problem, and various equalization and sequence estimation techniques can be applied.

We assume the following throughout the sequel.

1) The symbol interval T is known and is an integer multiple of the sampling period.
2) The impulse response $h(\cdot)$ has finite support.
3) $\{s_k\}$ is zero mean, and $E(s_k s_l^*) = \delta(k-l)$ where $\delta(t)$ is the discrete-time impulse function.
4) $n(\cdot)$ is zero mean, uncorrelated with $\{s_k\}$, and $E(n(t_1)n^*(t_2)) = \sigma^2\delta(t_1 - t_2)$.

As a general notational convention, symbols for matrices (in capital letters) and vectors are in boldface. The notations $(\cdot)^H$, $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^\dagger$ stand for hermitian, transpose, complex conjugate, and the Moore–Penrose generalized inverse, respectively. The symbol $\boldsymbol{I}$ ($\boldsymbol{0}$) stands for the identity (zero) matrix of an appropriate dimension. $E(\cdot)$ denotes expectation.

II. Exploitation of Cyclostationarity: A Vector Representation

We first examine the idea that the *spectral redundancy* in a cyclostationary signal [8]–[10], a second-order statistical property, can be used to identify a nonminimum phase channel. A complete discussion of the issue of channel identifiability in the frequency domain is developed in [22]. Nonetheless, the following brief discussion provides some motivation for our method of exploiting the cyclostationary nature of the observation.

We shall assume that the received signal $x(t)$ is wide-sense cyclostationary, i.e.,

$$E(x(t_1)x^*(t_2)) = E(x(t_1+T)x^*(t_2+T)). \tag{3}$$

The discrete-time sequence $x(i)$ obtained by sampling $x(\cdot)$ may or may not be cyclostationary, depending on the sampling rate. In fact, if the received signal is sampled at the baud rate $1/T$, then

$$x(iT) = \sum_{k=-\infty}^{\infty} s_k h((i-k)T) + n(iT) \tag{4}$$

will be a wide sense *stationary* process. In this case, it is well known that only minimum-phase channels can be identified from the second-order statistics of such a wide sense stationary sequence. In other words, the phase information of the channel is lost in the second-order statistics when its output is sampled at the baud rate. On the other hand, if the sampling rate is higher than the baud rate, the resulting output sequence is wide sense *cyclostationary*. The second-order statistics of the over-sampled observation do contain the phase information of the channel.

Consider the noiseless case and an *oversampled* discrete-time signal $x(i\Delta)$, denoted as $x(i)$,

$$x(i) = \sum_{k=-\infty}^{\infty} s_k h(i-kT), \tag{5}$$

$$= u(i) * h(i), \tag{6}$$

where $*$ stands for convolution, and $u(i)$ is a nonstationary sequence given by

$$u(i) = \sum_{k=-\infty}^{\infty} s_k \delta(i-kT). \tag{7}$$

Let

$$R_x(i_1, i_2) = E(x(i_1)x^*(i_2)), \tag{8}$$

$$R_u(i_1, i_2) = E(u(i_1)u^*(i_2)). \tag{9}$$

It can be shown (see, e.g., [18]) that, in the frequency domain,

$$\Gamma_x(\omega, \nu) = \Gamma_u(\omega, \nu)H(\omega)H^*(-\nu) \tag{10}$$

where $\Gamma_x(\omega, \nu)$, $\Gamma_u(\omega, \nu)$, and $H(\omega)$ are the Fourier transforms of $R_x(i_1, i_2)$, $R_u(i_1, i_2)$, and $h(i)$, respectively. When $u(i)$ is stationary, so is $x(i)$. Denoting

$$R_u(\tau) = R_u(t_1, t_2), \qquad \tau = t_1 - t_2, \tag{11}$$

$$S_u(\omega) = \sum_{\tau} R_u(\tau)e^{-j\omega\tau}, \tag{12}$$

we then have the familiar form

$$S_x(\omega) = S_u(\omega)H(\omega)H^*(\omega) \tag{13}$$

which shows that it is not possible to recover $H(\omega)$ completely from $S_x(\omega)$ and $S_u(\omega)$ unless $H(\omega)$ is a minimum phase system. However, when $u(i)$ is nonstationary, it is clear from (10) that $\Gamma_x(\omega, \nu)$ contains the phase information of the channel. The question now is how to identify the channel given $R_x(i_1, i_2)$ using the fact that $u(i)$ is cyclostationary. While the problem can be solved in the frequency domain [22], the approach presented here is based on a time-domain vector representation of the cyclostationary process $x(\cdot)$.

A. Vector Representation

An important observation is that, under the first assumption, the signal space of the observation restricted to any finite time interval I (referred to as an *observation interval*) is a *linear* space spanned by a basis comprised of time-shifted copies of $h(\cdot)$. Specifically, the signal space $\mathcal{R}_{(\mathcal{S},h)}(I)$, or simply $\mathcal{R}(I)$, is defined by

$$\mathcal{R}(I) = \left\{ r(t) \mid r(t) = \sum_{k=-\infty}^{\infty} s_k h(t-kT),\ s_k \in \mathcal{S},\ t \in I \right\}. \tag{14}$$

Moreover, when the observation interval I is finite and the channel impulse response has only finite support, $\mathcal{R}(I)$ is finite dimensional and its dimension can be computed directly from knowledge of I and the duration of $h(\cdot)$. Consider, for example, the received (noise free) signal $r(\cdot)$ in an observation interval $I_0 = (t_0, t_0 + L)$ for some finite L. The signal space $\mathcal{R}(I_0)$ is spanned by $h(t - K_0T), \cdots, h(t - (K_0 + (d-1))T)$ with the $h(t - iT)$ being restricted to $(t_0, t_0 + L)$, as illustrated by Fig. 2.

The integer K_0 and the dimension d depend on t_0, L and the duration L_h of the impulse response $h(\cdot)$, and can be roughly

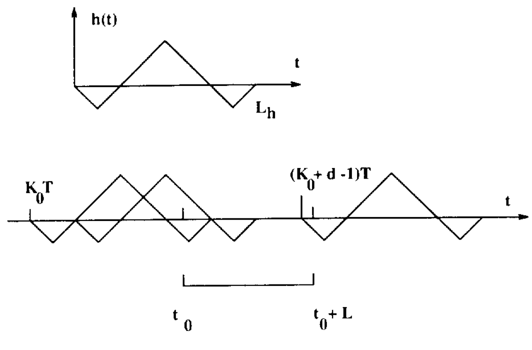

Fig. 2. Basis functions in a sampling interval.

estimated as follows:

$$K_0 = \left\lceil \frac{t_0 - L_h}{T} \right\rceil, \quad (15)$$

$$d = \left\lfloor \frac{t_0 - (K_0 - 1)T + L}{T} \right\rfloor \quad (16)$$

where $\lceil x \rceil$ ($\lfloor x \rfloor$) stands for the smallest (greatest) integer that is greater than or equal to (less than or equal to) x.

If $x(\cdot)$ is sampled in $(t_0, t_0 + L)$ with a sampling period Δ, we will have

$$x(t_0 + i\Delta) = \sum_{k=0}^{d-1} s_{K_0+k} h(t_0 + i\Delta - (K_0 + k)T) + n(t_0 + i\Delta), \quad i = 1, \cdots, m. \quad (17)$$

A matrix formulation of the above will be convenient. We can write

$$\boldsymbol{x}(t_0) = \boldsymbol{H}(t_0)\boldsymbol{s}(t_0) + \boldsymbol{n}(t_0) \quad (18)$$

where (see (20) below)

$$\boldsymbol{x}(t_0) = [x(t_0 + \Delta), \cdots, x(t_0 + m\Delta)]^T, \quad (19)$$

$$\boldsymbol{s}(t_0) = [s_{K_0}, \cdots, s_{K_0+d-1}]^T, \quad (21)$$

$$\boldsymbol{n}(t_0) = [n(t_0 + \Delta), n(t_0 + 2\Delta), \cdots, n(t_0 + m\Delta)]^T. \quad (22)$$

Similarly, if the sampling is performed in any interval of the form $(t_0 + nT, t_0 + nT + L)$, replacing t_0 and K_0 by $t_0 + nT$ and $K_0 + nT$ in (18) will yield

$$\boldsymbol{x}(t_0 + nT) = \boldsymbol{H}(t_0 + nT)\boldsymbol{s}(t_0 + nT) + \boldsymbol{n}(t_0 + nT). \quad (23)$$

Making a closer examination of the (i, j)th entry of $\boldsymbol{H}(t_0 + nT)$, we see that

$$h(t_0 + nT + i\Delta - (K_0 + n + j - 1)T) = h(t_0 + i\Delta - (K_0 + j - 1)T). \quad (24)$$

Hence, for all integers n,

$$\boldsymbol{H}(t_0) = \boldsymbol{H}(t_0 + nT). \quad (25)$$

We then have

$$\boldsymbol{x}(t_0 + nT) = \boldsymbol{H}(t_0)\boldsymbol{s}(t_0 + nT) + \boldsymbol{n}(t_0 + nT), \quad n = 0, 1, \cdots. \quad (26)$$

When no confusion arises, we shall drop t_0 (or let $t_0 = 0$) for the sake of simplicity, with the understanding that a certain t_0 is fixed throughout. Then the vector representation of the received signal is

$$\boldsymbol{x}(iT) = \boldsymbol{H}\boldsymbol{s}(iT) + \boldsymbol{n}(iT), \qquad i = 0, 1, \cdots, \quad (27)$$

where $\boldsymbol{x}(iT)$ and $\boldsymbol{n}(iT)$ are m-dimensional vectors formed from the m samples of $x(\cdot)$ and $n(\cdot)$ inside the interval $(t_0 + iT, t_0 + iT + L)$, respectively, $\boldsymbol{H} = \boldsymbol{H}(t_0)$, independent of i, is given by (20), and $\boldsymbol{s}(iT)$ is a d-dimensional vector consisting of symbols that have "contributions" to the received signal inside the observation interval $(t_0 + iT, t_0 + iT + L)$ given by (2). The vector $\boldsymbol{x}(iT)$ can be obtained via time-division demultiplexing [9].

B. Relations Between Scalar and Vector Representations

It is clear from the previous development that the vector representation merely rearranges the data samples of the (scalar) observation. Such a rearrangement should not result in any loss of information. Nevertheless, it is important to note the one-to-one correspondence between the matrix $\boldsymbol{H}(t_0)$ and the sampled impulse response $h(i\Delta)$. It is clear from (20) that $\boldsymbol{H}(t_0)$ can be formed from the sampled $h(\cdot)$ once t_0, L, and Δ are chosen, and L_h and T are known. It is less obvious that one can form the sampled $h(\cdot)$ from $\boldsymbol{H}(t_0)$ when t_0, L, Δ, and T are known, but L_h is not explicitly known. Indeed, L_h is unknown in practice, while t_0, Δ, and L can be chosen and the symbol interval T is known *a priori*. However, there are no conceptual difficulties in reconstructing $h(i\Delta)$ from $\boldsymbol{H}(t_0)$ as suggested in Fig. 2. Keep in mind that the columns of $\boldsymbol{H}(t_0)$ are simply time-shifted copies of the sampled $h(\cdot)$, truncated by the observation window, and time-shifted exactly by integer multiples of T. The simplest case is when the observation interval is longer than the duration of the impulse response, i.e., $L \geq L_h$. In such a case, the column of $\boldsymbol{H}(t_0)$ with the maximum 2-norm gives the sampled impulse response. When $L < L_h$, one can imagine that the columns of $\boldsymbol{H}(t_0)$ are sequences of "snapshots" of the shifted impulse responses, observed through an observation window of size L. The explicit construction is as follows. Given an $m \times d$ matrix $\boldsymbol{H}(t_0)$ with its (i, j)th entry denoted as H_{ij}, suppose that the sampling rate is T_s times faster than the baud rate, i.e., $T = T_s\Delta$. Then a sampled channel impulse response h_i,

$$\boldsymbol{H}(t_0) = \begin{pmatrix} h(t_0 + \Delta - K_0T) & \cdots & h(t_0 + \Delta - (K_0 + d - 1)T) \\ \vdots & \cdots & \vdots \\ h(t_0 + m\Delta - K_0T) & \cdots & h(t_0 + m\Delta - (K_0 + d - 1)T) \end{pmatrix}, \quad (20)$$

written as a vector $\boldsymbol{h} = [h_1, \cdots, h_{(d-1)T_s}]$, can be recovered from a vector $\tilde{\boldsymbol{h}}$ constructed from $\boldsymbol{H}(t_0)$ as

$$\tilde{\boldsymbol{h}} = [H_{1d} \cdots H_{T_s d} H_{1(d-1)} \cdots H_{T_s(d-1)} \cdots H_{12} \cdots H_{T_s 2} H_{11} \cdots H_{m1}], \quad (28)$$

and

$$\tilde{\boldsymbol{h}} = [\boldsymbol{0}, \boldsymbol{h}, \boldsymbol{0}]. \quad (29)$$

In other words, $\tilde{\boldsymbol{h}}$ is an extension of $\boldsymbol{h}$ in the sense that zeros are added to both sides of the actual sampled impulse response.

Despite the algebraic equivalence between the two representations, there is an important difference between the original scalar model representation (1) and the vector model representation (26). It is easily shown that the received signal $r(\cdot)$ is *cyclostationary* with period T while the "vectorized" process $\boldsymbol{x}(i)$ is *stationary*; this is not surprising because the process of (over) sampling of the received signal is equivalent to the so-called *translation series representation* or *time series representation* of the cyclostationary process $r(\cdot)$ [9].

III. A New Channel Identification and Equalization Method

A. Channel Identifiability

To simplify the presentation of the key ideas with respect to channel identifiability, we ignore the noise for the moment. Similar results can be obtained when the noise (second-order) statistics are known.

Following the development in the previous section, the *blind* channel identification problem can be restated as follows.

Consider a vector process $\boldsymbol{x}(i)$, $i = 1, \cdots$, obtained from a linear model

$$\boldsymbol{x}(i) = \boldsymbol{H}\boldsymbol{s}(i), \qquad i = 0, 1, \cdots \quad (30)$$

along with the following constraints.

1) H is an $m \times d$ complex matrix of full column rank.
2) $\boldsymbol{s}(i)$ is a zero mean stationary process with autocorrelation function

$$\boldsymbol{R_s}(k) = E(\boldsymbol{s}(i)\boldsymbol{s}^H(i-k)) \quad (31)$$

of the following form:

$$\boldsymbol{R_s}(k) = \boldsymbol{J}^k, \qquad k \geq 0, \quad (32)$$

$$\boldsymbol{R_s}(k) = (\boldsymbol{J}^H)^{|k|}, \qquad k < 0, \quad (33)$$

where $\boldsymbol{J}$ is a $d \times d$ "shifting" matrix

$$\boldsymbol{J} = \begin{pmatrix} 0 & 0 \cdots & 0 & 0 \\ 1 & 0 \cdots & 0 & 0 \\ 0 & 1 \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 \cdots & 1 & 0 \end{pmatrix}. \quad (34)$$

The objective of blind channel identification and equalization is to identify $\boldsymbol{H}$ (channel identification) and estimate $\boldsymbol{s}(i)$ from $\boldsymbol{x}(i)$ (channel equalization).

Remarks:

1) In most practical cases, the rank constraint on $\boldsymbol{H}$ can be satisfied easily by oversampling. In fact, it can be shown that the full rank condition is necessary for the channel to be identifiable from the second-order statistics [23], [22].
2) While the original received signal $r(\cdot)$ is a *cyclostationary* process, the vectorized process $\boldsymbol{x}(i)$ is a wide sense *stationary* process whose correlation matrices can be estimated conveniently by time averaging.
3) The constraint on the autocorrelation function is obtained by observing from (21) that

$$\boldsymbol{s}(t_0) = [s_{K_0}, s_{K_0+1}, \cdots, s_{K_0+d-1}]^T, \quad (35)$$

$$\boldsymbol{s}(t_0 + T) = [s_{K_0+1}, s_{K_0+2}, \cdots, s_{K_0+d}]^T. \quad (36)$$

Equation (32) is a direct consequence of the above and the i.i.d. assumption on the transmitted symbols. Note also that the assumption that $E(s_k s_k^*) = 1$ can be achieved without loss of generality by scaling $h(\cdot)$.

In establishing channel identifiability and developing channel identification and equalization algorithms, an algebraic approach is used. To provide some intuition behind what follows, we note the special "forward shift" structure of the correlation matrices of the source. Due to this special structure, the rank of $\boldsymbol{R_x}(k)$ decreases as k increases. On the other hand, the range space of $\boldsymbol{R_x}(k)$ is spanned by columns of the channel parameter matrix $\boldsymbol{H}$. It is the change in the rank of $\boldsymbol{R_x}(k)$ that provides the information for the identification of the column vectors of $\boldsymbol{H}$, and consequently, the identification of the channel.

Theorem 1: Suppose that $\boldsymbol{H}$ and $\boldsymbol{s}(i)$ satisfy the linear model (30) and its constraints. Then $\boldsymbol{H}$ is uniquely determined up to a constant by $\boldsymbol{R_x}(0)$ and $\boldsymbol{R_x}(1)$.

Proof: Suppose that there are two sets of channels and sources satisfying the linear model and its constraints, i.e. both $\boldsymbol{H}$, $\boldsymbol{s}(i)$ and $\tilde{\boldsymbol{H}}$, $\tilde{\boldsymbol{s}}(i)$ satisfy (30) and its constraints. Then we have

$$\boldsymbol{H}\boldsymbol{H}^H = \tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^H, \quad (37)$$

$$\boldsymbol{H}\boldsymbol{J}\boldsymbol{H}^H = \tilde{\boldsymbol{H}}\boldsymbol{J}\tilde{\boldsymbol{H}}^H. \quad (38)$$

Equation (37) implies that [27]

$$\tilde{\boldsymbol{H}} = \boldsymbol{H}\boldsymbol{Q} \quad (39)$$

where $\boldsymbol{Q}$ is an orthogonal matrix. Substituting (39) into (38), we have

$$\boldsymbol{Q}\boldsymbol{J}\boldsymbol{Q}^H = \boldsymbol{J}. \quad (40)$$

Denote the ith column of $\boldsymbol{Q}$ as $\boldsymbol{q}_i$. Equation (40) gives, not surprisingly, a Jordan chain of length d associated with eigenvalue 0 of matrix $\boldsymbol{J}$

$$\begin{aligned} \boldsymbol{J}\boldsymbol{q}_1 &= \boldsymbol{q}_2, \\ &\vdots \\ \boldsymbol{J}\boldsymbol{q}_{n-1} &= \boldsymbol{q}_n, \\ \boldsymbol{J}\boldsymbol{q}_n &= \boldsymbol{0}. \end{aligned} \quad (41)$$

Since $\|q_n\|_2 = 1$, the last equation in the above implies $q_n = [0, \cdots, 0, e^{j\phi}]^T$ for some real ϕ, and consequently $Q = e^{j\phi}I$. We finally have, from (39), $H = \tilde{H}e^{j\phi}$. □

The following lemma provides an alternative constructive proof of the above theorem. Its significance is that it offers a computational method, in closed form, to identify H from $R_x(0)$ and $R_x(1)$.

Lemma 1: Let $R_x(0)$ have the following singular value decomposition,

$$U^H R_x(0) U = \text{diag}(\sigma_1^2, \cdots, \sigma_d^2, 0, \cdots, 0). \quad (42)$$

Let u_i denote the ith column of U, and let

$$U_s = [u_1, \cdots, u_d], \quad (43)$$

$$\Sigma = \text{diag}(\sigma_1, \cdots, \sigma_d), \quad (44)$$

$$F = \Sigma^{-1} U_s^H. \quad (45)$$

Suppose

$$R = F R_x(1) F^H \quad (46)$$

has a singular value decomposition of the form

$$[y_1, \cdots, y_d]^H R [z_1, \cdots, z_d] = \text{diag}(\gamma_1^2, \cdots, \gamma_d^2). \quad (47)$$

Then there exists a real phase ϕ such that

$$H = U_s \Sigma Q e^{j\phi} \quad (48)$$

where

$$Q = [y_d, R y_d, \cdots, R^{(d-1)} y_d], \quad (49)$$

or equivalently,

$$Q = [(R^\dagger)^{(d-1)} z_d, (R^\dagger)^{(d-2)} z_d, \cdots, z_d]. \quad (50)$$

Proof: H satisfies the constraint imposed by $R_x(0)$

$$HH^H = R_x(0). \quad (51)$$

From (42), we have

$$H = U_s \Sigma V \quad (52)$$

where $V = [v_1, \cdots, v_d]$ is an orthogonal matrix. Thus,

$$FH = \Sigma^{-1} U_s^H U_s \Sigma V = V. \quad (53)$$

Since $R_x(1) = HJH^H$, one obtains

$$R = F R_x(1) F^H = FHJH^H F^H = VJV^H. \quad (54)$$

Keeping in mind that V is an orthogonal matrix, the right side of the above equation is a Jordan decomposition of R, and we have the familiar Jordan chain

$$R v_k = v_{k+1}, \qquad k = 1, \cdots, d-1, \quad (55)$$

$$R v_d = 0. \quad (56)$$

Unfortunately, it is not computationally reliable to obtain V from the Jordan decomposition of R. This difficulty is alleviated by the fact that v_d is also a singular vector of R. Compute $R^H R$ and we have

$$R^H R = V \text{diag}(1, \cdots, 1, 0) V^H. \quad (57)$$

It is clear from the above that i) the matrix R has one and only one singular value equal to 0; ii) v_d is a right singular vector of R associated with the zero singular value. Now if R has an SVD as in (47), i.e., if

$$[y_1, \cdots, y_d]^H R [z_1, \cdots, z_d] = \text{diag}(\gamma_1^2, \cdots, \gamma_d^2) \quad (58)$$

then there is a ϕ such that

$$v_d = z_d e^{j\phi}. \quad (59)$$

Now the problem is to solve (55)–(56) for v_i given v_d as in (59). Consider the equation involving v_d and v_{d-1}

$$R v_{d-1} = v_d. \quad (60)$$

Although R is singular (rank $d-1$), two observations imply that v_{d-1} is *uniquely* determined from the above equation. First,

$$\|R\|_2 = 1, \quad (61)$$

which is evident from (57). Second, V is orthogonal, hence

$$\|v_d\|_2 = \|v_{d-1}\|_2 = 1. \quad (62)$$

These two observations lead to

$$v_{d-1} = R^\dagger v_d \quad (63)$$
$$= R^\dagger z_d e^{j\phi}. \quad (64)$$

This leads to (50). The derivation of (49) is similar. Compute RR^H and we have

$$RR^H = V \text{diag}(0, 1, \cdots, 1) V^H. \quad (65)$$

It follows that v_1 is a left singular vector of R associated with the zero singular value and (49) is then evident. □

B. Algorithm Implementation with Noisy Data

Lemma 1 provides the essential parts of the proposed blind channel identification algorithm. The previous development was based on a noise-free model. However it is not hard to think of ways of handling additive white noise. In particular, we follow ideas now widely used in sensor array processing, see, e.g., [26]. Thus, we consider the vectorized process $x(i)$ satisfying

$$x(i) = Hs(i) + n(i). \quad (66)$$

The correlation matrix $R_x(k)$ satisfies

$$R_x(k) = HR_s(k)H^H + R_n(k). \quad (67)$$

Under the assumption of white noise, the noise correlation matrix $R_n(k)$ has the form

$$R_n(k) = E(n(i)n^H(i-k)) \quad (68)$$
$$= \sigma^2 J^{kT_s} \quad (69)$$

where σ^2 is the *unknown* noise variance, $\boldsymbol{J}$ is the forward shift matrix given by (34), and T_s is an integer such that $T = T_s\Delta$.

Although neither the noise covariance, nor the signal space dimension d is known *a priori,* they can be obtained from the data covariance matrix $\boldsymbol{R_x}(0)$. It can be shown that the SVD of $\boldsymbol{R_x}(0)$ must have the following form:

$$\boldsymbol{U}^H\boldsymbol{R_x}(0)\boldsymbol{U} = \text{diag}\,(\lambda_1+\sigma^2, \cdots, \lambda_d+\sigma^2, \sigma^2, \cdots, \sigma^2) \quad (70)$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d > 0$. Therefore, both σ^2 and d can in theory be obtained by determining the most significant singular values of $\boldsymbol{R_x}(0)$. In practice, a threshold test can be employed to determine d, and then to estimate σ^2 from the singular values of the estimated data covariance matrix. Readers are referred to [16] and [28] for suitable methods.

Once the noise covariance σ^2 is determined, the identification procedure suggested in Lemma 1 can be easily extended to handle noisy data: subtract the corresponding noise correlation matrices from the observation correlation matrices.

We first outline the algorithm and then address certain technical points of the algorithm.

A Blind Channel Identification and Equalization Algorithm:

1) Select t_0, Δ such that $T = T_s\Delta$, $L = m\Delta$, and form the vectorized observation $\boldsymbol{x}(i) = [x(t_0+\Delta+iT), \cdots, x(t_0+m\Delta+iT)]^T$, $i = 0, 1, \cdots,$
2) Estimate $\boldsymbol{R_x}(0)$ and $\boldsymbol{R_x}(1)$ from $\boldsymbol{x}(i)$ via, for example, time averaging,

$$\hat{\boldsymbol{R}}_{\boldsymbol{x}}(0) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}(i)\boldsymbol{x}^H(i), \quad (71)$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{x}}(1) = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}(i)\boldsymbol{x}^H(i-1). \quad (72)$$

3) From $\hat{\boldsymbol{R}}_{\boldsymbol{x}}(0)$, estimate the noise covariance σ and the dimension d of the signal space.
4) Compute the SVD of $\boldsymbol{R}_0$,

$$\boldsymbol{R}_0 = \hat{\boldsymbol{R}}_{\boldsymbol{x}}(0) - \hat{\sigma}^2\boldsymbol{I} \quad (73)$$

and form $\boldsymbol{U_s}$ which consists of the singular vectors associated with the d largest singular values, $\boldsymbol{\Sigma}$ which consists of the positive square-root of the d largest singular values, and then $\boldsymbol{F} = \boldsymbol{\Sigma}^{-1}\boldsymbol{U_s}^H$.
5) Compute the SVD of $\boldsymbol{R}$,

$$\boldsymbol{R} \doteq \boldsymbol{F}(\hat{\boldsymbol{R}}_{\boldsymbol{x}}(1) - \boldsymbol{R_n}(1))\boldsymbol{F}^H \quad (74)$$

where

$$\boldsymbol{R_n}(1) = \hat{\sigma}^2\boldsymbol{J}^{T_s}, \quad (75)$$

$$\boldsymbol{J} = \begin{pmatrix} 0 & 0\ldots & 0 & 0 \\ 1 & 0\ldots & 0 & 0 \\ 0 & 1\ldots & 0 & 0 \\ v\vdots & \ddots & \vdots & \vdots \\ 0 & 0\cdots & 1 & 0 \end{pmatrix}. \quad (76)$$

Let $\boldsymbol{y}_d$ and $\boldsymbol{z}_d$ denote the left and right singular vectors corresponding to the smallest singular value.

6) Form an estimate of $\boldsymbol{H}$ (and consequently $h(\cdot)$ if necessary) as

$$\hat{\boldsymbol{H}} = \boldsymbol{U_s\Sigma Q} \quad (77)$$

where

$$\boldsymbol{Q} = [\boldsymbol{y}_d, \boldsymbol{R}\boldsymbol{y}_d, \cdots, \boldsymbol{R}^{(d-1)}\boldsymbol{y}_d] \quad (78)$$

or

$$\boldsymbol{Q} = [(\boldsymbol{R}^\dagger)^{(d-1)}\boldsymbol{z}_d, (\boldsymbol{R}^\dagger)^{(d-2)}\boldsymbol{z}_d, \cdots, \boldsymbol{z}_d] \quad (79)$$

or a certain combination of the above.
7) Extract the information symbols

$$\hat{\boldsymbol{s}}(i) = \hat{\boldsymbol{H}}^\dagger\boldsymbol{x}(i) \quad (80)$$
$$= \boldsymbol{Q}^H\boldsymbol{F}\boldsymbol{x}(i) \quad (81)$$

or by implementing various types of equalizer or maximum-likelihood estimation schemes based on the estimated channel.

Discussion of Implementation Issues:

1) *Timing recovery and the selection of t_0 and L:* Uncertainties associated with timing recovery may cause significant transmission error in baud-rate equalizers. One advantage of the proposed approach is that such uncertainties do not affect the proposed estimation scheme. As one can see from previous sections, the selection of t_0 and L is arbitrary as long as the resulting channel parameter matrix is of full column rank. Different t_0, sampling period Δ, and L will result in a different channel parameter matrix $\boldsymbol{H}_{t_0,\Delta,L}$. Nonetheless, once t_0, Δ, and L are *fixed* and the resulting $\boldsymbol{H}_{t_0,\Delta,L}$ is of full column rank, the construction of observation vector sequence $\boldsymbol{x}_{t_0,\Delta,L}(n)$ is *fixed* and the source vector $\boldsymbol{s}_{t_0,\Delta,L}(n)$ is uniquely specified. Although we do not know the timing, i.e., the exact time instant when a symbol is transmitted from the source, we can still obtain the source vector $\boldsymbol{s}_{t_0,\Delta,L}(n)$ from the observation vector $\boldsymbol{x}_{t_0,\Delta,L}(n)$. It is important to note that the source vector sequence is constructed by adding one *new* symbol and removing the *oldest* symbol in a first-in first-out (FIFO) fashion. Consequently, by observing a fixed component of $\boldsymbol{s}_{t_0,\Delta,L}(n)$, one obtains the whole sequence of source symbols.

There are subtleties in choosing t_0 and L as far as the implementation is concerned. The choice of t_0, Δ, and L affects the signal space dimension d, as shown in (16), and also the condition number of the matrix $\boldsymbol{H}(t_0)$. The effects of these selections on the algorithm performance require further investigation. In our simulation study, the observation window length L is chosen as multiples of T. Various t_0's were chosen in the simulation study and the differences seem to be negligible.
2) *The selection of sampling period Δ:* The sampling period Δ needs to be an integer fraction of the symbol interval T. In addition, to ensure the full rank of the channel parameter matrix, the sampling rate has to be higher than the baud rate. For any fixed t_0 and L, the signal

space dimension d, independent of the sampling, is determined. The vectorized observation $\boldsymbol{x}(i)$ needs to have a dimension higher than d. (The dimension of the vector $\boldsymbol{x}(i)$ needs to be at least $d+1$ if the noise covariance is unknown.) In practice, the full rank condition can be ensured if the sampling frequency f_s satisfies

$$f_s > \frac{d+1}{L}. \tag{82}$$

For example, suppose that $L = T$, and there are d_0 symbols having contributions to the received signal in the observation window. Then the sampling frequency needs to be at least $(d_0+1)/T$. If the length of the observation window is increased to $L = 2T$, then there must be d_0+1 symbols having contributions to the received signals in the observation window. The sampling frequency can be decreased to $(d_0+2)/2T$. In general, if $L = kT$, the sampling frequency should be at least $(d_0+k)/k$ times faster than the baud rate. Hence the burden of oversampling is not very significant.

3) *Channel estimation:* Theoretically, either (78) or (79) can be used to obtain an estimate of $\hat{\boldsymbol{H}}$. Practically, it is better to use a combination of the two estimates. Since $\boldsymbol{R}_{\boldsymbol{x}}(0)$ and $\boldsymbol{R}_{\boldsymbol{x}}(1)$ can only be estimated, estimation error is inevitable. This error is likely to be propagated in forming the columns of $\hat{\boldsymbol{H}}$. The error propagation can be somewhat limited if the first $d/2$ columns of $\hat{\boldsymbol{H}}$ are constructed from (78) and the last $d/2$ columns of $\hat{\boldsymbol{H}}$ are constructed from (79).

4) *Channel equalization, least squares and minimum variance estimation:* Various equalization techniques can be employed once the channel estimation is available. Difficulties arise when the channel has *nonminimum* phase. In such a case, some channel inversion algorithms may not be stable. Interestingly, this is not so for the new algorithm presented here. The source symbols can be extracted by using the *causal* FIR filter obtained by taking the inverse of the matrix $\hat{\boldsymbol{H}}$. Moreover, the source estimate (81) can be achieved by the least squares method for computational simplicity (no additional matrix inversion is necessary). In principle, the so-called total least squares approach [13] is probably more justified because of the estimation error in $\hat{\boldsymbol{H}}$; it can also be implemented with some added computational burden. Slightly different from the least squares solution is the minimum-variance estimate given by

$$\hat{\boldsymbol{s}}(i) = \hat{\boldsymbol{H}}^H(\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^H + \hat{\sigma}^2\boldsymbol{I})^{-1}\boldsymbol{x}(i). \tag{83}$$

5) *Fast signal subspace decomposition:* The major computational cost in the proposed algorithm comes from the eigendecomposition of the correlation matrices. The recent development of fast signal subspace decomposition (FSD) techniques can provide computationally efficient, easily parallelizable methods for subspace determination of correlation matrices [29]. In our case, if m samples are used in an observation interval, and if the dimension of the signal space is d, the computational cost is of the order $O(m^2d)$ as opposed to $O(m^3)$ for standard eigendecomposition techniques.

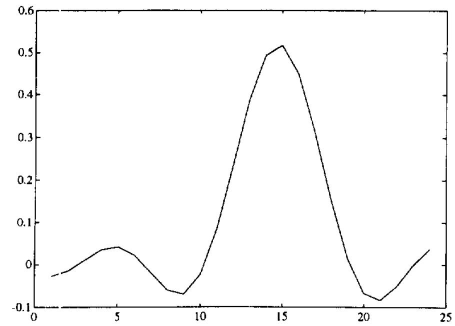

Fig. 3. A three-ray multipath channel impulse response.

TABLE I
CHANNEL IMPULSE RESPONSE

n	1	2	3	4	5	6
$h(n)$	−0.02788	−0.01556	0.009773	0.0343	0.04142	0.0216
n	7	8	9	10	11	12
$h(n)$	−0.01859	−0.06035	−0.07025	−0.0241	0.08427	0.2351
n	13	14	15	16	17	18
$h(n)$	0.3874	0.4931	0.5167	0.4494	0.3132	0.152
n	19	20	21	22	23	24
$h(n)$	0.01383	−0.06754	−0.08374	−0.05137	−0.001258	0.03679

IV. A SIMULATION EXAMPLE

In this simulation, as an approximation of a two-ray multipath environment, the channel[1] was generated from two delayed pulses and is given by Table I and shown in Fig. 3. Among 23 zeros of the system, there are 21 nonminimum phase zeros. The source symbols were drawn from a 16 QAM signal constellation with a uniform distribution.

The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = 20\log\frac{\|r(\cdot)\|_2}{\|n(\cdot)\|_2}\ (\text{dB}). \tag{85}$$

Fig. 4 is a plot of 1000 output symbols of the unequalized channel (obtained by sampling the received signal at kT, $k = 1, 2, \cdots$). The signal-to-noise ratio was 30 dB. Clearly, the intersymbol interference is severe and a high error rate is expected.

In the implementation of our algorithm, we chose $t_0 = 0$, and an observation interval of length $L = 4T$. A simple calculation according to (16) gave $d = 10$. The sampling frequency was chosen to be 4 times faster than the baud rate, i.e., $T = 4\Delta (T_s = 4)$. The vector representation of the received signal $\boldsymbol{x}(i)$ was a 20×1 complex vector. In

[1] The impulse response is obtained from delayed raised cosine pulses. A single pulse is described by $c(t, \alpha)$ where α is the roll-off factor.

$$h(t) = (0.2c(t, 0.11) + 0.4c(t-2.5, 0.11))W_{6T}(t) \tag{84}$$

where W_{6T} is a square window of duration 6 symbol intervals, i.e., $L_h = 6T$.

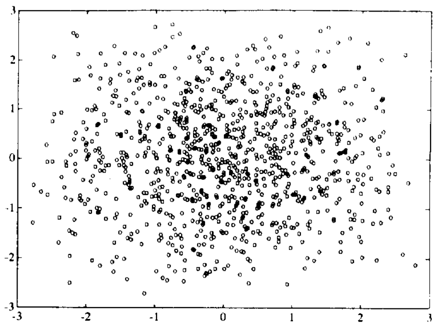

Fig. 4. The output of the unequalized channel; SNR=30 dB.

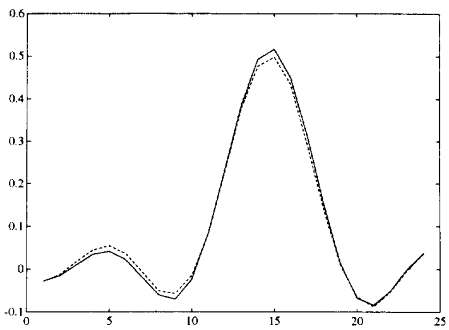

Fig. 5. The actual channel (solid) and the estimated channel (dashed). One hundred symbols were used for estimation. SNR=30 dB.

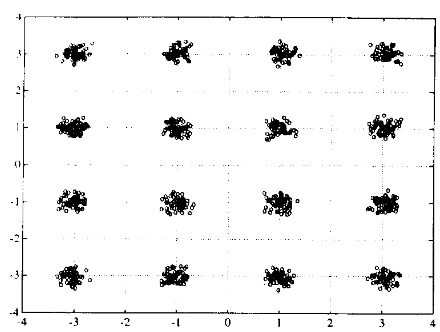

Fig. 6. The output of the equalized channel. 1000 symbols were plotted. One hundred symbols were used for estimation. SNR=30 dB.

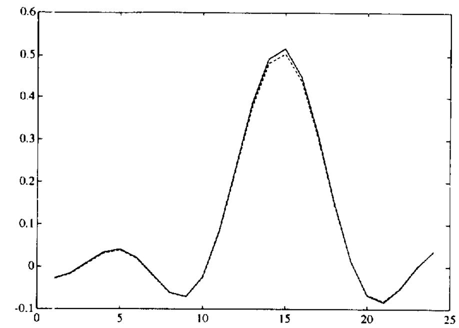

Fig. 7. The actual channel (solid) and the sample mean of 100 estimates (dashed). One hundred symbols were used for each estimate. SNR=30 dB.

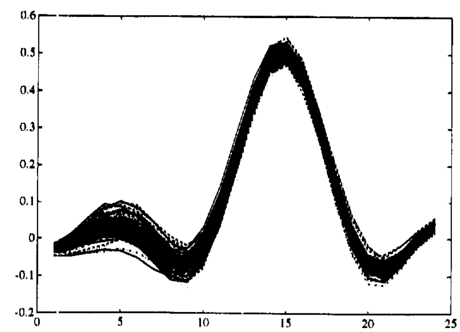

Fig. 8. 100 estimates of the channel. One hundred symbols were used for each estimate. SNR=30 dB.

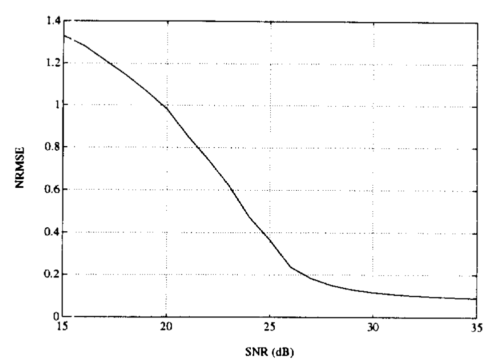

Fig. 9. NRMSE versus SNR. 100 Monte Carlo runs and 100 symbols were used in each run.

estimating the channel parameter matrix, 100 symbols were used to estimate $\boldsymbol{R_x}(0)$ and $\boldsymbol{R_x}(1)$. Fig. 5 shows the actual channel and its estimate based on a single realization. 1000 symbols were then transmitted and the equalized channel output is shown in Fig. 6, which indicates that the channel is well equalized.

A Monte Carlo simulation of 100 independent trials was conducted under the same simulation scenario. Fig. 7 shows the sample mean of 100 estimates of the channel, and also the actual channel. Fig. 8 shows the 100 estimates of the channel.

To obtain a performance measure of the channel estimation, the normalized root-mean-square error (NRMSE) of the estimator is defined by

$$\text{NRMSE} = \frac{1}{||h||}\sqrt{\frac{1}{M}\sum_{i=1}^{M}||\hat{h}_{(i)} - h||^2} \tag{86}$$

where M is the number of Monte Carlo trials (100 in our case), and $\hat{h}_{(i)}$ is the estimate of the channel from the ith trial. Fig. 9 shows the NRMSE versus SNR in a series of 100 Monte Carlo runs for different SNR's.

The bit error rate (BER) was also tested against the SNR. In this case a BPSK source was used to estimate the channel.

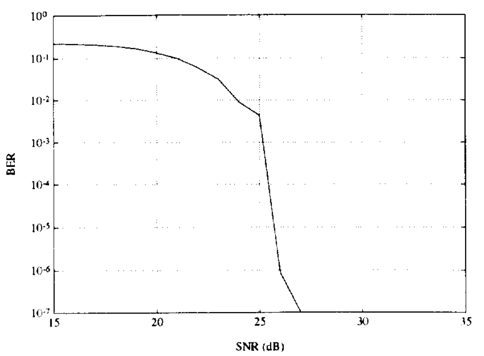

Fig. 10. Bit-error-rate versus SNR. 100 Monte Carlo runs. One hundred symbols were used for the channel estimate.

At each Monte Carlo run, the channel was estimated and a minimum-variance equalizer given by (83) was implemented and the probability of error was evaluated. The BER is defined here as the probability of error averaged over 100 Monte Carlo runs. Fig. 10 shows the effect of noise on the BER. While the performance is excellent at high SNR's, the error rate is high (above 10^{-3}) when SNR is below 25 dB. The achieve better performance, a larger sample size will be necessary.

V. Conclusion

Blind equalization is of significant value in many communication problems. A new approach to blind identification and equalization is developed in this paper. By exploiting the cyclostationarity of the received signal via oversampling, we are able to identify possibly nonminimum-phase channels using only second-order statistics. This leads to more accurate estimation with a smaller sample size than methods using higher-order statistics. In addition, it can be easily incorporated into various existing equalization methods.

Although the proposed method provides asymptotically exact channel identification, the performance of the algorithm when a small number of symbols is used needs to be further examined from both theoretical and experimental points of view. In our simulation, the proposed algorithm performs well when SNR is high. However, there seems to be a "breakdown" point when the SNR is below a threshold. For the proposed algorithm to be effective at low SNR, a larger number of symbols is necessary, which limits the effectiveness of the algorithm for rapidly varying channels.

Acknowledgment

The authors would like to thank the referees, Prof. W. Gardner, and Associate Editor J. Gibson for their constructive comments and careful reviews.

References

[1] A. Benveniste and M. Goursat, "Blind equalizers," *IEEE Trans. Commun.*, vol. COM-32, pp. 871–883, Aug. 1984.

[2] L. R. Cerrato and B. A. Eisenstein, "Deconvolution of cyclostationary signals," *IEEE Trans. Acoust., Speech, and Signal Processing,* vol. ASSP-25, pp. 466–476, Dec. 1977.

[3] 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.

[4] E. Eleftheriou and D. D. Falconer, "Adaptive equalization techniques for HF channels," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 238–247, Feb. 1987.

[5] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Inform. Theory,* vol. IT-18, pp. 363–378, May 1972.

[6] G. J. Foschini, "Equalizing without altering or detecting data," *Bell Syst. Tech. J.*, vol. 64, pp. 1885–1911, Oct. 1985.

[7] W. Gardner, "A new method of channel identification," *IEEE Trans. Commun.*, vol. COM-39, pp. 813–817, June 1991.

[8] ——, "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Signal Processing Mag.*, vol. 8, pp. 14–36, Apr. 1991.

[9] ——, *Introduction to Random Processes with Applications to Signals and Systems.* New York: Macmillan, 1985; (2nd ed: New York, McGraw-Hill, 1990).

[10] ——, *Statistical Spectral Analysis: A Nonprobabilistic Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[11] W. A. Gardner and W. A. Brown, "Frequency-shift filtering theory for adaptive co-channel interference removal," in *Proc. 23rd Asilomar Conf. Signals, Syst., and Comput.*, Pacific Grove, CA, Oct. 1989, pp. 562–567.

[12] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867–1875, Nov. 1980.

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

[14] D. Hatzinakos and C. Nikias, "Estimation of multipath channel response in frequency selective channels," *IEEE J. Select. Areas Commun.*, vol. SAC-7, pp. 12–19, Jan. 1989.

[15] C. R. Johnson, "Admissibility in blind adaptive channel equalization," *IEEE Contr. Syst. Mag.*, pp. 3–15, Jan. 1991.

[16] K. Konstantinides and K. Yao, "Statistical analysis of effective singular values in matrix rank determination," *IEEE Trans. Acoust., Speech, and Signal Processing,* vol. 36, pp. 757–763, May 1988.

[17] J. Kormylo and J. M. Mendel, "Identifiability of nonminimum phase linear stochastic systems," *IEEE Trans. Automat. Contr.*, vol. AC-28, pp. 1081–1090, Dec. 1983.

[18] A. Papoulis, "Identification of systems driven by nonstationary noise," *IEEE Trans. Inform. Theory,* vol. IT-24, pp. 240–244, Mar. 1978.

[19] G. Picci and G. Prati, "Blind equalization and carrier recovery using a stop-and-go decision directed algorithm," *IEEE Trans. Commun.*, vol. COM-35, pp. 877–887, Sept. 1987.

[20] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation," *IEEE Trans. Commun.*, vol. COM-23, pp. 679–682, June 1975.

[21] 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.

[22] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind identification and equalization of multipath channels: A frequency domain approach," *IEEE Trans. Inform. Theory,* to be published.

[23] L. Tong, G. Xu, and T. Kailath, "Necessary and sufficient conditions for channel identification based on second-order statistics," in *Proc. 1993 Int. Symp. Inform. Theory,* San Antonio, TX, Jan. 1993.

[24] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, and Signal Processing,* vol. ASSP-31, pp. 459–472, Apr. 1983.

[25] G. Ungerboeck, "Fractional tap-spacing equalizer and consequences for clock recovery in data modems," *IEEE Trans. Commun.*, vol. COM-24, pp. 856–864, Aug. 1976.

[26] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. Signal Processing,* vol. 39, pp. 12436–12449, Nov. 1991.

[27] B. Vinograde, "Canonical positive definite matrices under internal linear transforms," *Proc. Amer. Math. Soc.*, vol. 1, pp. 159–161, 1950.

[28] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, and Signal Processing,* vol. ASSP-33, pp. 387–392, Apr. 1985.

[29] G. Xu and T. Kailath, "Fast subspace decomposition," *IEEE Trans. Signal Processing,* Mar. 1994.