A Blind Fractionally Spaced Equalizer Using Higher Order Statistics

Jie Zhu, *Member, IEEE*, Xi-Ren Cao, *Fellow, IEEE*, and Ruey-wen Liu, *Life Fellow, IEEE*

***Abstract*—In this paper, we introduce a new blind fractionally spaced equalizer based on the fourth-order statistics of the input symbol sequence. The input symbol sequence is assumed to come from an independent identically distributed finite alphabet with nonzero fourth-order cumulants. We formulate the equalizer as a column vector and compute it by simultaneously diagonalizing a set of matrices obtained from the fourth-order cross-cumulants of the input and output of the equalizer. Simulation results show that this equalizer works well with a short symbol sequence, even if the channel time span is not accurately estimated.**

I. Introduction

A HIGH-SPEED digital communication system is always subject to intersymbol interference (ISI) caused by channel amplitude and phase distortions. In order to improve the transmission performance, it is important for the receiver to remove ISI through equalization technology. Traditionally, the design of an equalizer is achieved either by sending a known sequence (training sequence) or by using *a priori* knowledge of the channel. Under most communication environments, little *a priori* channel knowledge is available, and the training sequence therefore plays a key role in channel equalization. With the received signal as its input, the equalizer adapts its parameters by comparing its current output with the desired training sequence. When the channel varies rapidly, the training sequence has to be applied frequently, which results in loss of communication efficiency. Furthermore, such equalization techniques may not work when the receiver has no access to the training sequence. The goal of blind equalization is to recover the original sequence from the received signal that is corrupted by noise and ISI, without the help of a training sequence and *a priori* knowledge of the channel.

It has been clear that almost all man-made communication signals exhibit a statistical property called *cyclostationarity*. Gardner [2] showed that the second-order statistics (SOS) of a cyclostationary signal contains the phase information of the channel it goes through, and this phase information can be used to identify the channel, which is possibly a nonminimum phase channel, up to a complex constant. Since Tong *et al.* [3], [4] proposed the first blind equalizer by using the phase information contained in the SOS of the oversampled output sequence, the blind fractionally spaced equalizer, based on the cyclostationarity and data structures involved in the oversampled received sequence, has attracted considerable research attention [5]–[12]. Although these methods can obtain an acceptable equalization performance within 100 or more symbols, they are generally sensitive to the error in channel-order estimation. In these algorithms, it is generally assumed that the channel order is either known or can be estimated by other algorithms. As we know, when the signal-to-noise ratio (SNR) becomes smaller, to determine a correct channel order from the channel output is a difficult task. In order to weaken the dependence on channel-order estimate, Slock [13], [14] proposed a linear prediction algorithm (LPA), which is still based on the SOS of channel outputs, and Gesbert *et al.* [15] and Abed-Meraim *et al.* [16], [17] later gave a detail study of this algorithm. Comparing with the aforementioned algorithms, LPA is more robust to the estimate error of the channel order. However, LPA requires a large leading coefficient of the channel transfer function. When the leading coefficient is small, which is very common due to a limited bandwidth, LPA performs a poor channel identification. Recently, Ding [18] extended LPA by using the full outer-product decomposition of the channel parameter vector. This LPA generalization is also robust to the over-modeling error of channels and improves the performance of channel identification. The most popular shortcoming of SOS-based methods is that they cannot be applied for the co-channel interference (CCI) cancelation in multi-user systems.

In comparison with SOS, the higher order statistics (HOS) of a signal offers an appealing benefit: insensitivity to an additive Gaussian noise. This benefit is very useful in communication systems because most noises in communication system can be described approximately by Gaussian distribution. However, in order to exploit its HOS, a non-Gaussian symbol sequence with independent and identically distributed (i.i.d.) functions is commonly assumed. Although the i.i.d. condition is stricter than the *cyclostationarity* used in SOS-based methods, two facts render it applicable: 1) the real input symbol sequence tends to be i.i.d. and 2) if the HOS is utilized only up to the fourth order, then a qualified input symbol sequence is only required to satisfy i.i.d. condition up to the fourth order (i.e., not to *infinity*.) Many blind channel identification and equalization methods based on HOS have been developed

Manuscript received May 19, 1997; revised December 7, 1998. This work was supported in part by RGC under Grant HKUST 690/95E. This paper was recommended by Associate Editors W. Siu and J. Choma.

J. Zhu is with the Center for Signal Processing, Nanyang Technological University, Singapore 639798.

X.-R. Cao is with the Department of Electrical & Electronic Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.

R. Liu is with the Department of Electrical and Computer Engineering, University of Notre Dame, Notre Dame, IN 46556 USA.

Publisher Item Identifier S 1057-7130(99)04894-6.

[19]–[28]. Most of them can be used for CCI cancelation in multiuser systems. In these methods, a class of criteria which is sometimes called the *Wiggine-Donoho-Shalvi-Weinstain* class is widely used, and a gradient-based algorithm is usually adopted to optimize these criteria, so as to estimate the channel impulse response. However, as we know, the steepest gradient is very sensitive to modeling errors; when the symbol size is not large enough, the steepest gradient cannot be estimated within a reasonable error range. Consequently, these HOS-based blind identification and equalization techniques suffer from the weakness that hundreds and typically thousands of received data are processed before the performance of the equalizer reaches an acceptable level. In mobile digital communication, this weakness becomes so severe that such techniques may not work due to the rapid variation of the channel.

In this paper, we introduce another blind fractionally spaced equalization scheme that provides an acceptable estimate with a short input symbol sequence and is resistant to the noise and error in the channel-order estimate. We first convert a blind equalization problem to a blind source separation problem by oversampling the received baseband signal, then propose a criterion based on the fourth-order cross-cumulants of the input and output of the equalizer to be designed. We prove that the maximum of this criterion results in a reliable equalizer. To solve the optimization problem, a *Jacobi* algorithm for a set of matrices proposed by Cardoso [33] is applied. The contribution of this paper is to propose an HOS-based equalizer which is robust to the limited received samples and numerical errors. Throughout this paper, we make the following assumptions on the channel, the input symbol sequence, and the noise.

1) The channel can be approximately modeled as a time-invariant linear FIR filter with finite order.

2) The input symbols are in a finite alphabet, and are drawn from a set of zero-mean unit-variance i.i.d. random variables with a nonzero fourth-order cumulant.

3) The noise is white with a Gaussian distribution and is independent of the symbols.

The paper is organized as follows. In Section II, we formulate the problem by oversampling the received signal; the blind equalization algorithm is developed in Section III; several simulation experiments, which illustrate the performance of the equalizer, are described in Section IV. Finally, the conclusion is presented in Section V.

II. Problem Formulation

Consider a single-user digital communication system with baud period T. Let $\{\alpha_i, i = 0, 1, 2, \cdots\}$ be its input symbol sequence. Denoting the received signal (baseband) as a continuous function $y(t)$ of time t, we have

$$y(t) = \sum_{i=-\infty}^{+\infty} \alpha_i h(t - iT) + w(t) \tag{1}$$

where $h(t)$ is the channel transfer function and $w(t)$ the additive noise. The received signal is then sampled as a discrete sequence $\{y_k\}$. Let the sampling frequency be N/T, $N > 1$. Taking into account that the channel has a finite time span MT, we then have

$$y_{k+nN} = \sum_{i=0}^{M-1} \alpha_{n-i} h\left(\left(\frac{k}{N} + i\right)T\right) + w_{k+nN} \tag{2}$$

where w_k is the kth discrete sample of the noise $w(t)$. Denote superscript ^{T} as the *transpose* operator. Let

$$\begin{aligned}\mathbf{o}(n) &= \left[y_{nN}, \cdots, y_{N-1+nN}, \cdots, y_{(n+\lambda)N-1}\right]^T\\ \mathbf{s}(n) &= \left[\alpha_{n-M+1}, \cdots, \alpha_n, \cdots, \alpha_{n+\lambda}\right]^T\\ \mathbf{w}(n) &= \left[w_{nN}, \cdots, w_{N-1+nN}, \cdots, w_{(n+\lambda)N-1}\right]^T\end{aligned}$$

where λ is the smallest integer that is not less than $(M-1)/(N-1)$. Equation (2) can then be represented in a matrix form as

$$\mathbf{o}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{w}(n) \tag{3}$$

where $\mathbf{A}$ is called the *channel convolution matrix*. Let $N' = \lambda N$, $M' = M + \lambda - 1$. $\mathbf{A}$ is an $N' \times M'$ matrix and is defined by

$$\mathbf{A} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_M & 0 & \cdots & 0\\ 0 & \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_M & \cdots & 0\\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots\\ 0 & 0 & \cdots & \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_M \end{bmatrix}$$

where $\mathbf{h}_j = [\tilde{a}_{1j}, \tilde{a}_{2j}, \cdots, \tilde{a}_{Nj}]^T$ for $j = 1, 2, \cdots, M$ and $\tilde{a}_{ij}$ is given by

$$\tilde{a}_{i,j} = h\left(\left(\frac{i-1}{N} - j + M\right)T\right). \tag{4}$$

Observe that the vector sequence $\mathbf{s}(n)$ for $n = 1, 2, \cdots$, has a particular structure: each of its component sequences is the same as the input symbol sequence except for their time shift. Under the i.i.d. assumption of the input symbol sequence, it is easy to see that the components of $\mathbf{s}(n)$ for any fixed n are mutually independent. Therefore, the problem of recovering $\mathbf{s}$ and $\mathbf{A}$ by $\mathbf{o}$ without any knowledge about $\mathbf{A}$ and $\mathbf{s}$ belongs to the category of blind source separation problems [1], [29], [30], [31]. Each component of $\mathbf{s}$ can be viewed as a "source" and that of $\mathbf{o}$ as an observation. Many approaches exist to solve this problem when $\mathbf{A}$ is a full-column rank matrix. Note that for a full-column rank matrix $\mathbf{A}$, a necessary condition is $N' \geq M'$. When $N > 1$, there always exists a positive integer λ, so that $N' \geq M'$. In what follows, we always assume that $N' \geq M'$ and $\mathbf{A} = [a_{i,j}]$ is an $N' \times M'$ full-column rank matrix.

III. The Algorithm

In Section II, we formulated the blind equalization as a blind source separation problem. Principally, if the symbol sequence is not Gaussian distributed, we can employ any suitable blind source separation algorithm to extract the "source signals," i.e., the vector sequence $\mathbf{s}(n)$. We note, however, that all the source sequences in this blind equalization problem are identical to the input symbol sequence except for a different time delay. This special feature may simplify the algorithm. In fact, we only need to separate out one of the source signals. In this section, we will develop a procedure that extracts such a source sequence. For notational convenience, we drop the index n in (3), and thereby have

$$\mathbf{o} = \mathbf{A}\mathbf{s} + \mathbf{w}. \tag{5}$$

Denote superscript † as the *Hermitian transpose* operator. Suppose there exists a column vector $\mathbf{b} = [b_1, b_2, \cdots, b_{N'}]^T$, by which we have an estimate

$$e = \mathbf{b}^\dagger \mathbf{o} = \mathbf{b}^\dagger(\mathbf{A}\mathbf{s} + \mathbf{w}) = \mathbf{d}^\dagger \mathbf{s} + \mathbf{b}^\dagger \mathbf{w} \tag{6}$$

where $\mathbf{d}^\dagger = \mathbf{b}^\dagger \mathbf{A} = [d_1, d_2, \cdots, d_{M'}]$. If $\mathbf{b}$ is orthogonal to all columns of matrix $\mathbf{A}$ but one, i.e., $\mathbf{d}$ has only one nonzero component, then the estimated sequence of e is an unbiased estimate of the input symbol sequence regardless of the scale. Such a vector $\mathbf{b}$ indeed exists and is not unique based on the assumption that $\mathbf{A}$ is a full-column rank matrix.

In order to determine $\mathbf{b}$, the algorithm in [22] by maximizing the fourth-order cumulant of the estimated output e can be applied. However, experiments show that this algorithm needs a large number of received symbols before it provides a reliable solution. To obtain an equalizer which works well with a short symbol sequence, in the following we propose a similar but different algorithm; the equalizer can be determined by simultaneously diagonalising a set of matrices.

A. A Criterion Based on Fourth-Order Cumulant

Recall that a fourth-order cross-cumulant of two random variables x, y is defined as

$$C_{xy} = E\{xx^*yy^*\} - E\{xx^*\}E\{yy^*\} - E\{xy^*\}E\{x^*y\} - E\{xy\}E\{x^*y^*\}$$

where superscript * is the *complex conjugate* operator and $E\{x\}$ stands for the expectation of x. We employ the fourth-order cross-cumulant of the input and output of the equalizer as a criterion to design the equalizer. Let o_k be the kth component of $\mathbf{o}$ in (6). We define

$$F_k(\mathbf{b}) = E\{ee^*o_k o_k^*\} - E\{ee^*\}E\{o_k o_k^*\} - E\{eo_k^*\}E\{e^*o_k\} - E\{eo_k\}E\{e^*o_k^*\} \tag{7}$$

for any k, $k = 1, 2, \cdots, N'$. Substituting (5) and (6) into (7) gives

$$F_k(\mathbf{b}) = F_k(\mathbf{d}) = c_s \sum_{i=1}^{M'} |d_i|^2 |a_{ki}|^2 \tag{8}$$

where c_s is the fourth-order cumulant of the input symbol sequence, and $c_s \neq 0$ by assumption **A2**. Obviously, for a fixed k, $k = 1, 2, \cdots, N'$, if there is only one component in a_{ki}, $i = 1, 2, \cdots, M'$ with the maximum absolute value, then the solution to

$$\begin{cases} \text{maximize:} & |F_k(\mathbf{d})| \\ \text{subject to:} & \mathbf{d}^\dagger \mathbf{d} = 1 \end{cases} \tag{9}$$

is a unit vector $\mathbf{d}$, whose components are all zeros except the one whose index is equal to that of the maximum $|a_{ki}|$. Consequently, the estimate of e in (6) is the input symbol sequence. Since

$$\mathbf{d}^\dagger \mathbf{d} = \mathbf{b}^\dagger \mathbf{A}\mathbf{A}^\dagger \mathbf{b}$$

by (8), (9) is equivalent to

$$\begin{cases} \text{maximize:} & |F_k(\mathbf{b})| \\ \text{subject to:} & \mathbf{b}^\dagger \mathbf{A}\mathbf{A}^\dagger \mathbf{b} = 1. \end{cases} \tag{10}$$

Its solution $\mathbf{b}$ is the desired filter. Now, the problem becomes to determine the optimum solution to (10) with an unknown $\mathbf{A}$. By substituting $e = \mathbf{b}^\dagger \mathbf{o}$ into (7), we have

$$F_k(\mathbf{b}) = \sum_{i=1}^{N'} \sum_{j=1}^{N'} b_i^* b_j p_{ij}^k$$

where

$$p_{ij}^k = E\{o_i o_j^* o_k o_k^*\} - E\{o_i o_j^*\}E\{o_k o_k^*\} - E\{o_i o_k^*\}E\{o_j^* o_k\} - E\{o_i o_k\}E\{o_j^* o_k^*\}. \tag{11}$$

Let $\mathbf{P}_k = [p_{ij}^k]_{i,j=1}^{N'}$, we then have

$$F_k(\mathbf{b}) = \mathbf{b}^\dagger \mathbf{P}_k \mathbf{b}. \tag{12}$$

Since $p_{ij}^k = p_{ji}^{k*}$ for any $i, j = 1, 2, \cdots, N'$, $\mathbf{P}_k$ is an $N' \times N'$ *Hermitian* matrix and thereby $F_k(\mathbf{b})$ is a quadratic form of $\mathbf{b}$. By assumption **A2**, the covariance matrix of the received vector sequence is

$$\mathbf{Z}_o = E\{\mathbf{o}\mathbf{o}^\dagger\} = \mathbf{A}\mathbf{A}^\dagger + \sigma^2 \mathbf{I}_{N'} \tag{13}$$

which is equivalent to

$$\mathbf{A}\mathbf{A}^\dagger = \mathbf{Z}_o - \sigma^2 \mathbf{I}_{N'} \tag{14}$$

where $\mathbf{I}_{N'}$ is an $N' \times N'$ identity matrix and σ^2 is the variance of the noise. Here we assume that each component of $\mathbf{w}$ has the same variance. This is reasonable because all the noise components come from the same samples of $w(t)$ at the receiver. By (12) and (14), the problem (10) becomes

$$\begin{cases} \text{maximize:} & |\mathbf{b}^\dagger \mathbf{P}_k \mathbf{b}| \\ \text{subject to:} & \mathbf{b}^\dagger (\mathbf{Z}_o - \sigma^2 \mathbf{I}_{N'}) \mathbf{b} = 1. \end{cases} \tag{15}$$

Let $\mathbf{Z}_o = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^\dagger$ be the singular value decomposition (SVD) of $\mathbf{Z}_o$, then

$$\mathbf{A}\mathbf{A}^\dagger = \mathbf{\Gamma}(\mathbf{\Lambda} - \sigma^2 \mathbf{I}_{N'})\mathbf{\Gamma}^\dagger \tag{16}$$

where $\mathbf{\Gamma}$ is a unitary matrix and $\mathbf{\Lambda}$ a diagonal matrix with its entries on the main diagonal in a descending order. Define $\mathbf{\Lambda}' = \mathbf{\Lambda} - \sigma^2 \mathbf{I}_{N'}$. Since $\mathbf{A}$ is an $N' \times M'$ full-column rank matrix, the last $N' - M'$ entries on the main diagonal of $\mathbf{\Lambda}'$ are all zeros. By removing the last $N' - M'$ rows and columns of

$\mathbf{\Lambda}'$ and the last $N' - M'$ columns of $\mathbf{\Gamma}$, we have an $M' \times M'$ matrix $\tilde{\mathbf{\Lambda}}$ and an $N' \times M'$ matrix $\tilde{\mathbf{\Gamma}}$. Let

$$\mathbf{L} = \tilde{\mathbf{\Gamma}}\tilde{\mathbf{\Lambda}}^{1/2} \tag{17}$$

then (15) becomes

$$\begin{cases} \text{maximize:} & |\mathbf{u}^\dagger \mathbf{R}_k \mathbf{u}| \\ \text{subject to:} & \mathbf{u}^\dagger \mathbf{u} = 1 \end{cases} \tag{18}$$

where $\mathbf{u} = \mathbf{L}^\dagger \mathbf{b}$ and $\mathbf{R}_k = \mathbf{L}^{-1}\mathbf{P}_k\mathbf{L}^{-\dagger}$. Here $\mathbf{L}^{-1}$ is any pseudoinverse of matrix $\mathbf{L}$. Because $\mathbf{P}_k$ is an $N' \times N'$ Hermitian matrix, $\mathbf{R}_k$ is an $M' \times M'$ Hermitian matrix. Suppose that the eigenvector associated with the maximum absolute eigenvalue of $\mathbf{R}_k$ is $\mathbf{u}_m$. According to *Rayleigh's Principle* [34, p. 420], the solution to (18) is the eigenvector $\mathbf{u}_m$. The desired vector $\mathbf{b}$ is then computed by $\mathbf{b} = \mathbf{L}^{-\dagger}\mathbf{u}_m$. In summary, when there is only one element with the maximum absolute value on the selected kth row of $\mathbf{A}$, $\mathbf{b}$ can be found by the following procedure.

1) For a given k, compute $\mathbf{P}_k$ and $\mathbf{Z}_o$ with (11) and (13), respectively, and estimate the noise variance σ^2.
2) Make the singular value decomposition $\mathbf{Z}_o = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^\dagger$, then generate $\tilde{\mathbf{\Gamma}}$ and $\tilde{\mathbf{\Lambda}}$, and compute $\mathbf{L} = \tilde{\mathbf{\Gamma}}\tilde{\mathbf{\Lambda}}^{\frac{1}{2}}$.
3) Construct the matrix $\mathbf{R}_k = \mathbf{L}^{-1}\mathbf{P}_k\mathbf{L}^{-\dagger}$ and find the eigenvector $\mathbf{u}_m$ associated with its maximum absolute eigenvalue.
4) Compute the desired vector $\mathbf{b} = \mathbf{L}^{-\dagger}\mathbf{u}_m$.

B. Joint Criterion for Estimate Improvement

In the above development, we assume that there is only one element with the maximum absolute value on the kth row of $\mathbf{A}$. If this assumption does not hold, then $\mathbf{b}$ obtained by the above procedure may not give an acceptable estimate of the input symbol sequence. Unfortunately, in practice we cannot verify this assumption due to the lack of information about $\mathbf{A}$. In order to have a reliable solution, in this subsection we propose an algorithm based on a joint criterion for the computation of $\mathbf{b}$. The criterion is composed of all functions in (7) for $k = 1, 2, \cdots, N'$.

Let $\mathbf{X}$ be an $M' \times M'$ unitary matrix. We defined a function of $\mathbf{X}$ as

$$G(\mathbf{X}) = \sum_{i=1}^{N'} \|\operatorname{diag}(\mathbf{X}\mathbf{R}_i\mathbf{X}^\dagger)\|^2 \tag{19}$$

where $\mathbf{R}_i$ is the same in (18) and $\|\operatorname{diag}(\mathbf{X}\mathbf{R}_i\mathbf{X}^\dagger)\|$ is the Euclidean norm of the main diagonal of $\mathbf{X}\mathbf{R}_i\mathbf{X}^\dagger$.

Lemma: Let $\mathbf{D} = \mathbf{X}\mathbf{L}^{-1}\mathbf{A}$ and denote $\mathbf{Q}_k(\mathbf{D}) = \mathbf{D}\mathbf{\Lambda}_k\mathbf{D}^\dagger = [q_{ij}^k]_{i,j=1}^{M'}$ for $k = 1, 2, \cdots, N'$ where $\mathbf{\Lambda}_k$ is a diagonal matrix $\mathbf{\Lambda}_k = c_s \times \operatorname{diag}\{|a_{k1}|^2, |a_{k2}|^2, \cdots, |a_{kM'}|^2\}$, then

$$G(\mathbf{X}) = \sum_{k=1}^{N'} \left(\sum_{i=1}^{M'} |c_s|^2 |a_{ki}|^4 - \sum_{i=1}^{M'} \sum_{j \neq i}^{M'} \left| q_{ij}^k \right|^2 \right). \tag{20}$$

The proof of the lemma is presented in the Appendix. With the lemma, it is easy to see that for any $M' \times M'$ unitary matrix $\mathbf{X}$

$$G(\mathbf{X}) \le |c_s|^2 \sum_{i=1}^{N'} \sum_{j=1}^{M'} |a_{ij}|^4.$$

Now suppose $\mathbf{V} = [v_{ij}]_{i,j=1}^{M'}$ is such a unitary matrix that $\mathbf{D} = \mathbf{V}\mathbf{L}^{-1}\mathbf{A}$ is a permutation matrix, then any component of $\mathbf{e} = \mathbf{V}\mathbf{L}^{-1}\mathbf{o}$ can be viewed as an equalized result. In this case, denoting $\mathbf{v}_j$ as the jth column of $\mathbf{V}$, we have

$$\begin{aligned} G(\mathbf{V}) &= \sum_{i=1}^{N'} \sum_{j=1}^{M'} |\mathbf{v}_j^\dagger \mathbf{R}_i \mathbf{v}_j|^2 = \sum_{i=1}^{N'} \sum_{j=1}^{M'} |F_i(\mathbf{d}_j)|^2 \\ &= |c_s|^2 \sum_{i=1}^{N'} \sum_{j=1}^{M'} |a_{ij}|^4 \end{aligned} \tag{21}$$

which claims that $G(\mathbf{V})$ is the maximum of $G(\mathbf{X})$. The problem now becomes that if $G(\mathbf{V})$ is the maximum, is $\mathbf{D} = \mathbf{V}\mathbf{L}^{-1}\mathbf{A}$ still a permutation matrix? In order to answer this problem, we need a definition: two vectors with the same dimension are said to be symmetric if all their corresponding components have the same absolute value.

Theorem: Let $\mathbf{X}$ be an $M' \times M'$ unitary matrix and $\mathbf{D} = \mathbf{X}\mathbf{L}^{-1}\mathbf{A}$ where $\mathbf{L}$ is defined in (17) and $\mathbf{A}$ is the channel convolution matrix. If there do not exist two column vectors of $\mathbf{A}$ being symmetric, then $G(\mathbf{X})$ reaches its maximum if, and only if, $\mathbf{D}$ is a permutation matrix.

Proof: The sufficient condition has been proved in the above discussion. Now we prove that it is also necessary. By the above lemma, $q_{ij}^k = 0$ for $i \neq j$, $k = 1, 2, \cdots, N'$. Therefore, the expansion of $\mathbf{D}\mathbf{\Lambda}_k\mathbf{D}^\dagger = [q_{ij}^k]_{i,j=1}^{M'}$ gives

$$q_{ij}^k = \sum_{l=1}^{M'} d_{il} d_{jl}^* c_s |a_{kl}|^2 = 0, \quad \text{for } i, j = 1, 2, \cdots, M',\ i \neq j \tag{22}$$

where d_{il} is the (i, l)th entry of $\mathbf{D}$. On the other hand, by $\mathbf{A}\mathbf{A}^\dagger = \mathbf{L}\mathbf{L}^\dagger$ we have $\mathbf{D}\mathbf{D}^\dagger = \mathbf{I}_{M'}$. Therefore

$$\sum_{l=1}^{M'} d_{il} d_{jl}^* = 0, \quad \text{for } i \neq j \tag{23}$$

and

$$\sum_{l=1}^{M'} |d_{il}|^2 = 1, \quad \text{for } i = 1, 2, \cdots, M'. \tag{24}$$

Note that $\mathbf{A}$ and $\mathbf{L}$ have rank M', and $\mathbf{X}$ is an $M' \times M'$ unitary matrix. $\mathbf{D}$ is therefore a full rank matrix. Without loss of generality, consider the first row of $\mathbf{D}$. Suppose that $d_{1s} \neq 0$ and its cofactor is not zero. By (22), we have

$$\sum_{l \neq s}^{M'} d_{1l} d_{jl}^* |a_{kl}|^2 = -d_{1s} d_{js}^* |a_{ks}|^2, \quad \text{for } j = 2, 3, \cdots, M'. \tag{25}$$

If for some $k \in \{1, 2, \cdots, N'\}$, $a_{ks} = 0$, then (25) is a homogeneous equation system; thereby $d_{1l}|a_{kl}|^2 = 0$ for

$l \neq s$ because of the nonzero cofactor of d_{1s}. Since there is at least one $a_{kl} \neq 0$ for each l, we have $d_{1l} = 0$ for $l = 1, 2, \cdots, M'$ and $l \neq s$. On the other hand, if $a_{ks} \neq 0$ for all $k = 1, 2, \cdots, N'$, then (25) can be rewritten as

$$\sum_{l \neq s}^{M'} d_{jl}^* d_{1l} \left| \frac{a_{kl}}{a_{ks}} \right|^2 = -d_{1s} d_{js}^*, \quad \text{for } j = 2, \cdots, M'. \tag{26}$$

However, by (23) we have

$$\sum_{l \neq s}^{M'} d_{jl}^* d_{1l} = -d_{1s} d_{js}^*, \quad \text{for } j = 2, \cdots, M'. \tag{27}$$

A comparison of (26) with (27) yields

$$d_{1l} = d_{1l} \left| \frac{a_{kl}}{a_{ks}} \right|^2, \quad \text{for } l = 1, 2, \cdots, M', \, l \neq s.$$

If $d_{1l} \neq 0$, then $|a_{kl}| = |a_{ks}|$. This means that the lth and sth column vectors are symmetric. Consequently, d_{1s} is the unique nonzero entry on the first row of $\mathbf{D}$. By applying the above derivation to another row of $\mathbf{D}$, we have the conclusion. □

With this theorem, the equalization problem becomes finding the solution to

$$\begin{cases} \text{maximize:} & G(\mathbf{X}) \\ \text{subject to:} & \mathbf{X}\mathbf{X}^\dagger = \mathbf{I}_{M'}. \end{cases} \tag{28}$$

If, in the channel convolution matrix $\mathbf{A}$, there do not exist any two-column vectors being symmetric, any column of the solution $\mathbf{X}$ can be used to compute the desired vector $\mathbf{b} = \mathbf{L}^{-\dagger}\mathbf{x}_k$, where $\mathbf{x}_k$ is the kth column of $\mathbf{X}$. Note that the solution $\mathbf{X}$ is a unitary matrix rather than a column vector. Although computing $\mathbf{X}$ may require more computation than computing one column vector, there are some benefits behind. As any column of $\mathbf{X}$ can provide a qualified equalizer $\mathbf{b}$ and the performances of these equalizers are dramatically different when there is an additive noise, it is the matrix $\mathbf{X}$ that provides a possibility to select a suitable equalizer. We shall discuss the selection of the column of $\mathbf{X}$ in the next subsection. Because of the special structure of $\mathbf{A}$, the "symmetric columns" condition on $\mathbf{A}$ never occurs by carefully constructing $\mathbf{A}$. For example, if $N = 2$ (i.e., the oversampling factor is 2) and $N' > M'$ (i.e., the dimension of vector $\mathbf{o}$ is N'), then there will be no symmetric columns in $\mathbf{A}$.

To solve the above optimization problem, let us consider the set of matrices $\mathbf{X}\mathbf{R}_k\mathbf{X}^\dagger$ for $k = 1, 2, \cdots, N'$. By substituting (5) into (11), we have

$$p_{ij}^k = c_s \sum_{l=1}^{M'} a_{il} a_{jl}^* |a_{kl}|^2$$

by which

$$\mathbf{P}_k = \mathbf{A}\mathbf{\Lambda}_k\mathbf{A}^\dagger$$

where $\mathbf{A}$ is the channel convolution matrix and $\mathbf{\Lambda}_k$ is defined in the Lemma. Therefore, for $k = 1, 2, \cdots, N'$

$$\begin{aligned} \mathbf{X}\mathbf{R}_k\mathbf{X}^\dagger &= \mathbf{X}\mathbf{L}^{-1}\mathbf{P}_k\mathbf{L}^{-\dagger}\mathbf{X}^\dagger \\ &= (\mathbf{X}\mathbf{L}^{-1}\mathbf{A})\mathbf{\Lambda}_k(\mathbf{X}\mathbf{L}^{-1}\mathbf{A})^\dagger \\ &= \mathbf{D}\mathbf{\Lambda}_k\mathbf{D}^\dagger = \mathbf{Q}_k(\mathbf{D}). \end{aligned} \tag{29}$$

From the theorem, when $\mathbf{X}$ is the solution to (28), all the off-diagonal entries of $\mathbf{Q}_k(\mathbf{D})$ for $k = 1, 2, \cdots, N'$ are zeros. Thus, from (29), $\mathbf{X}$ simultaneously diagonalizes the set of matrices $\mathbf{R}_k$ for $k = 1, 2, \cdots, N'$. By (20) and (22), this is also sufficient for the maximum solution. In order to diagonalize these matrices, an algorithm in [33] can be employed. This algorithm is designed by extending the *Jacobi* technique for diagonalising a single Hermitian matrix to a set of Hermitian matrices. The key point in this algorithm lies on the computation of the Givens rotation. Its computation complexity is the same as that of Jacobi algorithm for diagonalizing a single matrix.

C. Practical Implementation

In practice, because of the noise and the short length of the received samples, we need to solve several problems. First, in order to compute matrix $\mathbf{L}$, it is required to estimate the covariance matrix $\mathbf{Z}_o$ and the noise variance σ^2. $\mathbf{Z}_o$ is approximately equal to the ensemble-averaged covariance matrix of the received sequence. When the length of received sequence is long enough, $\mathbf{Z}_o$ can be precisely determined. By (16), the noise variance is the smallest eigenvalue of $\mathbf{Z}_o$. However, for a short received sequence, it is almost impossible to have a reasonable estimate of the noise variance; the reason is that the noise contribution in the estimate of $\mathbf{Z}_o$ is comparable to the error caused by the limited received samples. Therefore, it is meaningless to determine the noise variance and remove it from $\mathbf{Z}_o$. Instead of the use of (17) for a short received sequence, matrix $\mathbf{L}$ can be simply computed by $\mathbf{\Gamma}\mathbf{\Lambda}^{\frac{1}{2}}$, i.e., we simply ignore the noise contribution in our algorithm.

Second, we choose M' as the dimension of matrix $\tilde{\mathbf{\Lambda}}$ in (17). However, because we ignore the noise in (16) (i.e., we set $\sigma^2 = 0$), we have $\tilde{\mathbf{\Lambda}} = \mathbf{\Lambda}$, thereby M' is always chosen as N'. This provides an over-estimated M' generally. One of the main advantages of our method based on (28) is that it is still applicable whenever M' is over-estimated. This point can be seen as follows. Suppose that M' is over-estimated as N', then $\mathbf{X}$ in (28) is an $N' \times N'$ matrix and $\mathbf{D} = \mathbf{X}\mathbf{L}^{-1}\mathbf{A}$ an $N' \times M'$ matrix. Although $\mathbf{D}$ is no longer a square matrix, it is easy to verify that the Lemma still holds, i.e.,

$$G(\mathbf{X}) = \sum_{k=1}^{N'} \left(\sum_{i=1}^{M'} |c_s|^2 |a_{ki}|^4 - \sum_{i=1}^{N'} \sum_{j \neq i}^{N'} \left| q_{ij}^k \right|^2 \right).$$

When $\mathbf{X}$ is a solution to (28), we can conclude that each column of $\mathbf{D}$ has no more than one nonzero entry. The proof can be simply done by applying $\mathbf{D}\mathbf{D}^\dagger = \text{diag}\{\mathbf{I}_{M'}, \underline{0}\}$ to the proof of the above theorem, where $\underline{0}$ is a zero vector with dimension $N' - M'$.

Third, because of the overestimate of M', $\mathbf{D}$ is a "tall" matrix, and some of its rows are zero vectors. Consequently, by

$$\mathbf{e} = \mathbf{X}\mathbf{L}^{-1}\mathbf{o} = \mathbf{D}\mathbf{s} + \mathbf{X}\mathbf{L}^{-1}\mathbf{w} \tag{30}$$

a component of $\mathbf{e}$ is either an unbiased estimate of the symbol sequence or consists only of noise. Notice that the noise is assumed to be Gaussian distributed, its higher order cumulants are all zeros. If the jth component of $\mathbf{e}$ contains noise only, then its fourth-order cumulant is null. This means that $F_k(\mathbf{d}_j) \approx 0$ for all $k = 1, 2, \cdots, N'$, i.e., $|\mathbf{x}_j^\dagger \mathbf{R}_k \mathbf{x}_j| \approx 0$ for all $k = 1, 2, \cdots, N'$. In fact, $\mathbf{x}_j^\dagger \mathbf{R}_k \mathbf{x}_j$ is the eigenvalue of $\mathbf{R}_k$ associated with eigenvector $\mathbf{x}_j$. It is a byproduct after $\mathbf{X}$ is obtained. For each matrix $\mathbf{R}_k$, there is an eigenvalue associated with $\mathbf{x}_j$. Hence, if there exists at least one nonzero eigenvalue associated with $\mathbf{x}_j$, then the jth component of $\mathbf{e}$ can be considered as a symbol sequence. Obviously, the eigenvector corresponding to the maximal absolute eigenvalue is always a candidate in the estimation of the symbol sequence; it will be used in our algorithm described in the next subsection.

D. The Algorithm

We now design a blind equalizer based on the above theorem. Although the equalizer requires an additional condition that the channel matrix $\mathbf{A}$ must have no symmetric columns, this condition is very mild or can be avoided in practice. The algorithm is stated as follows.

1) Collect an oversampled sequence $\{y_i, i = 0, 1, 2, \cdots\}$ and construct the vector samples $\{\mathbf{o}(n)\ n = 1, 2, \cdots\}$ based on N and an over-estimated channel order M.
2) Compute the covariance matrix of $\mathbf{o}$: $\mathbf{Z}_o = E\{\mathbf{o}\mathbf{o}^\dagger\}$.
3) Decompose the matrix $\mathbf{Z}_o$ into $\mathbf{\Gamma\Lambda\Gamma}^\dagger$, and compute $\mathbf{L} = \mathbf{\Gamma\Lambda}^{\frac{1}{2}}$.
4) For $k = 1, 2, \cdots, N'$ compute the matrix $\mathbf{P}_k = [p_{ij}^k]$ by using (11) and $\mathbf{R}_k = \mathbf{L}^{-1}\mathbf{P}_k\mathbf{L}^{-\dagger}$.
5) Find $\mathbf{X}$ (e.g., using the algorithm in [33]), a solution to
$$\begin{cases} \text{maximize:} & \sum_{i=1}^{N'} \|\text{diag}(\mathbf{X}\mathbf{R}_i\mathbf{X}^\dagger)\|^2 \\ \text{subject to:} & \mathbf{X}\mathbf{X}^\dagger = \mathbf{I}_{N'}. \end{cases}$$
For each $\mathbf{R}_i$, $i = 1, 2, \cdots, N'$ there are N' eigenvalues associated with $\mathbf{X}$. Denote Φ as the set of all these $(N')^2$ eigenvalues.
6) Let $\mathbf{x}_m$ be the column of $\mathbf{X}$, which is associated with the eigenvalue in Φ that has the maximal absolute value. The input symbol sequence then can be approximated by
$$e = \mathbf{x}_m^\dagger \mathbf{L}^{-1}\mathbf{o}.$$
7) End.

This algorithm is developed for a short received sequence. The noise variance σ^2 is not removed from $\mathbf{Z}_o$ in the computation of $\mathbf{L}$. In comparison with most SOS-based algorithms, this algorithm seems to be time-consuming, since a set of matrices $\mathbf{R}_i$ is computed and a joint diagonalization of these matrices has to be achieved. However, it is worthwhile to mention that the dimension of the matrices in this algorithm is generally smaller. The computation complexity of this algorithm is therefore comparable to the existing algorithms. The simulation examples in the next section show that this algorithm works well when the SNR is not too small.

IV. Simulation Examples

In this section, several *Monte Carlo* simulation examples are presented to illustrate the performance of the algorithm proposed above. The symbols were drawn from the QPSK signal constellation with a uniform distribution. One may check that the fourth-order cumulant of the symbol sequence is nonzero. We define

TABLE I
Channel Impulse Responses

Channel #1	Channel #2	Channel #3
-0.5232 - 0.4893i	-0.1761 - 0.1970i	0.1456 + 0.0786i
-0.2066 - 0.5700i	-0.6166 - 0.7580i	-0.7522 - 0.7205i
0.5075 + 0.0890i	0.5943 + 0.0064i	0.2615 - 0.2932i
-0.1803 - 0.0469i	-0.1080 - 0.0193i	0.2322 + 0.0644i
0.0911 + 0.0246i	0.0505 + 0.0110i	-0.1040 - 0.0315i

$$\text{SNR} = 10\log_{10}(\sigma_o^2/\sigma_w^2). \tag{31}$$

Here σ_o^2 denotes the variance of the noise-free part of the received signal, and σ_w^2 that of the additive Gaussian noise. The received sequence is generated by (3). The intersymbol interference (ISI) is employed as a performance measure of the estimated sequence which is defined by

$$\text{ISI} = \frac{\sum_{i=1}^{M'} |\hat{d}_i|^2 - \hat{d}_{\max}^2}{\hat{d}_{\max}^2} \tag{32}$$

where $\hat{d}_i$, $i = 1, 2, \cdots, M'$ are the combined impulse response of the cascade of the channel and the equalizer which are defined by $[\hat{d}_1, \hat{d}_2, \cdots, \hat{d}_{M'}] = \mathbf{x}_m^\dagger \mathbf{L}^{-1}\mathbf{A}$ in a *Monte Carlo* run, and $\hat{d}_{\max}$ is their maximum absolute value. We denote the number of the input symbols used in each *Monte Carlo* run as N_o.

A. Experiment 1: Varying the Number of Symbols

In this example, we investigated the performance improvement by the increase of the number of received vector samples N_o. The simulation channel was a three-ray multipath channel being truncated up to five symbol periods. Let the gain and time delay of the kth ray in this multipath channel be g_k and Δ_k, respectively. We choose $\Delta_1 = 0$, $\Delta_2 = 0.3T$ and $\Delta_3 = 1T$; and $g_1 = 1$, $g_2 = -0.7149 - 0.2375i$, $g_3 = -0.5138 - 0.6779i$. The transmitter waveform is a raised-cosine pulse with roll-off factor 11%. The received signal is sampled three times as fast as the symbol rate, i.e., $N = 3$. We then have three virtual channels whose impulse responses are listed in Table I. The SNR is fixed at 25 dB. Fig. 1 shows the sample constellations of these three virtual channel outputs and that of the equalized result with $N_o = 200$. By $M = 5$ and $N = 3$, $\lambda = 2$ or $M' = N' = 6$. The equalizer performance is determined by increasing N_o from 30 to 300 with a step size of 10. For each N_o 50 *Monte Carlo* runs were achieved. The input symbol sequence and the noise in each run are different. Fig. 2 exhibits the improvement of ISI behavior when the number of received samples increase. The curve tends to be flat when $N_o > 150$.

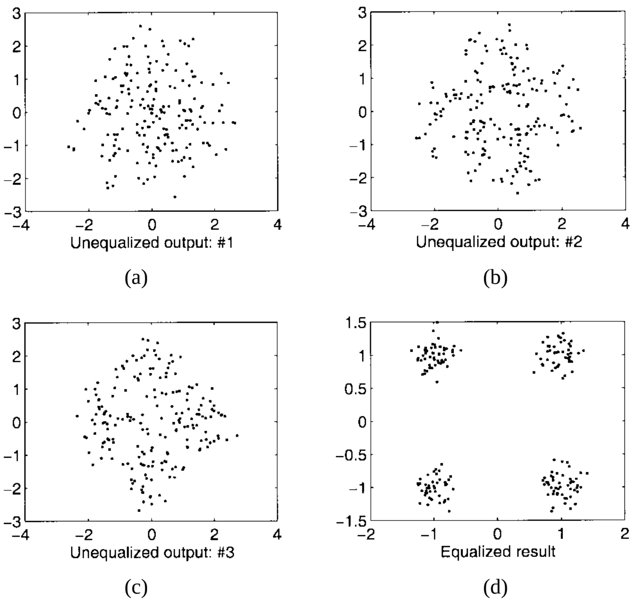

Fig. 1. The received sequences and equalized result with 200 symbols under SNR $= 25$ dB.

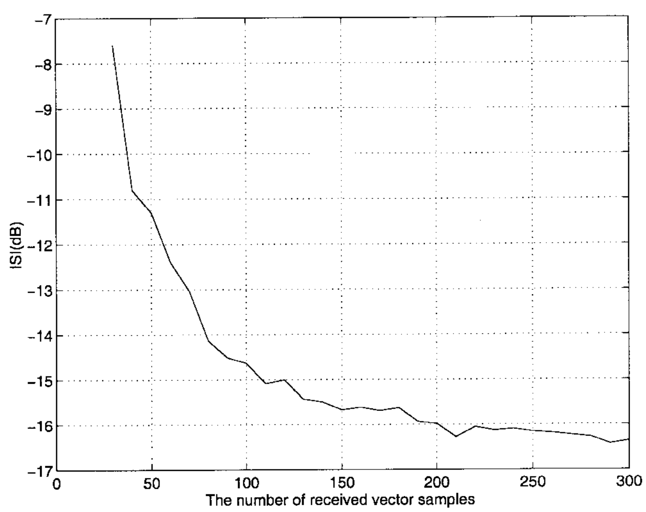

Fig. 2. ISI versus N_o under SNR $= 25$ dB.

B. Experiment 2: Varying the SNR

This experiment tests the performance of our algorithm subject to various levels of SNR. The channel is the same as that in the first experiment and the oversampling frequency is still three times as fast as the symbol rate, i.e., $M = 5$, $N = 3$, $M' = N' = 6$. Let $N_o = 150$ and the SNR varies from 0 to 40 dB with a step size of 2 dB. Again, for each noise level, 20 *Monte Carlo* runs were implemented. The result of the simulations is illustrated in Fig. 3. It is not surprising to see that the ISI is generally not affected by the noise when the SNR is not too small. The reason is that the noise is Gaussian distributed, and thereby its fourth-order cumulant is zero and contributes nothing to the criterion used by our approach. However, when the SNR is small, the ISI becomes bad. This is because in the computation of $\mathbf{L}$, we simply let $\sigma^2 = 0$ and it causes an inadmissible error.

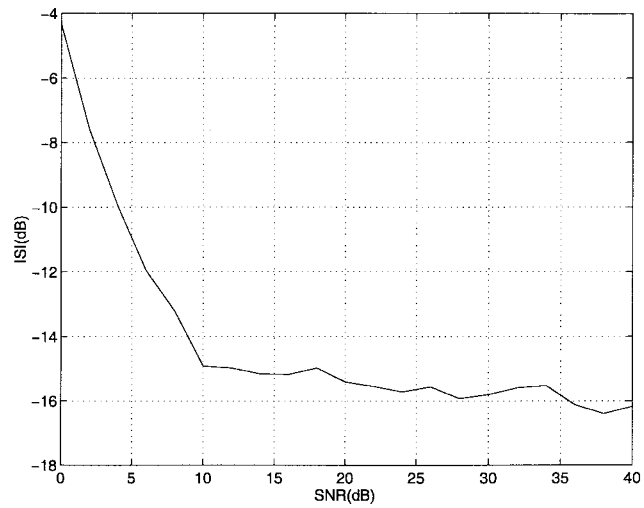

Fig. 3. ISI versus SNR with $N_o = 150$.

C. Experiment 3: Varying the Time Span of Channels

As previously mentioned, the main difficulty in the existing methods, that can reach an acceptable result with a small set of symbols, lies in the estimation of the channel order. In real implementation, the channel impulse response becomes smaller and smaller with the increase of its time span. Therefore, it is very hard to compute an accurate order of the channel based on a noisy and short received sample sequence. In order to demonstrate the insensitivity of our equalizer to the estimate error of channel order, in this experiment we fix the order of the simulation channel as 5 and change the estimate of the channel order from 2 to 12. As a result, when the oversampling frequency is threefold baud rate $N = 3$, the factor λ in (3) should be $1 \leq \lambda \leq 6$, i.e., $3 \leq N' \leq 18$ and $5 \leq M' \leq 10$. Again, the channel was composed of three rays, but we choose time delays as $\Delta_1 = 0$, $\Delta_2 = 0.5T$, and $\Delta_3 = 1.5T$, and the gains as $g_1 = 1$, g_2, and g_3 from a complex Gaussian random generator. Let SNR $= 25$ dB. For each particular value of the estimate of the channel order, 100 *Monte Carlo* runs were implemented. The results are shown in Fig. 4, where the solid curve is obtained by letting $N_o = 200$ and the dashed curve by $N_o = 400$. These curves tend to be flat when the estimated order is not less than the channel order 5. This means that our method indeed does not require an accurate estimate of the channel order.

D. Experiment 4: Fading Environments

The behavior of the equalizer in a fading environment was simulated. The channel was still a three-ray multipath channel, but we employed the *Jakes* Model [35] to simulate a fading environment. Suppose that the arrival angles of the received signal are uniformly distributed. Let f_D be the *Doppler* frequency, then the resultant complex fading envelope

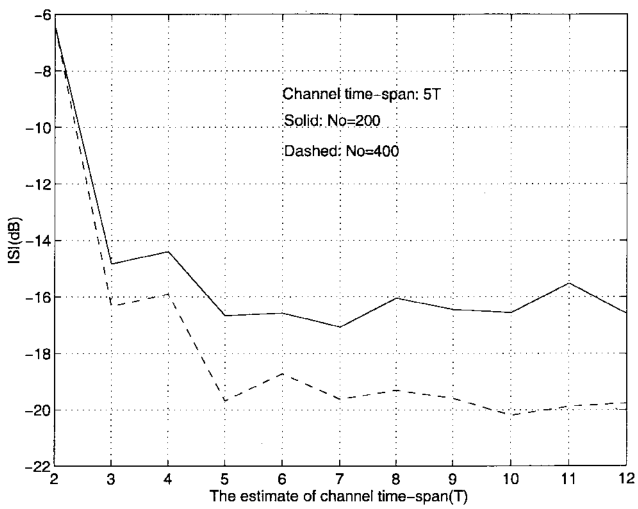

Fig. 4. ISI versus the mismatch of channel time span when the real channel time span is $5T$, under SNR $= 25$ dB.

TABLE II
PERFORMANCE IN A FADING ENVIRONMENT

Doppler Freq.	Burst error	Symbol error
f_D=50Hz	2.88%	0.0826%
f_D=5Hz	1.77%	0.0495%

of the signal from the nth path can be approximated by

$$g_n(t) = \sum_{k=1}^{8} \cos\left(2\pi f_D t \cos\left(\frac{\pi k}{17}\right) + \frac{\pi}{9}(2n+1)k\right) e^{j\frac{\pi k}{9}}$$

where $j = \sqrt{-1}$ and $n = 1, 2, 3$. A received signal from a fading channel can be expressed by

$$y(t) = g_1(t)z(t) + g_2(t-\Delta_2)z(t-\Delta_2) + g_3(t-\Delta_3)z(t-\Delta_3) + w(t) \quad (33)$$

where $w(t)$ is Gaussian noise and $z(t)$ is a single-path signal, defined by

$$z(t) = \sum_{i=-\infty}^{\infty} \alpha_i P(t - iT).$$

Here, $P(t)$ is the raised-cosine function. Each 150-symbol sequence is said to be a burst. In our experiments, 10 000 bursts were transmitted continuously through a fading channel; thereby a long received sequence could be obtained with (33). The channel was equalized once after each burst was received. When the roll-off factor of $P(t)$ was 0.9, the baud rate of the input symbol sequence was 270 kHz (i.e., used in GSM system), and $N = 2$, SNR $= 20$ dB, $\Delta_1 = 0$, $\Delta_2 = 0.3T$, $\Delta_3 = 0.9T$, the equalizer was tested by letting $M = 7$ (i.e., $M' = N' = 12$) for both $f_D = 50$ Hz and $f_D = 5$ Hz. After 10 000 bursts transmitted, 9712 bursts or 1 498 761 symbols were recovered correctly for $f_D = 50$ Hz and 9823 bursts or 1 499 258 symbols for $f_D = 5$ Hz. The simulation results are listed in Table II. There are, in total, 288 bursts containing error symbol(s) for $f_D = 50$ Hz, and 177 bursts for $f_D = 5$ Hz. Fig. 5 is the histogram of these error bursts. The first bar depicts the number of bursts containing one error symbol only, the second bar depicts the number of bursts containing two error symbols only, and so on. The last bar is an exception; it presents the total number of error bursts, each of which has more than six error symbols. Fig. 5 implies that among these error bursts, most of them include a very small number of error symbols.

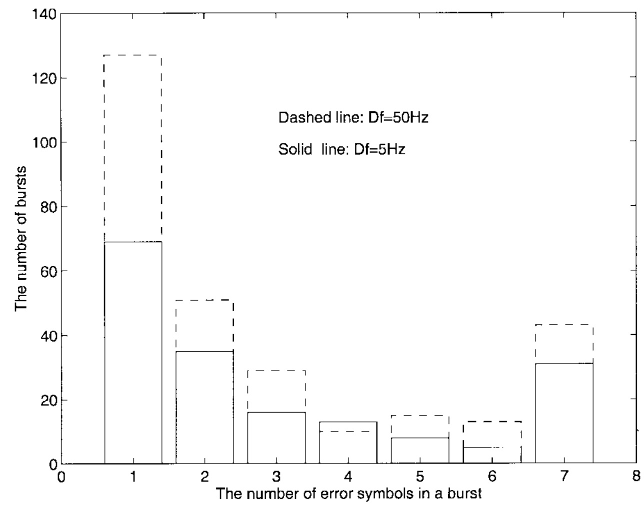

Fig. 5. The histogram of error bursts under SNR $= 20$ dB in 10 000 *Monte Carlo* runs.

V. CONCLUSION

In this paper, we developed a blind fractionally spaced equalizer in digital communications. The equalizer is designed on the fourth-order statistics of input symbol sequence and is resistant to the noise and errors in computation and modeling. Simulation results show that it works well with a short symbol sequence, even when the channel time span is not exactly known.

APPENDIX
PROOF OF THE LEMMA

Proof: For any unitary matrix $\mathbf{X}$

$$G(\mathbf{X}) = \sum_{k=1}^{N'}\sum_{j=1}^{M'} |\mathbf{x}_j^\dagger \mathbf{R}_k \mathbf{x}_j|^2 = \sum_{k=1}^{N'}\sum_{j=1}^{M'} |F_k(\mathbf{d}_j)|^2 = \sum_{k=1}^{N'} ||\mathrm{diag}(\mathbf{D}\mathbf{\Lambda}_k\mathbf{D}^\dagger)||^2 = \sum_{k=1}^{N'} ||\mathrm{diag}(\mathbf{Q}_k(\mathbf{D}))||^2$$

where $\mathbf{x}_j$ and $\mathbf{d}_j$ are the jth column of $\mathbf{X}$ and $\mathbf{D}^\dagger$, respectively. By $\mathbf{X}^\dagger\mathbf{X} = \mathbf{I}_{M'}$, we have $\mathbf{D}^\dagger\mathbf{D} = \mathbf{I}_{M'}$. Therefore

$$\mathbf{Q}_k(\mathbf{D})\mathbf{Q}_k(\mathbf{D})^\dagger = \mathbf{D}\mathbf{\Lambda}_k\mathbf{\Lambda}_k^\dagger\mathbf{D}^\dagger. \quad (34)$$

Since any $\mathbf{\Lambda}_k$ is a diagonal matrix, and $\mathbf{D}\mathbf{D}^\dagger = \mathbf{I}_{M'}$(if M' is over-estimated as $\hat{M}$, $\mathbf{D}\mathbf{D}^\dagger = \mathrm{diag}\{\mathbf{I}_{M'}, \underline{0}\}$ where $\underline{0}$ is a zero vector with dimension $\hat{M} - M'$), the trace of the left in (34) is equal to that of $\mathbf{\Lambda}_k\mathbf{\Lambda}_k^\dagger$

$$\sum_{i=1}^{M'}\sum_{j=1}^{M'}\left|q_{ij}^k\right|^2 = \sum_{i=1}^{M'}|c_s|^2|a_{ki}|^4 \tag{35}$$

i.e.,

$$\sum_{i=1}^{M'}\left|q_{ii}^k\right|^2 = \sum_{i=1}^{M'}|c_s|^2|a_{ki}|^4 - \sum_{i=1}^{M'}\sum_{j\neq i}^{M'}\left|q_{ij}^k\right|^2. \tag{36}$$

By

$$||\mathrm{diag}(\mathbf{Q}_k(\mathbf{D}))||^2 = \sum_{i=1}^{M'}\left|q_{ii}^k\right|^2 \tag{37}$$

we have

$$G(\mathbf{X}) = \sum_{k=1}^{N'}\left(\sum_{i=1}^{M'}|c_s|^2|a_{ki}|^4 - \sum_{i=1}^{M'}\sum_{j\neq i}^{M'}\left|q_{ij}^k\right|^2\right). \tag{38}$$

□

References

[1] X.-R. Cao and R. Liu, "General approach to blind signal separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 562–571, Mar. 1996.

[2] W. A. Gardner, "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Signal Processing Mag.*, pp. 14–36, Apr. 1991.

[3] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, Mar. 1994.

[4] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind channel identification based on second-order statistics: A frequency-domain approach," *IEEE Trans. Inform. Theory*, vol. 41, pp. 329–334, Mar. 1995.

[5] Z. Ding and Y. Li, "On channel identification based on second order cyclic spectra," *IEEE Trans. Signal Processing*, vol. 42, pp. 1260–1264, May 1994.

[6] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.

[7] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Processing*, vol. 44, pp. 661–672, Mar. 1996.

[8] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing*, vol. 43, pp. 2982–2993, Dec. 1995.

[9] A. J. van der Veen, S. Talwar, and A. Paulraj, "Blind estimation of multiple digital signals transmitted over FIR channels," *IEEE Signal Processing Lett.*, vol. 2, pp. 99–102, May 1995.

[10] A. J. van der Veen, S. Talwar, and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 173–190, Jan. 1997.

[11] D. Gesbert, P. Duhamel, and S. Mayrargue, "Subspace-based adaptive algorithms for the blind equalization of multichannel FIR filters," in *Proc. 1994 Eur. Signal Processing Conf. (EUSIPCO'94)*, pp. 712–715.

[12] G. B. Giannakis and C. Tapedelenlioglu, "Direct blind equalizers of multiple FIR channels: A deterministic approach," in *Proc. 30th Annu. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, 1996, pp. 290–294.

[13] D. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction," in *Proc. IEEE ICASSP*, Adelaide, Australia, May 1994, pp. 585–588.

[14] D. Slock and C. Papadias, "Further results on blind identification and equalization in multiple FIR channels," in *Proc. IEEE ICASSP*, May 1995, pp. 1964-1967.

[15] D. Gesbert and P. Duhamel, "Robust blind channel identification and equalization based on multi-step predictors," in *Proc. IEEE ICASSP*, 1997, pp. 3621–3624.

[16] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Trans. Inform. Theory*, vol. 43, pp. 499–511, Mar. 1997.

[17] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Trans. Signal Processing*, vol. 45, pp. 694–705, Mar. 1997.

[18] Z. Ding, "Matrix outer-product decomposition method for blind multiple channel identification," *IEEE Trans. Signal Processing*, vol. 45, pp. 3053–3061, Dec. 1997.

[19] G. B. Giannakis, Y. Inouye, and J. M. Mendel, "Cumulant based identification of multichannel moving-average models," *IEEE Trans. Automat. Contr.*, vol. 34, pp. 783–787, July 1989.

[20] A. Swami, G. B. Giannakis, and S. Shamsunder, "Multichannel ARMA processes," *IEEE Trans. Signal Processing*, vol. 42, pp. 898–913, Apr. 1994.

[21] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Inform. Theory*, vol. 36, pp. 312–321, Mar. 1990.

[22] O. Shalvi and E. Weinstein, "Super-exponential methods for blind deconvolution," *IEEE Trans. Inform. Theory*, vol. 39, pp. 504–519, Mar. 1993.

[23] F. B. Ueng and Y. T. Su, "Adaptive blind equalization using second- and higher-order statistics," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 132–140, Jan. 1995.

[24] C. Y. Chi and M. C. Wu, "Inverse filter criteria for blind deconvolution and equalization using two cumulants," *Signal Processing*, vol. 43, pp. 55–63, Jan. 1995.

[25] J. K. Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher order statistics and inverse filter criteria," *IEEE Trans. Signal Processing*, vol. 45, pp. 658–672, Mar. 1997.

[26] D. Hatzinakos and C. L. Nikias, "Blind equalization based on higher-order statistics (HOS)," in *Blind Deconvolution*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[27] G. B. Giannakis and M. K. Tsatsanis, "Restoring identifiability of fractionally sampled blind channel estimators using HOS," in *Proc. Int. Conf. HOS*, Barcelona, Spain, June 1995, pp. 409–413.

[28] L. Tong, "Identification of multivariate FIR systems using higher-order statistics," in *Proc. IEEE ICASSP*, May 1996, pp. 3037–3040.

[29] L. Tong, R. Liu, V. C. Soon, and Y. F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 499–509, May 1991.

[30] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, pp. 287–314, 1994.

[31] J. Zhu, X.-R. Cao, and M. L. Liou, "A unified algorithm for blind separation of independent sources," in *Proc. IEEE ISCAS*, Atlanta, GA, May 1996, vol. 2, pp. 153–156.

[32] J. Zhu, X.-R. Cao, and R. Liu, "A matrix decomposition method for blind equalization of FIR channels," in *Proc. IFAC Symp. System Identification (SYSID'97)*, Japan, July 1997, vol. 3, pp. 1081–1086.

[33] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Matrix Anal., Applicat.*, vol. 17, no. 1, pp. 161–164, Jan. 1995.

[34] B. Noble and J. W. Daniel, *Applied Linear Algebra*, 3rd Edition, Prentice-Hall, 1988.

[35] W. C. Jakes Jr., *Microwave Mobile Communications.* New York: McGraw-Hill, 1982.

Jie Zhu (S'95–M'98) received the B.Eng. and M.Eng. degrees in computer engineering from Southwest Jiaotong University, China, and Northwestern Polytechnic University, China, in 1985 and 1988, respectively, and the Ph.D. degree in electrical and electronic engineering from Hong Kong University of Science and Technology (HKUST), Hong Kong, in 1997.

From 1988 to 1992, he was an Engineer and Project Manager at the Aeronautical Computing Technique Institute of China, working on speech recognition system. From 1993 to 1997, he was a Teaching and Research Assistant in the Department of Electrical and Electronic Engineering, HKUST. Currently, he is a Research Fellow at the Centre for Signal Processing, Nanyang Technological University, Singapore. His research interests include signal processing in communications, blind signal separation, higher order statistics, neural networks, and pattern recognition.

Xi-Ren Cao (SM'89–F'96) received the M.S. and Ph.D. degrees from Harvard University, Cambridge, MA, in 1981 and 1984, respectively.

He was a Research Fellow at Harvard University from 1984 to 1986. He then worked as a Principal and Consultant Engineer at Digital Equipment Corporation until October 1993. Since then, he has been a Professor with the Hong Kong University of Science and Technology. He has held visiting positions at Harvard University, the University of Massachusetts at Amherst, AT&T Labs, and the University of Maryland at College Park. He has two patents in data communications and published two books, *Realization Probabilities—the Dynamics of Queuing Systems* (New York: Springer Verlag, 1994), and *Perturbation Analysis of Discrete-Event Dynamic Systems* (Norwell, MA: Kluwer, 1991), co-authored with Y. C. Ho. His current research interests include signal processing, discrete event systems, data communication networks, stochastic processes, and optimization techniques.

Dr. Cao received the Outstanding Transactions Paper Award from the IEEE Control System Society in 1987 and the Outstanding Publication Award from the Institution of Management Science in 1990. He has been an Associate Editor for a number of international journals and Chairman of a few technical committees of international professional societies.

Ruey-wen Liu (S'53–M'59–SM'78–F'81–LF'95) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Illinois, Urbana, IL.

Since 1960, he has been with the University of Notre Dame, Notre Dame, IN, and is currently the Frank M. Freimann Professor of Electrical and Computer Engineering. He was a Visiting Professor at the University of California at Berkeley during 1965–1966, 1977–1978, and 1984–1985, and at the Hong Kong University of Science and Technology, Hong Kong, in 1991–1992. He received Honorary Professorships from several institutions in China, including one from the Chinese Academy of Science, Beijing, China, in l988. His research interest are in the areas of nonlinear dynamics (circuits, systems, and control) and, most recently, signal processing (blind signal processing, neural networks and wireless communication).

Dr. Liu has held many positions with the IEEE, as well as in the IEEE Circuits and Systems Society. Among them, he has served as the Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS from l989 to 199l, and as President of the IEEE Circuits and Systems Society in 1995. He was Co-Guest Editor of PROCEEDINGS OF THE IEEE, Special Issue on Blind Identification and Estimation. In 1997, he was appointed Distinguished Lecturer by the IEEE Circuits and Systems Society, and was awarded the Circuits and Systems Society Meritorious Service Award in 1998.