

# Correspondence

## Lag-Windowing and Multiple-Data-Windowing Are Roughly Equivalent for Smooth Spectrum Estimation

Michael L. McCloud, Louis L. Scharf, and Clifford T. Mullis

**Abstract**—There is no fundamental difference between lag-windowing a correlation sequence and multiple-windowing a data sequence when the objective is to reduce the mean-squared error of a spectrum estimator. By analyzing the approximate low-rank factorization of a bandlimiting Toeplitz operator, we find that lag-windowed (or spectrally smoothed) spectrum estimators have multiple-data-windowed implementations. This makes the Blackman–Tukey–Grenander–Rosenblatt spectrogram equivalent to the Thomson spectrum estimator (and vice-versa), meaning BTGR spectrograms may be implemented in a multichannel filterbank version of the Thomson estimator.

### I. INTRODUCTION

Nonparametric estimators of the power spectrum are often computed as Blackman–Tukey–Grenander–Rosenblatt (BTGR) [1], [2] lag-windowed spectrograms or as Thomson multiple-data-windowed estimators [3]. Although these two procedures appear to be quite different, we will argue that they are essentially equivalent. That is, the net effect of lag-windowing is the same as multiple data-windowing. This result will follow from the factorization of the kernel matrices for lag-windowed spectrum estimators.

The estimation problem is as follows. We are given  $N$  samples  $\mathbf{y} = (y_0 \cdots y_{N-1})^T$  from a complex wide sense stationary (WSS) random process  $\{y_n\}$  with autocorrelation sequence  $\{r_k\}$  and power spectral density  $\{S(e^{j\theta}), -\pi < \theta \leq \pi\}$

$$r_k = E[y_s y_{s-k}^*] \quad (1)$$

$$S(e^{j\theta}) = \sum r_k e^{-j\theta k}. \quad (2)$$

We seek to estimate  $S(e^{j\theta})$  from this finite length realization  $\mathbf{y}$ .

We will only consider estimators  $\hat{S}(e^{j\theta})$ , which possess the following properties:

- P1)** quadratic in the data  $\hat{S}(e^{j\theta}) = \mathbf{y}^* \mathbf{Q}(\theta) \mathbf{y}$  ( $\mathbf{y}^*$  denotes the Hermitian transpose of  $\mathbf{y}$ );
- P2)** non-negative  $\hat{S}(e^{j\theta}) \geq 0$ ;
- P3)** modulation invariant  $(\mathbf{D}(e^{j\phi}) \mathbf{y})^* \mathbf{Q}(\theta) (\mathbf{D}(e^{j\phi}) \mathbf{y}) = \mathbf{y}^* \mathbf{Q}(\theta - \phi) \mathbf{y}$ , where  $\mathbf{D}(e^{j\phi}) = \text{diag}[1, e^{j\phi}, \dots, e^{j(N-1)\phi}]$  is a modulation matrix.

It has been shown [4] that any spectrum estimator that enjoys the preceding properties must be a quadratic form in the complex demodulated data  $\mathbf{D}(e^{-j\theta}) \mathbf{y}$ :

$$\hat{S}(e^{j\theta}) = \mathbf{y}^* \mathbf{D}(e^{j\theta}) \mathbf{Q}(0) \mathbf{D}(e^{-j\theta}) \mathbf{y}; \quad \mathbf{Q}(0) \geq 0. \quad (3)$$

This representation will allow us to explore the similarities between different estimation procedures.

Manuscript received February 4, 1998; revised August 28, 1998. This work was supported by the National Science Foundation under Contract MIP-9529050 and by the Office of Naval Research under Contract N00014-89-J-1070. The associate editor coordinating the review of this paper and approving it for publication was Prof. Antonio Cantoni.

The authors are with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309-0425 USA (e-mail: mccloud@ucsu.colorado.edu).

Publisher Item Identifier S 1053-587X(99)01346-X.

### II. MULTIPLE WINDOW SPECTRUM ESTIMATORS

Factoring the non-negative matrix  $\mathbf{Q}(0)$  from (3) into  $\mathbf{Q}(0) = \mathbf{V}^* \mathbf{V}$ , we obtain the equivalent representations

$$\begin{aligned} \hat{S}(e^{j\theta}) &= \|\mathbf{V} \mathbf{D}(e^{-j\theta}) \mathbf{y}\|^2 \\ &= \sum_{i=1}^M \left| \sum_{n=0}^{N-1} v_{i,n} y_n e^{-j\theta n} \right|^2 \end{aligned} \quad (4)$$

where  $\{v_{i,n}\}_{n=0}^{N-1}$  is the window sequence formed from the  $i$ th row of  $\mathbf{V}$ , and  $M = \text{rank}(\mathbf{Q}(0))$ . An implementation of this estimator employing a bank of windowed periodograms is shown in Fig. 1. Notice that this decomposition is valid for any spectrum estimator that satisfies conditions **P1)–P3)**.

When the windows are the  $M$ -dominant Slepian sequences [5] weighted by the square root of the corresponding eigenvalues, we have the Thomson estimator [3]. We will use the term Thomson estimator to mean any multiple data window estimator employing a small number of windows (relative to the data length).

### III. LAG-WINDOW SPECTRUM ESTIMATORS

We define the BTGR lag-window spectrum estimator with positive semidefinite window sequence  $\{w_k\}_{-(N-1)}^{N-1}$  by

$$\begin{aligned} \hat{S}_{\text{BT}}(e^{j\theta}) &= \sum_{k=-(N-1)}^{N-1} w_k \hat{r}_k e^{-j\theta k} \\ &= \int_{-\pi}^{\pi} W(e^{j(\theta-\phi)}) \hat{S}_P(e^{j\phi}) \frac{d\phi}{2\pi}. \end{aligned} \quad (5)$$

In these two formulas,  $\hat{r}_t$  is the biased estimator of the autocorrelation sequence, and  $\hat{S}_P(e^{j\theta})$  is the periodogram:

$$\begin{aligned} \hat{r}_k &= \frac{1}{N} \sum_{n=k}^{N-1} y_n y_{n-k}^* \quad 0 \leq k \leq N-1; \quad \hat{r}_{-k} = \hat{r}_k^* \\ \hat{S}_P(e^{j\theta}) &= \sum_{k=-(N-1)}^{N-1} \hat{r}_k e^{-j\theta k} \\ &= \frac{1}{N} \mathbf{y}^* \mathbf{\Psi}(e^{j\theta}) \mathbf{\Psi}^*(e^{j\theta}) \mathbf{y} \\ &\quad \text{where } \mathbf{\Psi}(e^{j\theta}) = [1 \quad e^{j\theta} \quad \dots \quad e^{j(N-1)\theta}]^T. \end{aligned} \quad (6)$$

The last expression for  $\hat{S}_P(e^{j\theta})$  may be substituted into (5) to produce the following representation for the BTGR spectrum estimator:

$$\begin{aligned} \hat{S}_{\text{BT}}(e^{j\theta}) &= \mathbf{y}^* \mathbf{D}(e^{j\theta}) \mathbf{W} \mathbf{D}(e^{-j\theta}) \mathbf{y} \\ \mathbf{W} &= \frac{1}{N} \int_{-\pi}^{\pi} W(e^{j\theta}) \mathbf{\Psi}(e^{j\theta}) \mathbf{\Psi}^*(e^{j\theta}) \frac{d\theta}{2\pi}. \end{aligned} \quad (7)$$

The BTGR spectrum estimator satisfies (3), with  $\mathbf{Q}(0) = \mathbf{W}$ : a positive-definite Toeplitz matrix.

Notice that the representation given in (7) is valid for estimators defined in either the time or frequency domain, as in (5). This

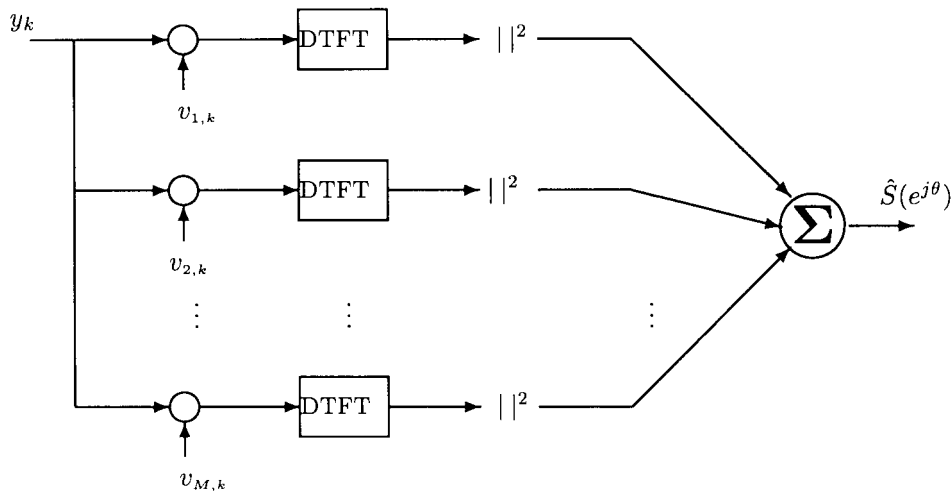


Fig. 1. Multichannel filterbank implementation of the multiple window estimator.

allows us to include in our discussion estimators, such as the Daniell spectrogram [6], that are based on manipulating the periodogram in the frequency domain.

#### IV. EQUIVALENCE OF LAG-WINDOWED AND MULTIPLE-WINDOWED SPECTRUM ESTIMATORS

We claim that lag-windowed spectrum estimators are basically equivalent to a multiple-data-window implementation using  $N\beta$  data windows, where  $2\pi\beta$  is the resolution bandwidth employed. We will defend this claim for “good” lag-window estimators that approximate the ideal lowpass filter of bandwidth  $2\pi\beta$ . We choose this characterization of good windows because it encapsulates the properties of a desirable window sequence: namely, out-of-band rejection and minimal distortion in the main lobe.

Let  $W(e^{j\theta})$  be the real, non-negative symmetric frequency response of a lag-window sequence  $\{w_k\}_{-(N-1)}^{N-1}$  of bandwidth  $2\pi\beta$ . Let  $B(e^{j\theta})$  be the frequency response of the ideal lowpass filter of bandwidth  $2\pi\beta$ . We assume that  $W(e^{j\theta})$  and  $B(e^{j\theta})$  are normalized so that

$$\int_{-\pi}^{\pi} W^2(e^{j\theta}) \frac{d\theta}{2\pi} = \int_{-\pi}^{\pi} B^2(e^{j\theta}) \frac{d\theta}{2\pi} = 1. \quad (8)$$

Define the real, symmetric error function

$$E(e^{j\theta}) = B(e^{j\theta}) - W(e^{j\theta}). \quad (9)$$

We say  $W(e^{j\theta})$  is a “good” lowpass filter (and, hence, window) if

$$\left( \int_{-\pi}^{\pi} E^2(e^{j\theta}) \frac{d\theta}{2\pi} \right)^{1/2} < \epsilon \quad (10)$$

for  $\epsilon$  a small positive number.

Each of these frequency responses defines a corresponding Toeplitz matrix:

$$\begin{aligned} \mathbf{W} &= \int_{-\pi}^{\pi} W(e^{j\theta}) \boldsymbol{\Psi}(e^{j\theta}) \boldsymbol{\Psi}^*(e^{j\theta}) \frac{d\theta}{2\pi} \geq 0 \\ \mathbf{B} &= \int_{-\pi}^{\pi} B(e^{j\theta}) \boldsymbol{\Psi}(e^{j\theta}) \boldsymbol{\Psi}^*(e^{j\theta}) \frac{d\theta}{2\pi} \geq 0 \\ \mathbf{E} &= \int_{-\pi}^{\pi} E(e^{j\theta}) \boldsymbol{\Psi}(e^{j\theta}) \boldsymbol{\Psi}^*(e^{j\theta}) \frac{d\theta}{2\pi}. \end{aligned} \quad (11)$$

For any eigenvalue  $\lambda \in \lambda(\mathbf{E})$  with corresponding eigenvector  $\mathbf{u} = (u_0, \dots, u_{N-1})^T$ , we have  $\lambda \mathbf{u} = \mathbf{E} \mathbf{u}$ , from which it follows that  $(\lambda \mathbf{u})_i = (\mathbf{E} \mathbf{u})_i$ . Therefore

$$\begin{aligned} |\lambda u_i| &= |(\mathbf{E} \mathbf{u})_i| \\ &= \left| \int_{-\pi}^{\pi} E(e^{j\theta}) U(e^{j\theta}) e^{j\theta i} \frac{d\theta}{2\pi} \right| \quad (\text{since } \boldsymbol{\Psi}^*(e^{j\theta}) \mathbf{u} = U(e^{j\theta})) \\ &\leq \int_{-\pi}^{\pi} |E(e^{j\theta}) U(e^{j\theta}) e^{j\theta i}| \frac{d\theta}{2\pi} \\ &\leq \left( \int_{-\pi}^{\pi} E^2(e^{j\theta}) \frac{d\theta}{2\pi} \right)^{1/2} \left( \int_{-\pi}^{\pi} |U(e^{j\theta})|^2 \frac{d\theta}{2\pi} \right)^{1/2} \end{aligned} \quad (12)$$

$$\leq \epsilon \quad (13)$$

where (12) follows from the Cauchy–Schwartz inequality, and (13) follows from the orthonormality of the eigenvectors of  $\mathbf{E}$  and the bound from (10). Since  $\mathbf{u}^* \mathbf{u} = 1$ , we see that at least one of the entries in  $\mathbf{u}$ , say  $u_k$ , satisfies  $|u_k| \geq 1/\sqrt{N}$ . Hence, our bound on each eigenvalue is

$$|\lambda| \leq \epsilon \sqrt{N}. \quad (14)$$

In particular, the matrix norm of  $\mathbf{E}$  satisfies

$$\|\mathbf{E}\|_2 = \max_i |\lambda_i| \leq \sqrt{N} \epsilon. \quad (15)$$

To complete our analysis, we use the Hoffman–Wielandt theorem for perturbed matrices, which is stated in the following theorem [7].

*Theorem 1:* Let  $\mathbf{A}, \mathbf{E}$  be two  $N \times N$  matrices. Assume that  $\mathbf{A}$  is Hermitian and that  $\mathbf{A} + \mathbf{E}$  is normal. Let  $\{\lambda_1, \dots, \lambda_N\}$  be the eigenvalues of  $\mathbf{A}$  arranged in decreasing order, and let  $\{\hat{\lambda}_1, \dots, \hat{\lambda}_N\}$  be the eigenvalues of  $\mathbf{A} + \mathbf{E}$  arranged so that  $\text{Re} \hat{\lambda}_1 \geq \text{Re} \hat{\lambda}_2 \geq \dots \geq \text{Re} \hat{\lambda}_N$ . Then

$$\left[ \sum_{i=1}^N |\hat{\lambda}_i - \lambda_i|^2 \right]^{1/2} \leq \|\mathbf{E}\|_2.$$

*Proof:* See [7].  $\square$

We may use this theorem to see that for  $\{\mu_i\} = \{\text{eigenvalues of } \mathbf{B}\}$  and  $\{\gamma_i\} = \{\text{eigenvalues of } \mathbf{W}\}$ , each arranged in increasing order, we have

$$\sqrt{\frac{1}{N} \sum_{i=1}^N |\gamma_i - \mu_i|^2} \leq \epsilon. \quad (16)$$

Therefore, for  $\epsilon$  small,  $\mathbf{B}$  and  $\mathbf{W}$  have equivalent eigenvalue distributions. We know [5] that  $\mathbf{B}$  has approximately  $N\beta$  dominant

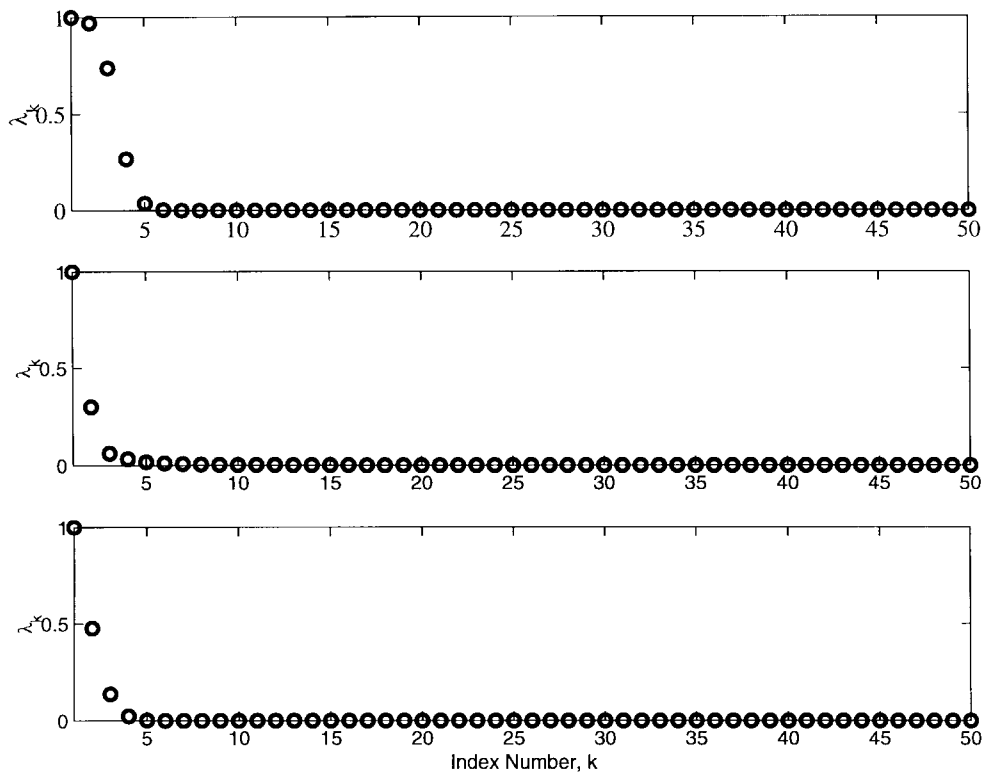


Fig. 2. First 50 eigenvalues corresponding to the three window sequences: Spectral Smoothing, Bartlett, and Kaiser for  $N = 300$ .

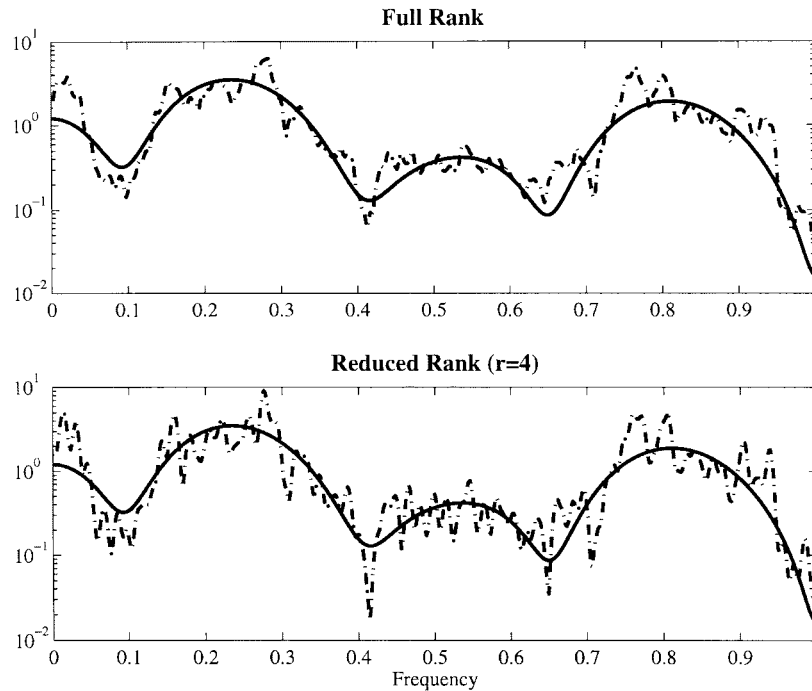


Fig. 3. Result of estimation for full rank and reduced rank Spectral Smoothing.

eigenvalues, and therefore, for  $\mathbf{W}$  corresponding to a small  $\epsilon$ , we have

$$\text{rank}(\mathbf{W}) \approx N\beta. \tag{17}$$

Hence, “good” lag-windowed spectrum estimators may be implemented by a multiple window estimator with  $\approx N\beta$  data windows corresponding to the weighted dominant eigenvectors of  $\mathbf{W}$ .

#### V. EXACTLY LOW-RANK BTGR SPECTRUM ESTIMATORS

The low-rank factorization is trivial for estimators based on sums of sinusoids (Hamming, Hanning, Blackman, etc.). This comes immediately from the definition of the window sequences. For example, the generalized Hamming window defined by

$$w_k = a + b \cos(\theta k) \tag{18}$$

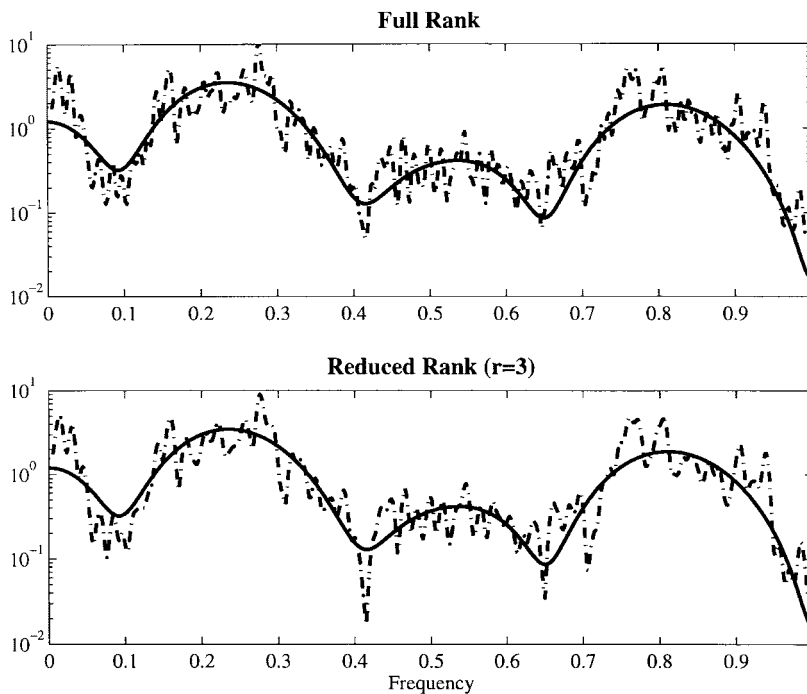


Fig. 4. Result of estimation for full rank and reduced rank Bartlett window.

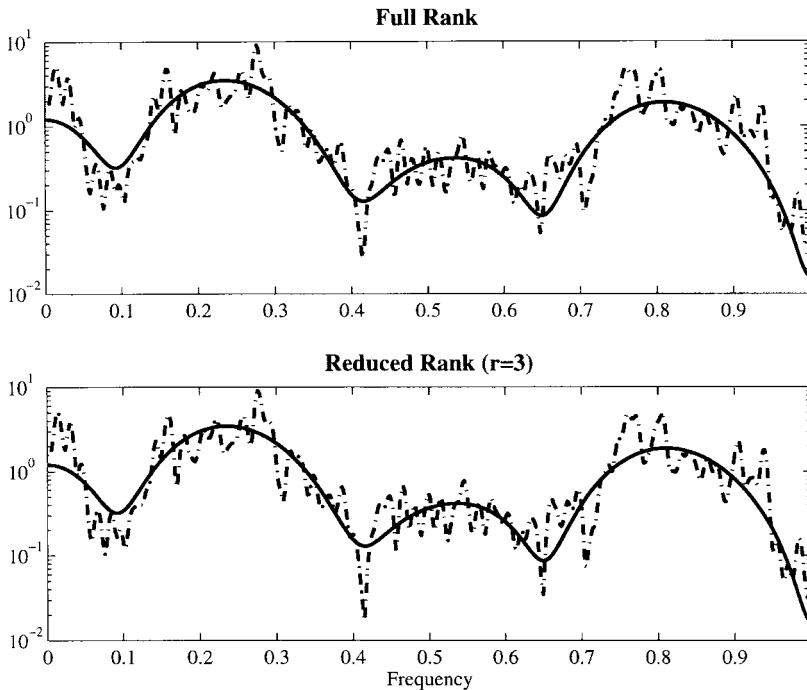


Fig. 5. Result of estimation for full rank and reduced rank Kaiser window.

has the estimator matrix

$$\mathbf{W} = a\mathbf{1}\mathbf{1}^T + \frac{b}{2}[\Psi(e^{j\theta})\Psi^*(e^{j\theta}) + \Psi(e^{-j\theta})\Psi^*(e^{-j\theta})] \quad (19)$$

where  $\mathbf{1} = [1, 1, \dots, 1]^T$  and, hence, has rank 3 independent of  $N$ . In a similar fashion, we find that the Blackman window has rank 7. The Daniell estimator [6] is given by

$$\hat{S}(e^{j\theta}) = \frac{1}{2K+1} \sum_{i=-K}^K \hat{S}_P(e^{j(\theta - \frac{2\pi i}{M})}) \quad (20)$$

$$= \int_{-\pi}^{\pi} W(e^{j(\theta-\phi)}) \hat{S}_P(e^{j\phi}) \frac{d\phi}{2\pi} \quad (21)$$

where

$$W(e^{j\theta}) = \frac{1}{2K+1} \sum_{i=-K}^K \delta\left(\theta - \frac{2\pi i}{M}\right). \quad (22)$$

The window sequence corresponding to this estimator is

$$w_k = \frac{1}{2K+1} \left( 1 + \sum_{i=-K}^K \cos\left(\frac{2\pi}{M} ik\right) \right) \quad (23)$$

so we see that  $\mathbf{W}$  has rank  $2K+1$ .

## VI. EXAMPLES

We will demonstrate the low-rank multiple window implementation of several commonly used BTGR window sequences for a 12th-order moving average process. In each case, our data length is 300 samples, which means the Rayleigh limit to resolution is  $2\pi/300$ . We will employ the Bartlett, Kaiser, and Spectral Smoothing windows. The latter is defined by  $w_k = \text{sinc}(\beta\pi k)$  where  $\beta$  is the normalized bandwidth over which the smoothing is to occur (we choose  $\beta = 0.01$  so that  $2\pi\beta$  is  $2\pi/100$ , which is three times the Rayleigh limit).

The eigenvalues of the matrices corresponding to each window are shown in Fig. 2. It is clear that each allows an extremely low-rank multiple window implementation. The results of the spectral estimation are shown in Figs. 3–5 for the full-rank BTGR implementation and for the reduced rank multiple window implementation. In all of the figures, the solid line is the exact MA spectrum. It can be seen that the low-rank approximations to the lag-windowed estimators give results that are very close to the full-rank estimates.

## VII. CONCLUSION

We have argued that BTGR spectrum estimators are essentially multiple-data-window estimators. This correspondence was shown for estimators employing “good” lag-window sequences, meaning they approximate an ideal lowpass filter with small  $L_2$  error. In this case, the kernel is low rank—an observation made experimentally in [8]—and a lag-window estimator may be implemented as a multiple-data-window estimator employing  $N\beta$  windows, where  $\beta$  is the normalized bandwidth over which the BTGR smoothes the periodogram estimate.

Does the argument go the other way? That is, can every multiple-data windowed spectrum estimator be realized as a lag-windowed or spectrally smoothed BTGR spectrum estimator? The answer is no because the representations of (3) and (4) allow for *any* non-negative definite (Hermitian) matrix or kernel  $\mathbf{Q}(0)$ . Only those kernels  $\mathbf{Q}(0) = \mathbf{V}^*\mathbf{V}$  that approximate Toeplitz matrices  $\mathbf{W}$  whose spectrum  $W(e^{j\theta})$  is narrowband will have BTGR representations. In fact, from the argumentation in [3] and [4], it is exactly these kernels that produce good bias-variance tradeoffs for reduction in mean-squared error. Therefore, we may say that *narrowband* lag-windowed and *good* multiple-data-windowed spectrum estimators are roughly equivalent, allowing for the exception that a bad multiple-window spectrum estimator may have no lag-windowed representation. This result shows that there is no fundamental difference between lag-windowing a correlation sequence and *multiple* windowing a data sequence when the objective is to reduce the mean-squared error of a spectrum estimator.

There is a slight computational advantage to the multiple window technique when the effective rank of the lag-window matrix is very small ( $<6$ ). This saving comes about from the fact that  $N$ -point FFT’s can be used for the multiple window procedure, whereas the lag-windowed estimation requires  $2N$ -point FFT’s to estimate the correlation sequence. The rank of the commonly used window sequences tends to meet this requirement.

Finally, mean squared error computations are somewhat easier to carry out for multiple windowed spectrum estimators than for lag-windowed estimators (see, for example, [4]). Therefore, one way to compute performance bounds for BTGR spectrograms is to analyze the equivalent low-rank Thomson estimator.

## REFERENCES

- [1] R. B. Blackman and J. W. Tukey, “The measurement of power spectra from the point of view of communications engineering,” *Bell Syst. Tech. J.*, vol. 33, pp. 185–282; 485–569, 1958.
- [2] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*. New York: Wiley, 1957.
- [3] D. J. Thomson, “Spectrum estimation and harmonic analysis,” *Proc. IEEE*, vol. 70, pp. 1055–1096, 1982.
- [4] C. T. Mullis and L. L. Scharf, “Quadratic estimators of the power spectrum,” in *Advances in Spectrum Analysis and Array Processing: Volume I*. Englewood Cliffs, NJ: Prentice-Hall, 1991, pp. 1–57.
- [5] D. Slepian, “Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: The discrete case,” *Bell Syst. Tech. J.*, vol. 57, pp. 1371–1430, 1978.
- [6] P. J. Daniell, “Discussion on the symposium of autocorrelation in time series,” *J. Roy. Stat. Soc.*, vol. 8, pp. 88–90, 1946.
- [7] R. Horn and C. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1985.
- [8] A. T. Walden, “Estimated cross spectrum matrices and their inverses,” submitted for publication.

## A Nonparametric Phase Estimation Method for SIMO Systems Based on Second-Order and Higher Order Statistics

Zhen Mao and Zhi Ding

**Abstract**—In this correspondence, we present a nonparametric phase estimation algorithm for linear single-input multiple-output (SIMO) channels. Given an unknown stationary input signal with known statistics, our approach is to obtain the joint minimum mean square phase estimation based on the polyspectra and the cross-spectra of the SIMO channel outputs. By utilizing both higher order and second-order statistics of the channel outputs, our approach is shown to be more accurate and reliable than methods based on higher order statistics alone. It can be applied to SIMO channels with common zeros.

## I. INTRODUCTION

The problem of channel identification for linear single-input multiple-output (SIMO) linear systems is often encountered in communications, surveillance, and geophysical signal processing. In blind SIMO system identification, both the channel input and the channel response are unknown. Channel identification must rely exclusively on the channel output and known statistics of channel input. Once the SIMO system is estimated, its input signal can be extracted by deconvolution.

For single-input single-output systems, phase information of the channel frequency response is contained in the higher order statistics of channel output signals. Existing nonparametric algorithms typically utilize the polyspectra of channel outputs for phase recovery of nonminimum-phase systems. In [3], a simple nonparametric FFT method was presented for channel phase recovery from polyspectra.

Manuscript received December 1, 1995; revised September 1, 1998. This work was supported in part by the National Science Foundation and by the US Army Research Office under Grant DAAH04-G4-G-0252. The associate editor coordinating the review of this paper and approving it for publication was Prof. Kon Max Wong.

The authors are with the Department of Electrical Engineering, Auburn University, Auburn, AL 36849 USA.

Publisher Item Identifier S 1053-587X(99)01347-1.

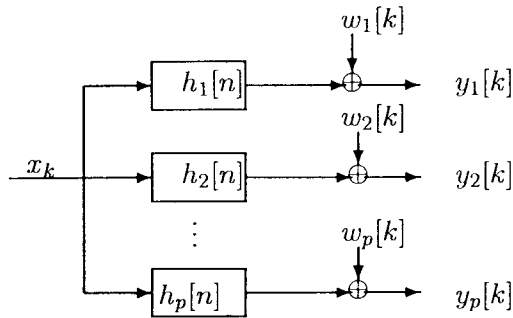


Fig. 1. Equivalent discrete SIMO system model.

Unfortunately, polyspectra tend to be unreliable when available data length is short. Moreover, errors in polyspectral phase unwrapping can be particularly exacerbating.

Fortunately, for SIMO systems, additional statistical information can be utilized. In particular, second-order statistics of SIMO system outputs contain additional phase information that can be used in the phase estimation of channel frequency response. Second-order cross-spectra of channel outputs are easier to estimate and tend to be more accurate for limited data length. In fact, a SIMO channel estimation based only on cross-spectra has been presented in [9]. Thus, an algorithm utilizing information from higher order statistics and second-order cross-spectra is expected to generate better performance than strictly higher order or second-order statistical algorithms.

In this correspondence, we will present a new nonparametric method for the phase identification of SIMO channels based on polyspectra and cross-spectra of the output signals. Our approach is derived to minimize the joint mean square phase error between the measured and the estimated spectral phases. As will become evident later, it is a simple linear approach that can be implemented through FFT directly.

## II. PROBLEM DESCRIPTION

The SIMO system model can be used for communication systems with multiple antenna elements. Given sufficient channel bandwidth, the analog channel output in quadrature amplitude modulation (QAM) systems can be sampled at higher than the baud rate to yield an equivalent SIMO linear system model [5], [6]. In geophysical exploration, multiple sensors can also be used to record multiple output signals driven by a common input.

Let  $x_k$  denote the i.i.d. non-Gaussian input of a system as shown in Fig. 1. Assuming that there are  $p$  subchannels in the SIMO system with impulse responses  $h_i[k]$ , their outputs  $y_i[k]$  can be written as

$$y_i[k] = \sum_{n=-\infty}^{\infty} x_k h_i[k-n] + w_i[k], \quad i = 1, 2, \dots, p \quad (2.1)$$

where  $w_i[k]$  are channel noises that are stationary, Gaussian, and independent of the channel input  $x_k$ . Both  $x_k$  and  $w_i[k]$  are white and zero mean with variance  $\sigma_x^2$  and  $\sigma_w^2$ , respectively.

Our goal is to identify the unknown linear system from output signals  $\{y_i[k]\}$ . Commonly known as blind channel identification, this task can rely on statistical knowledge of the channel input sequence as well as the measurable output signal. Note that if the subchannels are FIR of known order, "deterministic" algorithms can also be applied [4]. Here, we assume no prior knowledge on the model of the subchannels and use the statistical approach.

Based on the SIMO equivalent model, its second-order output statistics can be represented by the power spectra and cross spectra

of subsequences  $y_j[n]$ . The SIMO system is fully characterized by the magnitude and the phase of each transfer function  $H_i(\omega) = \sum_{k=-\infty}^{\infty} h_i[k]e^{jk\omega}$ . If the background noise level  $\sigma_w^2$  is known, the magnitude  $|H_i(\omega)|$  can typically be estimated from the power spectrum of  $y_i[k]$

$$P_i(\omega) = \sigma_x^2 |H_i(\omega)|^2 + \sigma_w^2.$$

Thus, the key obstacle in blind channel identification is the estimation of the channel phase  $\Phi(\omega) = \angle H_i(\omega)$  from its output statistics. For a SIMO linear system with a common stationary input, channel phase information is contained in both the polyspectra and the cross-spectra of output signals. If only the second-order statistics are used, then common zeros among subchannels cannot be identified [9]. To resolve the ambiguity caused by common zeros among subchannels, higher order statistical information should be exploited jointly with second-order statistics. In the next section, a minimum mean square phase error algorithm for SIMO channel phase estimation will be presented that exploit both the second-order and the higher order statistical information of output signals.

## III. MINIMUM MSE PHASE ESTIMATION ALGORITHM

### A. SIMO Channel Phase Information

Since subchannel noises  $w_i[k]$  are independent with zero mean, it then follows that the cross spectrum between subsequences  $y_i[n]$  and  $y_j[n]$  is given by

$$S_{ij}(\omega) = \sigma_x^2 H_i(\omega) H_j^*(\omega), \quad 1 \leq i, j \leq p.$$

In addition, higher order statistics also provide additional phase information. Therefore, our algorithm should focus on the bispectra or the trispectra of subchannel outputs as they can be estimated more accurately from a fixed length data sequence [1] and [2]. However, as will become apparent in the derivation, phase information from other polyspectra can be similarly exploited either jointly or separately.

For notational simplicity, we only consider the case of  $p = 2$ . Our derivation, however, does apply to more general SIMO systems. Since  $x_k$  is an i.i.d non-Gaussian random sequence and  $w_k$  is a Gaussian noise, the trispectra of the two outputs are given [1] by

$$T_i(\omega_1, \omega_2, \omega_3) = \gamma_{4x} H_i(\omega_1) H_i(\omega_2) H_i(\omega_3) H_i^*(\omega_1 + \omega_2 + \omega_3) \quad i = 1, 2. \quad (3.1)$$

Here, we denote  $\gamma_{4x} \neq 0$  as the fourth-order cumulant of the input signal  $x_k$  at zero lag. Consequently, the trispectral phase and the subchannel phases are related via

$$\Psi_i(\omega_1, \omega_2, \omega_3) = \sum_{k=1}^3 \Phi_i(\omega_k) - \Phi_i(\omega_1 + \omega_2 + \omega_3), \quad i = 1, 2. \quad (3.2)$$

Similarly, the cross-spectral phase  $\Psi_{12}(\omega)$  satisfies

$$\Psi_{12}(\omega) = \Phi_1(\omega) - \Phi_2(\omega). \quad (3.3)$$

Based on the above measurable phase information from trispectra and cross-spectra, subchannel phase functions  $\hat{\Phi}_1(\omega)$  and  $\hat{\Phi}_2(\omega)$  are to be estimated. Our phase identification is based on minimizing the mean square error between estimated spectral phases and measured spectral phases. Let

$$\Delta \Psi_i(\omega_1, \omega_2, \omega_3) = \Psi_i(\omega_1, \omega_2, \omega_3) - \sum_{j=1}^3 \hat{\Phi}_i(\omega_j) + \hat{\Phi}_i(\omega_1 + \omega_2 + \omega_3), \quad i = 1, 2. \quad (3.4)$$

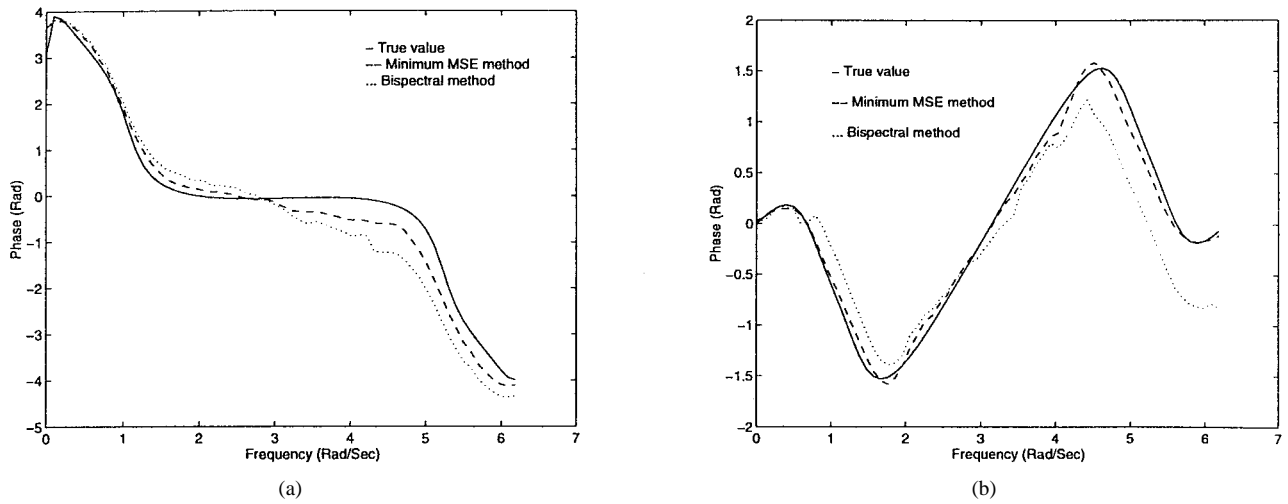


Fig. 2. Comparison of SIMO phase estimates.

Given the phase difference, the mean square error (MSE) of spectral phase can be defined as

$$\begin{aligned} \text{MSE} \triangleq & \frac{1}{2\pi} \int_{-\pi}^{\pi} |\Psi_{12}(\omega) - \hat{\Phi}_1(\omega) + \hat{\Phi}_2(\omega)|^2 d\omega \\ & + \frac{1}{(2\pi)^3} \int \int \int_{-\pi}^{\pi} [|\Delta\Psi_1(\omega_1, \omega_2, \omega_3)|^2 \\ & + |\Delta\Psi_2(\omega_1, \omega_2, \omega_3)|^2] d\omega_1 d\omega_2 d\omega_3. \end{aligned} \quad (3.5)$$

We shall derive a nonparametric algorithm that will estimate the SIMO phase responses to minimize the MSE.

Note that the channel noises are assumed to be independent so that they do not affect the cross-spectrum. This additional assumption is important to the accuracy of cross-spectral phase information. Our approach may not be suitable when subchannel noises are strong and correlated.

### B. Algorithm Development

Because discrete Fourier transform (DFT) is used to determine  $\Psi_1(\omega_1, \omega_2, \omega_3)$ ,  $\Psi_2(\omega_1, \omega_2, \omega_3)$ , and  $\Psi_{12}(\omega)$ , both trispectral phases are  $2\pi$ -periodic in  $\omega_1$ ,  $\omega_2$ , and  $\omega_3$ , whereas the cross-spectral phase is also  $2\pi$ -periodic in  $\omega$ . Moreover, the desired phase estimates  $\hat{\Phi}_1(\omega)$  and  $\hat{\Phi}_2(\omega)$  are also  $2\pi$ -periodic. Hence, we can define the (imaginary) cepstral sequences

$$\begin{aligned} \psi_i(n_1, n_2, n_3) = & \frac{1}{(2\pi)^3} \int \int \int_{-\pi}^{\pi} \Psi_i(\omega_1, \omega_2, \omega_3) e^{-j(n_1\omega_1 + n_2\omega_2 + n_3\omega_3)} d\omega_1 d\omega_2 d\omega_3 \\ & i = 1, 2. \end{aligned} \quad (3.6)$$

We also have

$$\psi_{12}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Psi_{12}(\omega) e^{-jn\omega} d\omega \quad (3.7)$$

$$\hat{\phi}_i(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{\Phi}_i(\omega) e^{-jn\omega} d\omega, \quad i = 1, 2. \quad (3.8)$$

Consequently, trispectrum phases and cross-spectrum phase can be written as

$$\begin{aligned} \Psi_i(\omega_1, \omega_2, \omega_3) = & \sum_{n_1} \sum_{n_2} \sum_{n_3} \psi_i(n_1, n_2, n_3) e^{jn_1\omega_1} e^{jn_2\omega_2} e^{jn_3\omega_3} \\ & i = 1, 2 \end{aligned} \quad (3.9)$$

$$\Psi_{12}(\omega) = \sum_{n=-\infty}^{\infty} \psi_{12}(n) e^{jn\omega} \quad (3.10)$$

$$\hat{\Phi}_i(\omega) = \sum_{n=-\infty}^{\infty} \hat{\phi}_i(n) e^{jn\omega}, \quad i = 1, 2. \quad (3.11)$$

By substituting (3.9)–(3.11) into (3.5) and using the orthogonality of sinusoidal harmonics, the minimum mean square error is given by

$$\begin{aligned} \text{MSE} = & \sum_{i=1}^2 \left[ \sum_{n_1} \sum_{n_2} \sum_{n_3} |\psi_i(n_1, n_2, n_3)|^2 - \sum_{n \neq 0} (|\psi_i(n, 0, 0)|^2 \right. \\ & + |\psi_i(0, n, 0)|^2 + |\psi_i(0, 0, n)|^2) + \sum_{n \neq 0} (|\psi_i(n, 0, 0) \\ & - \hat{\phi}_i(n)|^2 + |\psi_i(0, n, 0) - \hat{\phi}_i(n)|^2 + |\psi_i(0, 0, n) \\ & - \hat{\phi}_i(n)|^2) - \sum_n |\psi_i(n, n, n)|^2 + |\psi_i(0, 0, 0) \\ & - 2\hat{\phi}_i(0)|^2 + \sum_{n \neq 0} |\psi_i(n, n, n) + \hat{\phi}_i(n)|^2 \left. \right] + \sum_n |\psi_{12}(n) \\ & - \hat{\phi}_1(n) + \hat{\phi}_2(n)|^2. \end{aligned} \quad (3.12)$$

Thus, the optimum solutions  $\{\hat{\phi}_1^\dagger(n)\}$  and  $\{\hat{\phi}_2^\dagger(n)\}$  for this equation can be found as

$$\hat{\phi}_1^\dagger(0) = \frac{1}{8} \{3\psi_1(0, 0, 0) + \psi_2(0, 0, 0) + 2\psi_{12}(0)\} \quad (3.13)$$

$$\hat{\phi}_2^\dagger(0) = \frac{1}{8} \{(\psi_1(0, 0, 0) + 3\psi_2(0, 0, 0) - 2\psi_{12}(0))\} \quad (3.14)$$

$$\begin{aligned} \hat{\phi}_1^\dagger(n) = & \frac{1}{24} \{5[\psi_1(n, 0, 0) + \psi_1(0, n, 0) + \psi_1(0, 0, n) \\ & - \psi_1(n, n, n)] + [\psi_2(n, 0, 0) + \psi_2(0, n, 0) + \psi_2(0, 0, n) \\ & - \psi_2(n, n, n)] + 4\psi_{12}(n)\} \end{aligned} \quad (3.15)$$

$$\begin{aligned} \hat{\phi}_2^\dagger(n) = & \frac{1}{24} \{[\psi_1(n, 0, 0) + \psi_1(0, n, 0) + \psi_1(0, 0, n)] \\ & - \psi_1(n, n, n)] + 5[\psi_2(n, 0, 0) + \psi_2(0, n, 0) + \psi_2(0, 0, n) \\ & - \psi_2(n, n, n)] - 4\psi_{12}(n)\}. \end{aligned} \quad (3.16)$$

Similarly, if the bispectra of the channel outputs are not zero, then the channel phases can be estimated jointly from the phases of cross-spectra and bispectra. In particular, the phases of the bi-spectra are related to the phase of the subchannels via

$$\Psi_i(\omega_1, \omega_2) = \sum_{k=1}^2 \Phi_i(\omega_k) - \Phi_i(\omega_1 + \omega_2), \quad i = 1, 2.$$

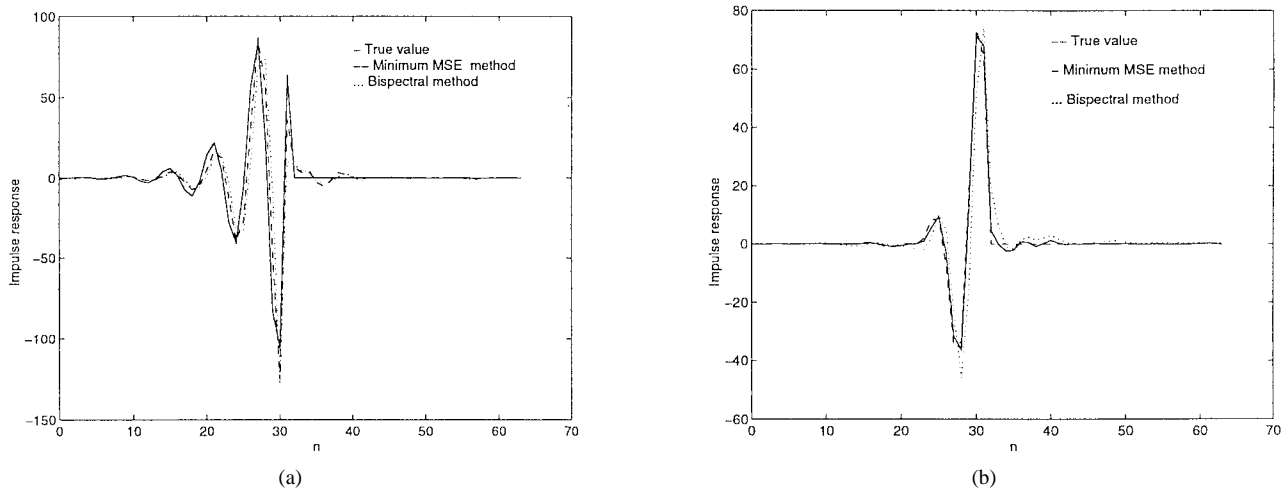


Fig. 3. Comparison of SIMO impulse response estimates.

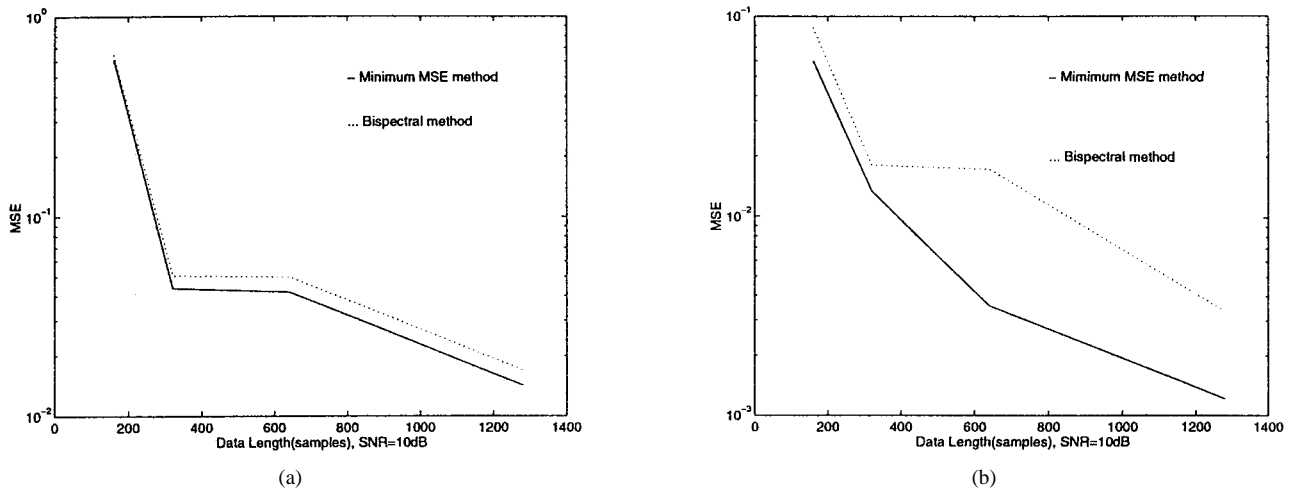


Fig. 4. Comparison of MSE in SIMO impulse response estimates for different length.

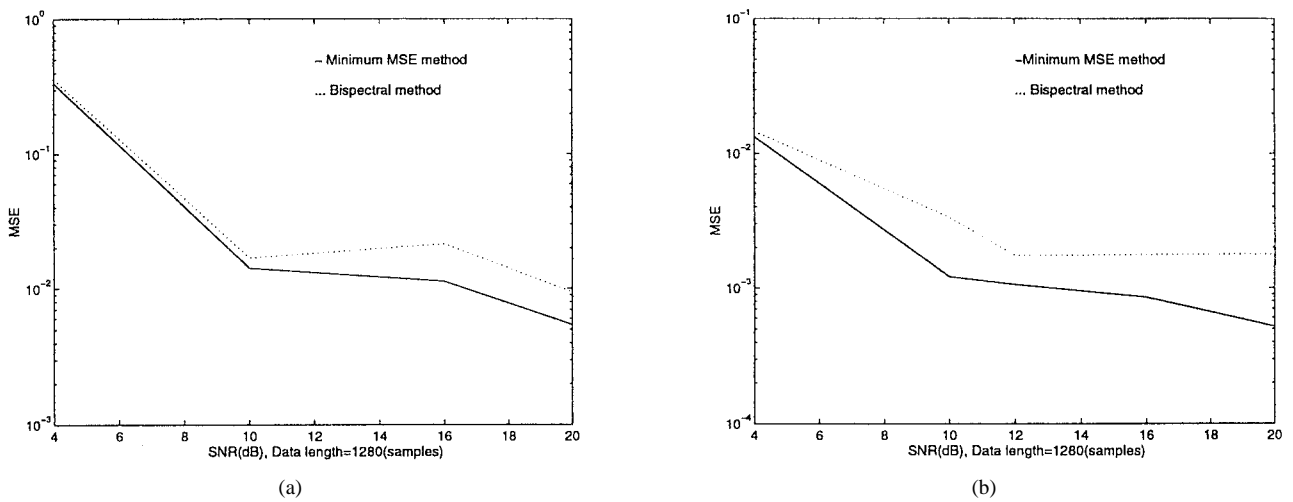


Fig. 5. Comparison of MSE in SIMO impulse response estimates at different SNR.



Define bi-cepstra as

$$\psi_i(n_1, n_2) = \frac{1}{(2\pi)^3} \int \int_{-\pi}^{\pi} \Psi_i(\omega_1, \omega_2) e^{-j(n_1\omega_1 + n_2\omega_2)} d\omega_1 d\omega_2 \quad i = 1, 2. \quad (3.17)$$

The spectral MSE can be determined as

$$\begin{aligned} \text{MSE} = & \sum_{i=1}^2 \left[ \sum_{n_1} \sum_{n_2} |\psi_i(n_1, n_2)|^2 - \sum_n (|\psi_i(n, 0)|^2 + |\psi_i(0, n)|^2 \right. \\ & + |\psi_i(n, n)|^2) + 2|\psi_i(0, 0)|^2 + \sum_{n \neq 0} (|\psi_i(n, 0) - \hat{\phi}_i(n)|^2 \\ & + |\psi_i(0, n) - \hat{\phi}_i(n)|^2 + |\psi_i(n, n) + \hat{\phi}_i(n)|^2) + |\psi_i(0, 0) \\ & \left. - \hat{\phi}_i(0)|^2 \right] + \sum_n |\psi_{12}(n) - \hat{\phi}_1(n) + \hat{\phi}_2(n)|^2. \quad (3.18) \end{aligned}$$

Thus, the optimum solutions  $\{\hat{\phi}_1^\dagger(n)\}$  and  $\{\hat{\phi}_2^\dagger(n)\}$  that minimizes the MSE can be obtained from

$$\hat{\phi}_1^\dagger(0) = \frac{1}{3} \{2\psi_1(0, 0) + \psi_2(0, 0) + \psi_{12}(0)\} \quad (3.19)$$

$$\hat{\phi}_2^\dagger(0) = \frac{1}{3} \{(\psi_1(0, 0) + 2\psi_2(0, 0) - \psi_{12}(0))\} \quad (3.20)$$

$$\begin{aligned} \hat{\phi}_1^\dagger(n) = & \frac{1}{15} \{4[\psi_1(n, 0) + \psi_1(0, n) - \psi_1(n, n)] \\ & + [\psi_2(0, n) + \psi_2(n, 0) - \psi_2(n, n)] + 3\psi_{12}(n)\} \quad (3.21) \end{aligned}$$

$$\begin{aligned} \hat{\phi}_2^\dagger(n) = & \frac{1}{15} \{4[\psi_2(n, 0) + \psi_2(0, n) - \psi_2(n, n)] \\ & + [\psi_1(0, n) + \psi_1(n, 0) - \psi_1(n, n)] - 3\psi_{12}(n)\}. \quad (3.22) \end{aligned}$$

These phase estimates can be straightforwardly extended for  $p$  subsystems. The MSE and a corresponding algorithm can be similarly defined and derived when other polyspectral information is to be exploited either jointly or separately. Because the algorithm is nonparametric, it does not require any channel model or order estimation. It is therefore simple and robust. Moreover, the DFT steps can be easily and efficiently implemented using FFT.

#### IV. SIMULATION RESULTS

To demonstrate the performance of this minimum MSE method, we present an example of SIMO identification based on bispectra and cross-spectra. In this example, a two subchannel system ( $p = 2$ ) is chosen, and the two subchannels are characterized by

$$H_1(z) = \frac{1 - 2.95z^{-1} + 1.9z^{-2}}{1 - 1.3z^{-1} + 1.05z^{-2} - 0.325z^{-3}} \quad (4.1)$$

$$H_2(z) = \frac{(1 - 0.2z^{-1} + 0.04z^{-2})}{(1 - 1.352z^{-1} + 1.338z^{-2} - 0.662z^{-3} + 0.24z^{-4})} \quad (4.2)$$

respectively. The input signal  $x_k$  is an independent random sequence with density function

$$f(x) = e^{-(x+1)} u(x+1).$$

The noise is i.i.d. Gaussian, and the overall signal-to-noise ratio is set at 12 dB. A total of  $64 \times 20$  input data samples are used for spectral estimation.

In our simulation, channel output data are used to estimate the bispectra and cross-spectra using  $5 \times 5$  and  $5 \times 1$  smoothing windows, respectively. The spectral phases are unwrapped under the continuity assumption. From phases of estimated bispectra and cross-spectrum, the two subchannel phases are estimated using the minimum MSE method. The results are used to compare with a nonparametric minimum MSE algorithm based on polyspectra alone

[3]. The phase estimates of the two subchannels are shown in Fig. 2. The impulse responses of the two subchannels based on the estimated phase and (ideally) known magnitude are shown in Fig. 3.

Next, we illustrate the impact of the data length and SNR on the result of the estimation algorithm. Results are averaged over 100 Monte Carlo simulation runs. Under SNR = 10 dB, we determine the MSE between the true channel impulse response and the estimated channel impulse response. The comparative results in Fig. 4 demonstrate the asymptotic improvement of the new method for larger number of data samples. Last, we illustrate the effect of the channel signal-to-noise ratio on the estimation result. The results are shown in Fig. 5.

#### V. CONCLUSION

In this correspondence, we presented a new nonparametric method for the phase identification of SIMO channels based on the polyspectra and the cross-spectra of subchannel outputs. Our approach is based on the minimization of a joint mean square phase error between the measured and the estimated phase of polyspectra and cross-spectra. Our method is linear and can be implemented using FFT. It is readily extendable to other spectral information sources as they become available.

#### REFERENCES

- [1] D. R. Brillinger, "An introduction to polyspectra," *Ann. Math. Stat.*, vol. 36, pp. 1351–1374, 1965.
- [2] D. R. Brillinger and M. Rosenblatt, "Computation and interpretation of  $K$ th order spectra," *Spectral Analysis of Time Series*, B. Harris, Ed. New York: Wiley, 1967.
- [3] Y. Li and Z. Ding, "A new nonparametric method for linear system phase recovery from bispectrum," *IEEE Trans. Circuits Syst. II*, vol. 41, pp. 415–419, 1994.
- [4] G. Xu, H. Liu, and L. Tong, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing*, vol. 43, pp. 2982–2993, Dec. 1995.
- [5] S. D. Halford and G. B. Giannakis, "Channel order determination based on sample cyclic correlations," in *Proc. 28th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, 1994, pp. 425–429.
- [6] M. I. Gürelli and C. L. Nikias, "Blind identification algorithms for co-channel systems using higher-order statistics" in *Proc. 28th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, 1994, pp. 435–439.
- [7] J. M. Mendel, "Use of higher-order statistics in signal processing and system theory: An update," *Proc. SPIE*, vol. 975, pp. 126–144, 1988.
- [8] J. M. Tribolt, "A new phase unwrapping algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 25, pp. 170–177, 1982.
- [9] H. Pozidis and A. P. Petropulu, "Cross-spectrum based blind identification," *IEEE Trans. Signal Processing*, vol. 45, pp. 2977–2992, Dec. 1997.

## Exploiting Input Cyclostationarity for Blind Channel Identification in OFDM Systems

Robert W. Heath, Jr. and Georgios B. Giannakis

**Abstract**— Transmitter-induced cyclostationarity has been explored recently as an alternative to fractional sampling and antenna array methods for blind identification of FIR communication channels. An interesting application of these ideas is in OFDM systems, which induce cyclostationarity due to the cyclic prefix. In this correspondence, we develop a novel subspace approach for blind channel identification using cyclic correlations at the OFDM receiver. Even channels with equispaced unit circle zeros are identifiable in the presence of any nonzero length cyclic prefix with adequate block length. Simulations of the proposed channel estimator along with its performance in OFDM systems combined with impulse response shortening and Reed–Solomon coding are presented.

### I. INTRODUCTION

Recently, there has been interest in filter bank precoding for communication systems impaired by frequency-selective fading channels [5], [14]. Although similar forms of precoding have been considered in the past [7], current interest focuses on the introduction of discrete-time cyclostationarity by the multirate precoder to enable blind channel identification at the receiver. Transmitter-induced cyclostationarity, which can be introduced either with filter banks [5], repetition coding [14], or modulation [13], allows for blind identification of arbitrary FIR multipath channels without zero-restrictions as in blind fractional sampling methods [4]. One application of transmitter-induced cyclostationarity for blind channel identification is orthogonal frequency division multiplexing (OFDM) systems that can be considered a special case of the precoding structure presented in [5].

OFDM has found renewed interest in recent years due to applications such as digital terrestrial TV [12], indoor wireless networks [2], and mobile communications [11], which are all systems that require communication in severe multipath channels. By inserting a cyclic prefix before each transmitted block longer than the order of the channel, OFDM effectively turns a frequency-selective channel into a flat-fading channel. This allows, for simple, one-tap vector equalization at the expense of a loss of 10–25% in efficiency due to the extra symbols required by the cyclic prefix [17] as well as increased sensitivity to timing, frequency synchronization [15], and transmission nonlinearities accentuated by the nonconstant modulus of OFDM signals. Channel estimation in these systems is performed by inserting training data in either the time or the frequency domain, further decreasing the efficiency. Alternatively, differential encoding may be used in place of equalization if a constant modulus constellation (say, QPSK) is transmitted at the expense of less bandwidth efficiency. Unfortunately, if the channel

Manuscript received August 14, 1997; revised June 23, 1998. This work was supported by the Office of Naval Research under Grant N00014-95-1-0908. The associate editor coordinating the review of this paper and approving it for publication was Prof. James A. Bucklew.

R. W. Heath, Jr. is with the Information Systems Laboratory, Stanford University, Stanford, CA 94305-9510 USA (e-mail: rheath@stanford.edu).

G. B. Giannakis was with the Department of Electrical Engineering, University of Virginia, Charlottesville, VA 22903-2442 USA. He is now with the Department of Electrical Engineering, University of Minnesota, Minneapolis, MN 55455 USA.

Publisher Item Identifier S 1053-587X(99)01348-3.

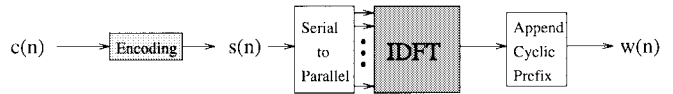


Fig. 1. OFDM transmitter.

impulse response is longer than the cyclic prefix, interference occurs, and the simple equalization property is lost.

In this correspondence, we use the cyclostationarity induced by the cyclic prefix in the OFDM system to develop an algorithm for blind channel estimation. The subspace approach of [13] is generalized here to multirate precoders and is proposed for OFDM systems as an alternative to the nonlinear matching approach of [5]. The blind nature of this estimation allows more data to be used for information transfer or for coding. As shown in [6], this approach is robust to the presence of stationary noise and channel overestimation error and does not require the cyclic prefix to be longer than the channel memory. Since we can estimate channels regardless of the cyclic prefix and without training data, the conclusion in [17] that channel coding is better than subchannel equalization must be re-evaluated. A blind algorithm for equalization was considered in [3], which uses a property of the digital-to-analog converter at the receiver and does not employ a cyclic prefix. With the absence of a cyclic prefix, however, this approach can only result in partial elimination of the intersymbol interference caused by the channel.

This correspondence is organized as follows. In the next section, we provide an overview of the OFDM system. Section III then considers the problem of blind channel identification using the cyclostationarity inherent in the OFDM transmitter. We present simulations of the proposed algorithms in Section IV.

### II. THE OFDM SYSTEM

Consider the OFDM transmitter in Fig. 1. The OFDM modulator takes the  $M$ -point IDFT of a block of  $M$  input symbols from the coder and appends a sequence of  $L < M$  symbols to the beginning of each block (see Fig. 2). In this way, the OFDM modulator can be viewed as a rate  $M/(M+L)$  block code operating in the real field. To describe the input/output relationship at various points, let  $P = M + L$ , and adopt the polyphase notations  $s_m(n) = s(nM + m)$ ,  $w_p(n) = w(nP + p)$ ,  $x_p(n) = x(nP + p)$  to denote the  $m$ th or  $p$ th symbol in the  $n$ th block of data at the input to the modulator, output of the modulator, and at the output of the channel, respectively (see Figs. 1–3). Then, the sequence to be transmitted is

$$w_p(n) = \sum_{m=0}^{M-1} s_m(n) e^{j \frac{2\pi}{M} m(p-L)}, \quad p = 0, \dots, P-1 \quad (1)$$

with the  $\exp(j2\pi m(-L)/M)$  accounting for the cyclic prefix, which is a repetition of the last  $L$  frequency domain symbols as in Fig. 4. During transmission,  $w_p(n)$  is pulse-shaped with  $g_c^{(tr)}(t)$ , propagates through an unknown frequency selective channel  $g_c^{(ch)}(t)$ , is degraded by additive white Gaussian noise (AWGN)  $\nu_c(t)$ , and is filtered by  $g_c^{(rec)}(t)$  on reception. With  $\star$  denoting linear convolution, let  $g_c(t) = g_c^{(tr)}(t) \star g_c^{(ch)}(t) \star g_c^{(rec)}(t)$  and  $h(n) = g_c(nT + \epsilon)$  denote the order  $L_h$  composite discrete-time channel, and let  $v(n) = g_c^{(rec)}(t) \star \nu_c(t)|_{t=nT+\epsilon}$  denote the received AWGN (assuming Nyquist pulse shaping). Note that the channel model explicitly incorporates the symbol timing error  $\epsilon \in [0, T)$ . Assuming that the block size

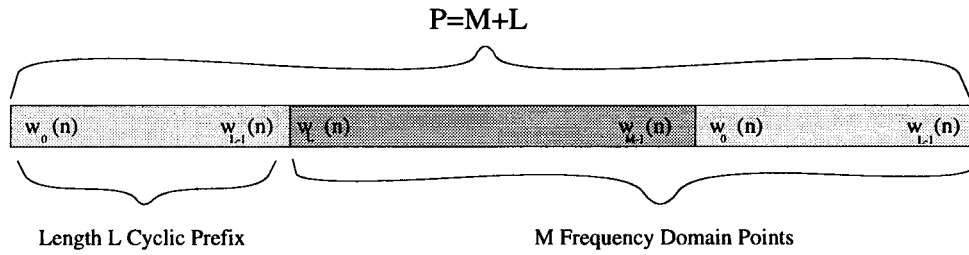


Fig. 2. OFDM frame structure.

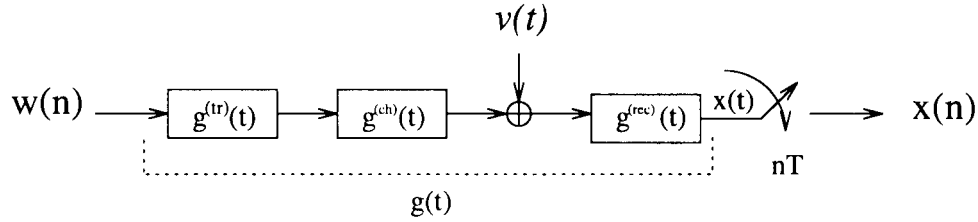


Fig. 3. Baseband equivalent transmission channel.

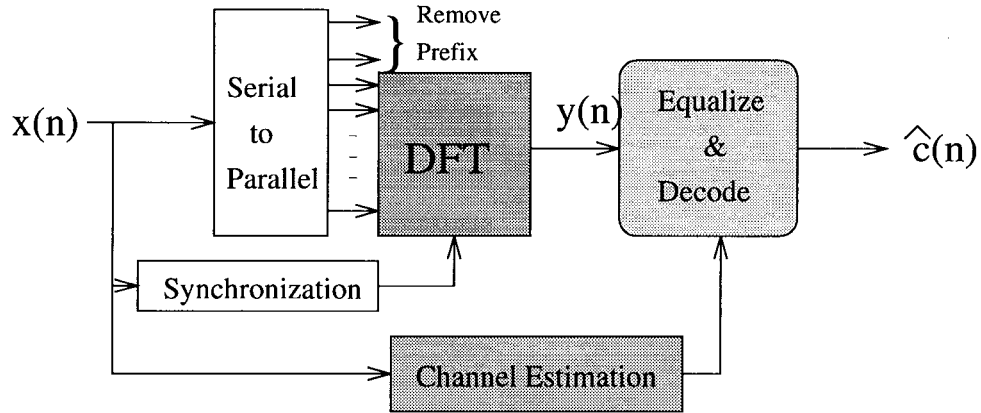


Fig. 4. OFDM receiver.

$M$  is greater than  $L_h$ , only two blocks can overlap due to ISI. We use polyphase notation to separate the channel output  $x(n) = \sum_{l=0}^{L_h} h(l)w(n-l) + v(n)$  into one part that depends on the present symbols  $s_m(n)$  and one part that depends on the past symbols  $s_m(n-1)$ ,  $m \in [0, M-1]$ . Then, the input to the OFDM demodulator for  $p \in [0, P-1]$  is

$$\begin{aligned}
 x_p(n) &= \sum_{m=0}^{M-1} s_m(n) e^{j\frac{2\pi}{M}m(p-L)} \sum_{l=0}^{L_h} h(l) e^{-j\frac{2\pi}{M}ml} \\
 &\times \sum_{r=0}^{P-1} \delta(r - (p-l)) + \sum_{m=0}^{M-1} s_m(n-1) e^{j\frac{2\pi}{M}m(p-L)} \\
 &\times \sum_{l=0}^{L_h} h(l) e^{-j\frac{2\pi}{M}m(l-P)} \delta(r - (p-l+P)) + v_p(n) \quad (2)
 \end{aligned}$$

where the summation over deltas constrains the ranges of  $(p-l)$  and  $(p-l+P)$  to lie in  $[0, P-1]$ . Assuming correction for frequency offset [15], the demodulator then removes the first  $L$  symbols corresponding to the cyclic prefix and takes the  $M$ -point DFT to obtain  $y_k(n) = \frac{1}{M} \sum_{p=0}^{M-1} x_{p+L}(n) e^{-j\frac{2\pi}{M}pk}$ , which, in general,

equals

$$\begin{aligned}
 y_k(n) &= \frac{1}{M} \sum_{m=0}^{M-1} s_m(n) \sum_{p=0}^{M-1} e^{j\frac{2\pi}{M}p(m-k)} \sum_{l=0}^{L_h} h(l) e^{-j\frac{2\pi}{M}ml} \\
 &\times \sum_{r=0}^{P-1} \delta(r - (p+L-l)) \\
 &= \frac{1}{M} \sum_{m=0}^{M-1} s_m(n-1) \sum_{p=0}^{M-1} e^{j\frac{2\pi}{M}p(m-k)} \sum_{l=0}^{L_h} h(l) e^{-j\frac{2\pi}{M}m(l-P)} \\
 &\times \sum_{r=0}^{P-1} \delta(r - (p+L-l+P)) + v_k(n) \quad (3)
 \end{aligned}$$

where  $v_k(n)$  is the transformed noise. By choosing the cyclic prefix to be as long as the order of the channel, e.g.,  $L \geq L_h$ , the second sum in (3) becomes zero because  $(p-l+L+p) \in [P, 2P+L_h-1] \notin [0, P-1]$ . The sum over  $p$  in the first term becomes  $\delta(m-k)$  since  $(p-l+L) \in [0, P-1]$ , and thus, (3) simplifies to (in the absence of noise)

$$y_k(n) = s_k(n) H\left(\frac{2\pi}{M}k\right). \quad (4)$$

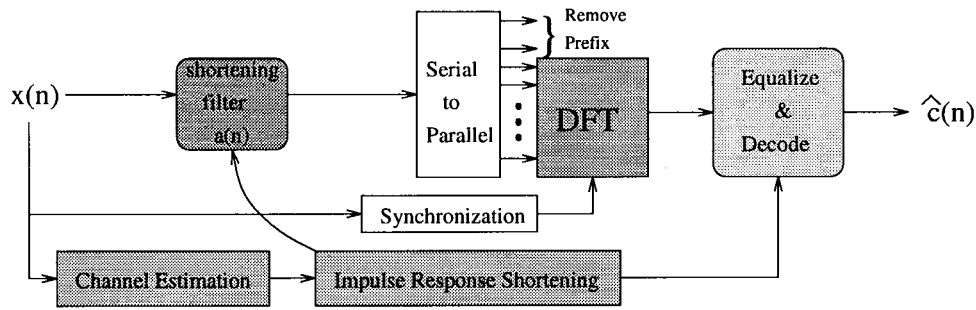
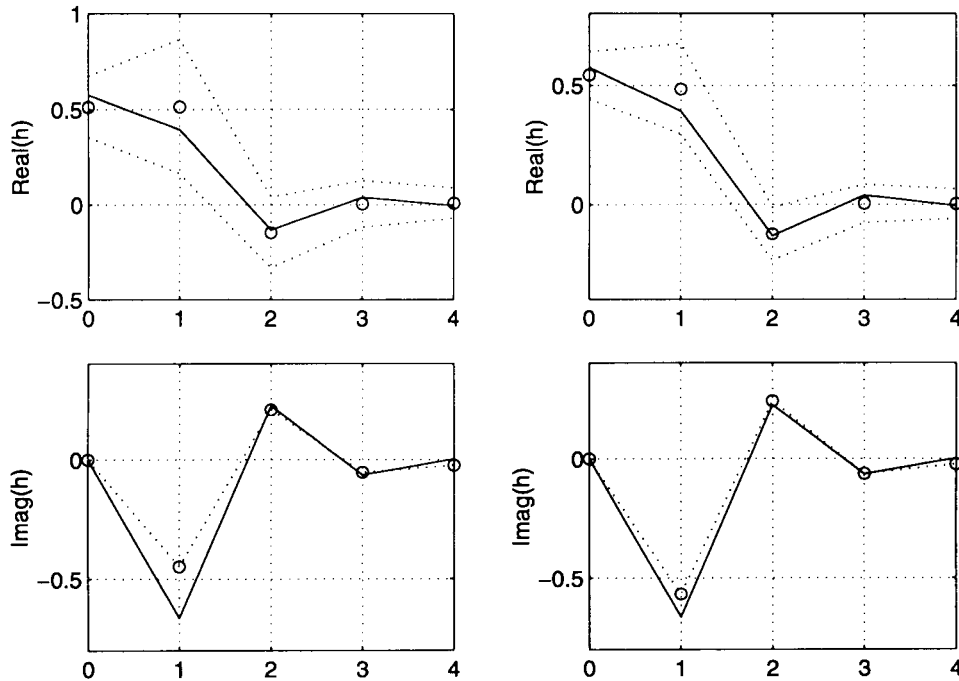


Fig. 5. OFDM receiver with impulse response shortening.

Fig. 6. Averaged channel estimates  $\pm$  standard deviation for  $P = 19$ ,  $M = 15$ ,  $I = 100$ , and  $\text{SNR} = 20$  dB.

The fact that OFDM effectively turns a frequency-selective fading channel into a flat-frequency fading channel is evident in (4). Clearly, equalization amounts to correcting  $y_k(n)$  for a phase and a scale factor for each  $k$ . In fact, for PSK constellations incorporated with subchannel differential encoding, channel estimation is unnecessary. Unfortunately, many applications use nonconstant modulus symbol sets to improve transmission efficiency, thus requiring the estimation of  $\{H(2\pi k/M)\}_{k=0}^{M-1}$ .

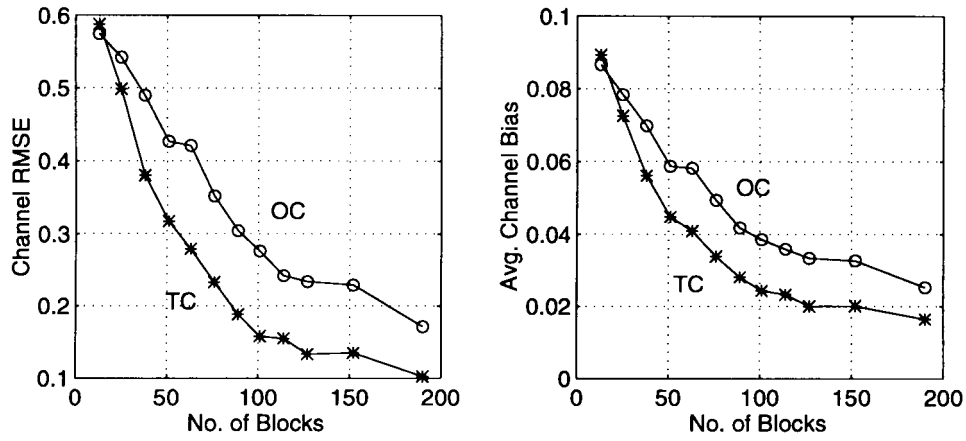
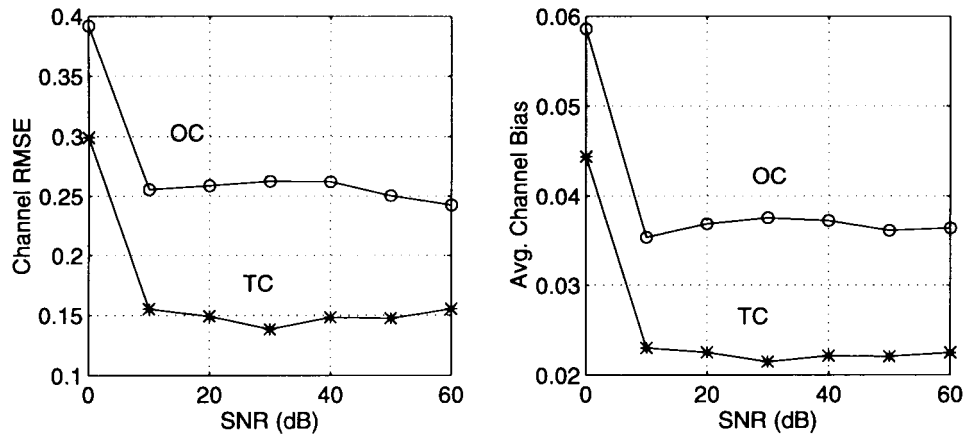
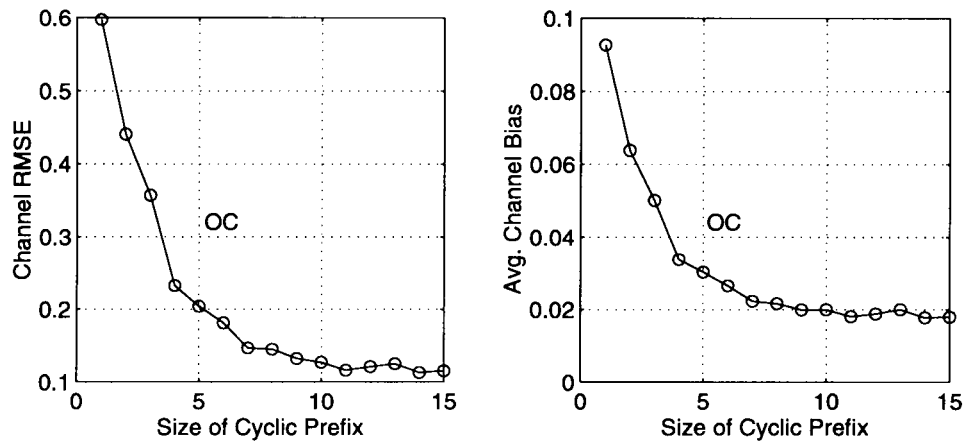
The ease in equalization due to (4) is not without drawbacks. The presence of the DFT at the receiver places stricter requirements on synchronization [15]. Additionally, (4) exhibits reduced performance in spectral nulls. In wireline transmission schemes such as discrete multitone transmission (DMT), the channel knowledge is used by the transmitter for adaptive loading and/or power control [1]. In broadcast and wireless OFDM, this problem is resolved by using coding. A variety of schemes for coding have been proposed, typically employing some sort of convolutional coding, interleaving, and concatenation [8], for the purpose of correcting the resulting bursty errors.

Since the length of the cyclic prefix is chosen *a priori*, interference occurs as shown in (3) due to the coefficients of  $h(l)$ , for  $l > L$ . This

creates problems in channel estimation that relies on the structure in (4) and results in symbol errors caused by the ISI term. This problem can be eliminated by employing decision feedback [16] or impulse response shortening [9]. Since decision feedback has a complexity that increases with the size of the DFT, we focus on impulse response shortening (see Fig. 5), which consists of designing a prefilter  $a(n)$  such that the composite channel  $a(n) * h(n)$  has energy concentrated in the desired duration chosen for the cyclic prefix [9]. In the next section, we develop an algorithm for blind channel estimation that does not require the channel to be shorter than the cyclic prefix. We apply impulse response shortening when the channel duration exceeds the duration of the cyclic prefix. This approach allows for a reduction of the cyclic prefix and elimination of training data so that the system can provide a higher data rate or greater reliability through increased coding.

### III. BLIND CHANNEL IDENTIFICATION

With the presence of a cyclic prefix in the OFDM transmission scheme, we are motivated to look for cyclostationarity in the output of the encoder as in [5]. Assume the symbols  $s_m(n)$  from the encoder in Fig. 1 are white and zero-mean with variance  $\sigma_s^2$ , take  $P > M$ , and,


 Fig. 7. Channel error versus number of symbols for  $M = 15$ ,  $P = 19$ ,  $I = 100$ ,  $\text{SNR} = 20$  dB.

 Fig. 8. Channel error versus SNR for  $M = 15$ ,  $P = 19$ ,  $I = 100$  120 M data.

 Fig. 9. Channel error versus size of cyclic prefix for  $M = 15$ ,  $I = 100$  120 M data (with  $P = M + L_h$  varying accordingly).

using (1), examine the time-varying correlation  $c_{ww}(nP + p; \tau) := E\{w_p(n)w_{p+\tau}^*(n)\}$  at the output of the modulator

$$c_{ww}(nP + p; \tau) = \sum_{m_1=0}^{M-1} \sum_{m_2=0}^{M-1} E\{s_{m_1}(n)s_{m_2}^*(n)\} e^{j\frac{2\pi}{M}m_1(p-L)} \times e^{-j\frac{2\pi}{M}m_2(\tau+p-L)} \sum_{r=0}^{P-1} \delta(r - (p + \tau)) \quad (5)$$

$$= \sigma_s^2 M \left[ \delta(\tau) + \delta(\tau - M) \sum_{r=0}^{P-M-1} \delta(p - r) + \delta(\tau + M) \sum_{r=M}^{P-1} \delta(p - r) \right]. \quad (6)$$

Since the right-hand side of (6) depends only on  $p$ , the output of the encoder is cyclostationary with period  $P$ . Note that if we take  $P = M$ , then (6) would instead become  $c_{ww}(nP + p; \tau) = \sigma_s^2 M \delta(\tau)$

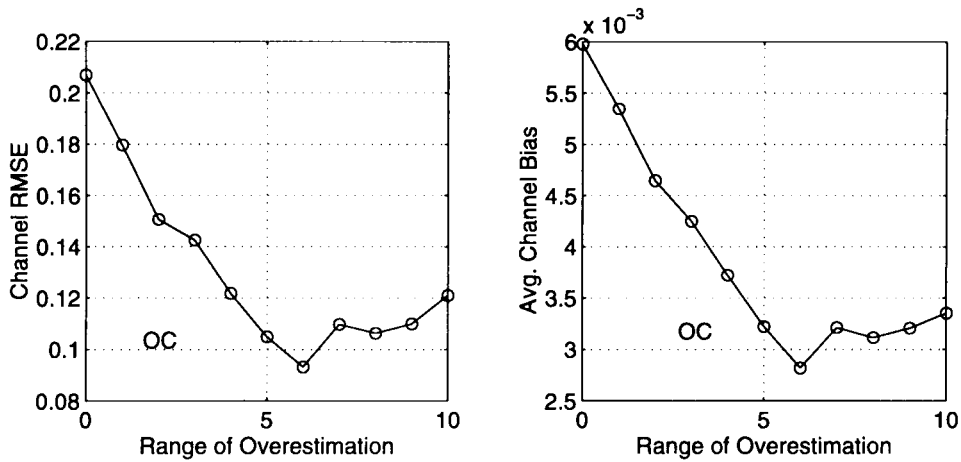


Fig. 10. Channel error versus overestimated order for  $M = 15$ ,  $I = 100$ , 120 M data (with  $P = M + L_h$  varying accordingly).

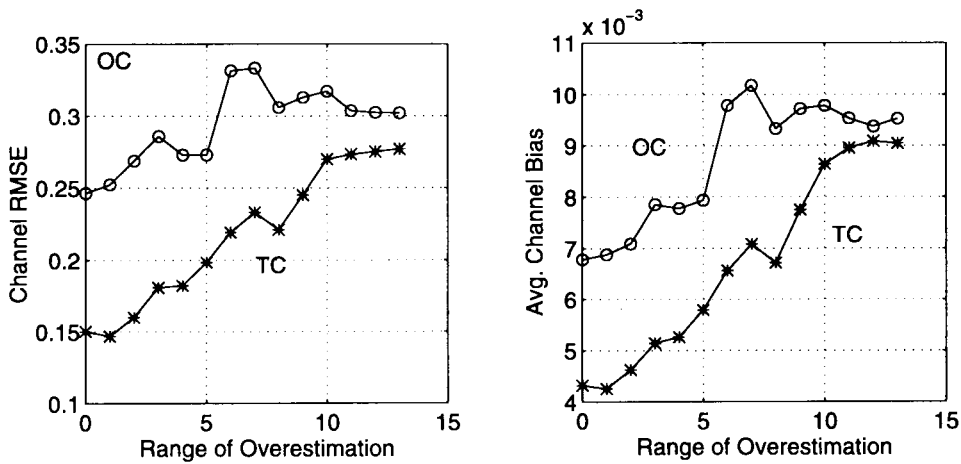


Fig. 11. Channel error versus overestimated order for  $M = 15$ ,  $P = 19$ ,  $I = 100$  120 M data.

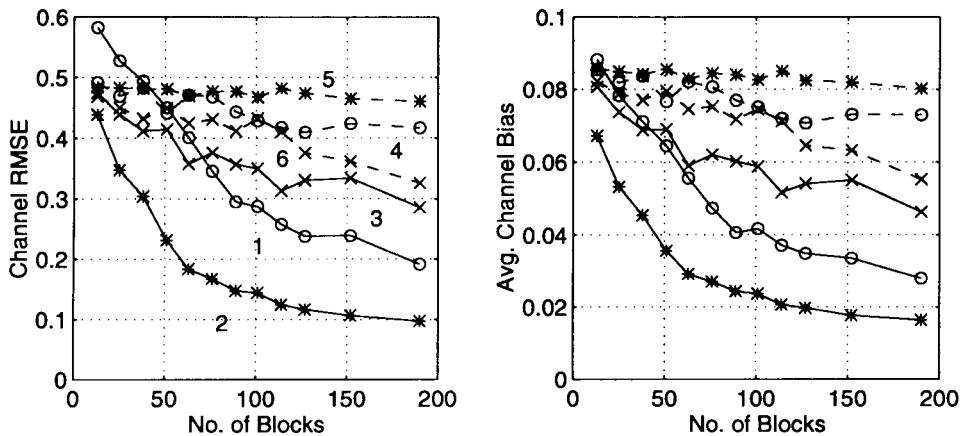


Fig. 12. Channel error versus number of blocks of data for  $M = 15$ ,  $P = 19$ ,  $I = 100$  for the one cycle approach with cycles 1...6.

because of the orthogonality of the exponentials, and periodicity is lost. Linear, time-invariant filtering does not change cyclostationarity; consequently, we expect periodicity in the time-varying correlation at the output of the channel

$$c_{xx}(nP + p; \tau) = \sum_{l=0}^{L_h} h(l)h^*(l + \tau - q)c_{ww}(n - l; q) + c_{vv}(nP + p; \tau) \quad (7)$$

which is indeed the case. If the noise is AWGN, then it has a time-varying correlation that simplifies to  $c_{vv}(nP + q; \tau) = \sigma_v^2 \delta(\tau)$  in (7). To avoid stationary noise, we consider the cyclic correlation, which is defined as the Fourier series expansion of the time-varying correlation  $C_{yy}(k; \tau) = (1/P) \sum_{p=0}^{P-1} c_{yy}(p; \tau) \exp(-j2\pi kp/P)$ . The cyclic correlation of the encoder output is  $C_{ww}(k; \tau) = (\sigma_s^2 M/P) \{ \delta(\tau)\delta(k) + [\delta(\tau + M) \exp(-j2\pi kM/P) + \delta(\tau - M)]E(k) \}$  with  $E(k) :=$

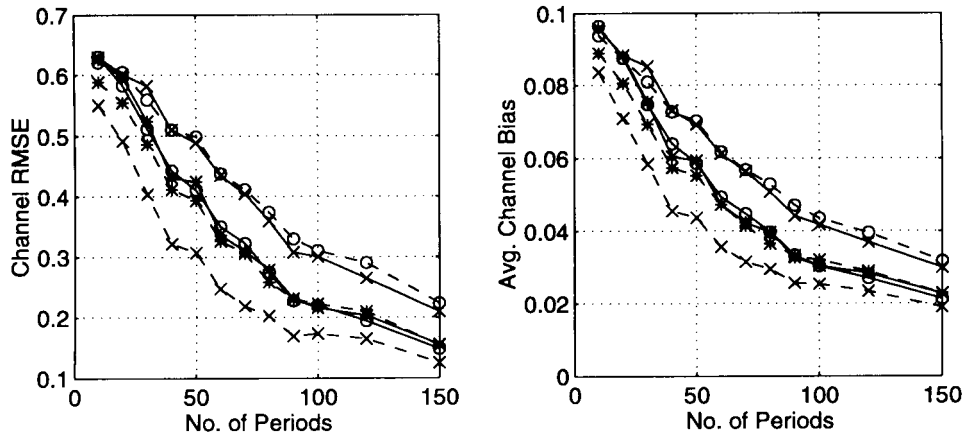


Fig. 13. Channel error versus number of blocks of data for  $M = 15$ ,  $P = 19$ ,  $I = 100$  for the two cycle approach with cycle 1 and cycles 2, ..., 7. Cycles are ordered 4, 3, 5, 1, 2, 6 from top to bottom at 120 M.

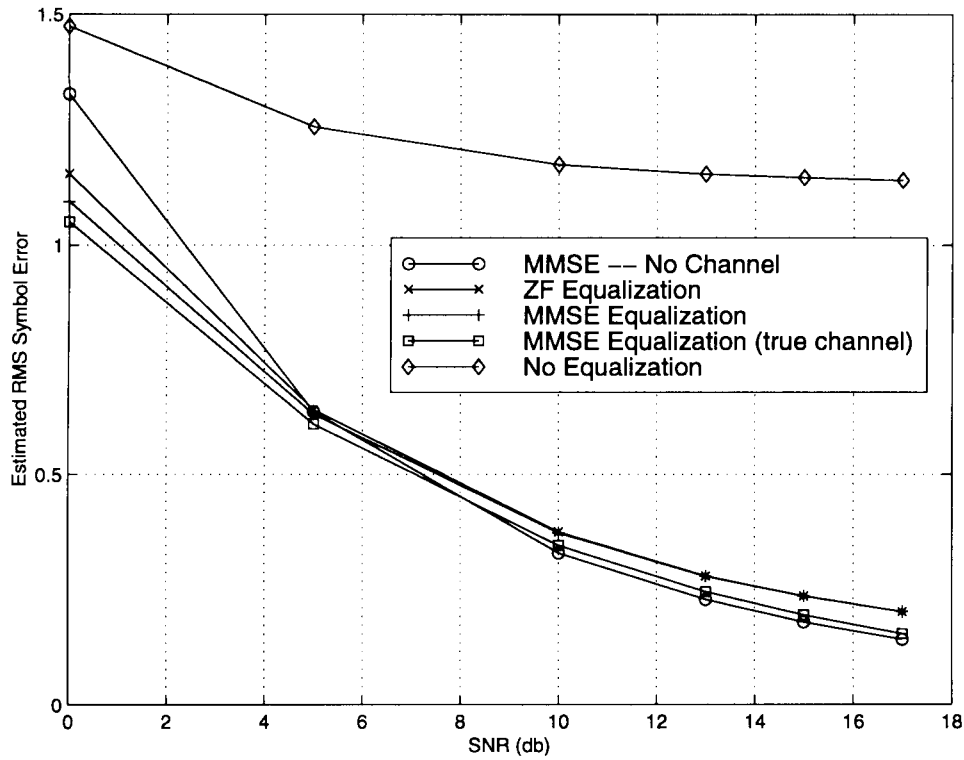


Fig. 14. RMS symbol estimation error—Adequate length cyclic prefix.

$\exp(-j\pi k(L-1)/P) \sin(\pi kL/P) / \sin(\pi k/P)$ . We write the cyclic correlation of the channel output in (7) as

$$C_{xx}(k; \tau) = \sum_{l=0}^{L_h} \sum_{q=-M}^M h(l)h^*(l+\tau-q)C_{ww}(k; q)e^{-j\frac{2\pi}{P}kl} + C_{vv}(k; \tau). \quad (8)$$

The cyclic correlation of the noise is  $C_{vv}(k; \tau) = \sigma_v^2 \delta(\tau) \delta(k)$ , which is zero for nonzero cycles  $k$ . Subsequently, we will assume that  $k \neq 0$  to avoid stationary noise.

The Z transform of the cyclic correlation with respect to  $\tau$  defines the cyclic spectrum. For a particular cycle  $k \neq 0$ , the output cyclic spectrum is  $S_{xx}(k; z) = S_{ww}(k; z) H(\exp(-j2\pi k/P)z^{-1})H^*(z^*)$ , where  $S_{ww}(k; z) = \sigma_s^2(M/P) \{\delta(k) + [z^{-M} \exp(-j2\pi kM/P) + z^M]E(k)\}$ . Interestingly, if we examine two cycles  $k_1$  and  $k_2$ , we can take the ratio of the cyclic

spectra to find

$$\frac{S_{xx}(k_1; z)S_{ww}(k_2; z)H(e^{j\frac{2\pi}{P}k_2z^{-1}})}{S_{xx}(k_2; z)S_{ww}(k_1; z)H(e^{j\frac{2\pi}{P}k_1z^{-1}})} = 1. \quad (9)$$

In the following development, we will only consider admissible cycles  $k$ , as defined in the following proposition.

*Proposition 1:* A cycle  $k$  is admissible if  $k \neq 0$ , and  $C_{ww}(k; \tau)$  is nonzero for at least one lag  $\tau$ . This occurs if  $k$  is chosen such that  $kL \bmod P \neq 0$ . The set of all admissible cycles is nonempty, e.g., choose  $L$  even and  $P$  odd.

*Proof:* If  $kL \bmod P \neq 0$ , it follows that  $\sin(\pi kL/P) \neq 0$ , which for  $k \neq 0$  implies that  $E(k) \neq 0$ , and hence,  $C_{ww}(k; \tau) \neq 0$  for at least one  $\tau$ , namely,  $\tau = M$  or  $\tau = -M$ .  $\square$

To solve for the channel, we write (9) in matrix form. With ' as transpose, let  $\mathbf{h} = [h(0) \cdots h(L_h)]'$  and  $\mathbf{D}_k(L_h) =$

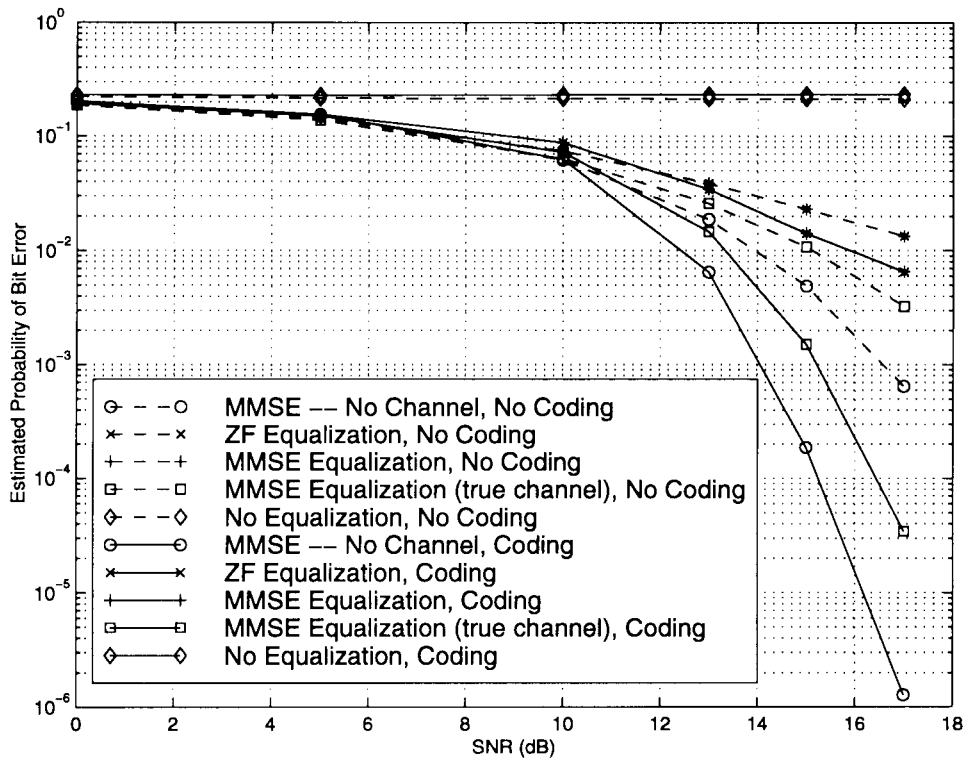


Fig. 15. Estimated probability of bit error with adequate-length cyclic prefix.

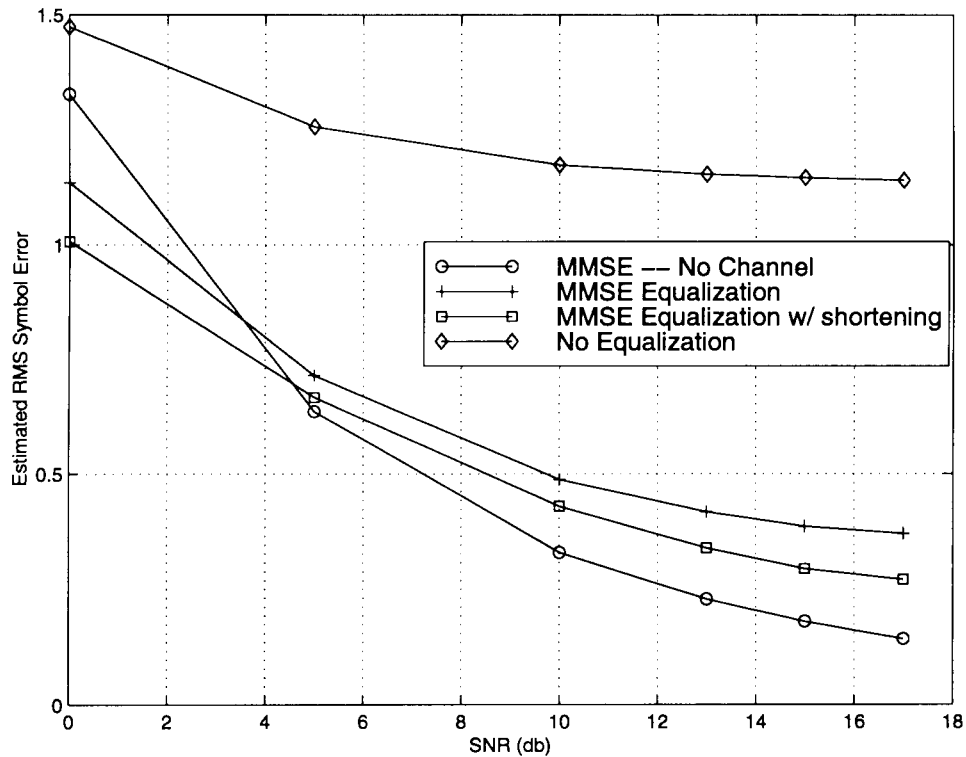


Fig. 16. RMS symbol error—Inadequate length cyclic prefix.

$\text{diag}(1, \dots, \exp(j2\pi k L_h/P))$ . Then, let  $T_k^{(w)}$  denote the  $(2M + L_h + 2) \times (L_h + 1)$  Toeplitz matrix with first column  $[C_{ww}(k; -M), \dots, C_{ww}(k; M), 0, \dots, 0]$  and first row  $[C_{ww}(k; -M), 0, \dots, 0]$ . Similarly, let  $T_k^{(x)}$  denote the  $(4M + 2L + L_h + 3) \times (2M + L_h + 2)$  matrix with first column  $[C_{xx}(k; -M - L_h),$

$\dots, C_{xx}(k; M + L_h), 0, \dots, 0]$  and first row  $[C_{xx}(k; -M - L_h), 0, \dots, 0]$ . Then, we rewrite (9) and can solve for the channel (within a scale nonidentifiable blindly) from

$$T\mathbf{h} = 0, \quad T := [T_{k_2}^{(x)} T_{k_1}^{(w)} \mathbf{D}_{k_1} - T_{k_1}^{(x)} T_{k_2}^{(w)} \mathbf{D}_{k_2}]. \quad (10)$$



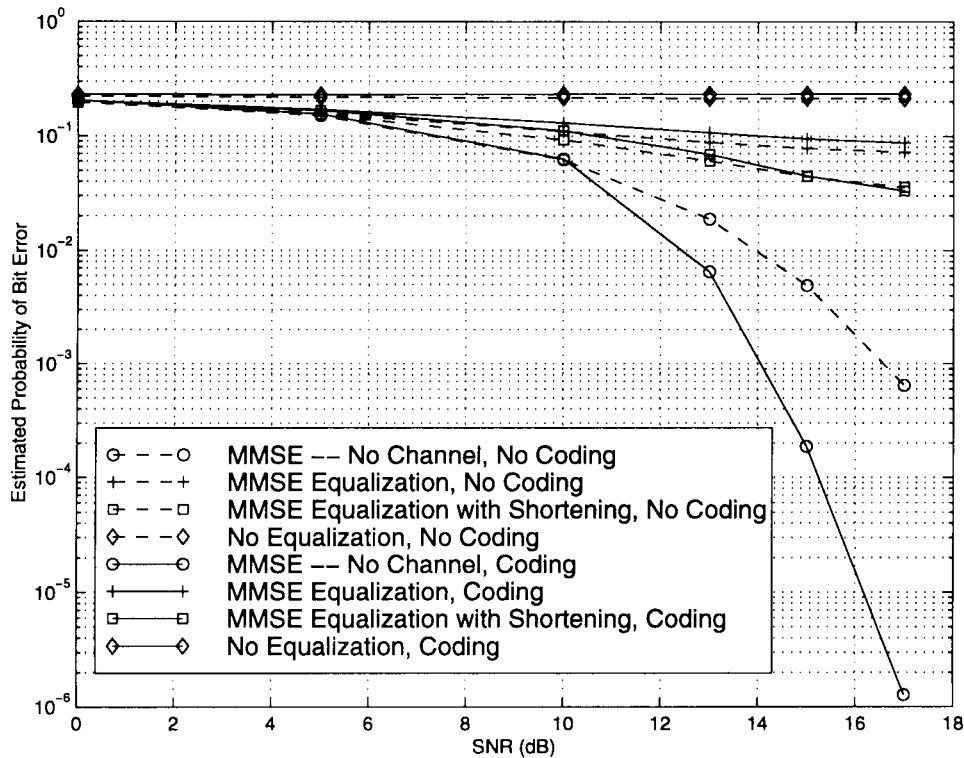


Fig. 17. Estimated probability of bit error with inadequate-length cyclic prefix.

Note that  $\mathbf{D}_k$  and  $\mathcal{T}_{k_i}^{(w)}$  are known *a priori* while the coefficients of  $\mathcal{T}_k^{(x)}$  are the cyclic correlations at the receiver, which can be consistently estimated using the sample cyclic correlation estimator  $\hat{C}_{xx}(k; \tau) = (1/N) \sum_{n=0}^{N-1} x(n)x^*(n + \tau) \exp(-j2\pi kn/P)$ . Identifiability of the channel from (10) is established in the following proposition.

*Proposition 2:* The channel  $\{h(l)\}_{l=0}^{L_h}$  is uniquely identifiable within a complex scalar from (10) if and only if  $k_1$  and  $k_2$  are both admissible, and there is no  $l \in [1, L_h]$  such that  $\exp(-j2\pi k_1 l/P) = \exp(-j2\pi k_2 l/P)$ .

*Proof:* The proof follows from the proof in [13] by noting that the product  $\mathcal{T}_{k_i}^{(x)} \mathcal{T}_{k_j}^{(w)}$  is a product of Toeplitz matrices, which is Toeplitz.  $\square$

One interesting pair of cycles is  $k$  and  $-k$ . Because  $C_{xx}(-k; \tau) = C_{xx}^*(k; -\tau) \exp(-j2\pi k\tau/P)$  in  $S_{xx}(k; z)$  and  $S_{xx}(-k; z)$ , we can solve for the channel from (10) using only one cyclic correlation, which we call the one-cycle (OC) approach. Alternatively, with two cycles, we can use a two-cycle (TC) approach and build three sets of equations like (10) to possibly increase estimation accuracy. Note that with slight modification, this channel identification formulation can be applied to any transmitter that has a cyclostationary outputs.

Given a set of admissible cycles, it is desirable to know which two cycles  $k_1$  and  $k_2$  to choose to find the channel. From the cyclic spectrum, we would like to pick the cycles such that  $S_{ww}(k_1; z)$  and  $S_{ww}(k_2; z)$  are the “most different.” a requirement that is difficult to quantify. Since  $S_{ww}(k_i; z)$  is only a function of  $C_{ww}(k_i; M)$  and  $C_{ww}(k_i; -M)$ , one option is to pick  $k_i$  such that  $|C_{ww}(k_i; M)|$  (which conveniently equals  $|C_{ww}(k_i; -M)|$ ) is maximum. From the expression for  $C_{ww}(k_i; M)$  (8), this occurs when  $k_i$  is chosen such that  $|E(k)|$  is maximized. Due to the odd symmetry of the sine function, for any  $k$  that maximizes  $|E(k)|$ ,  $-k$  will also maximize  $|E(k)|$ , and such a choice of  $k$  and  $-k$  results in the OC solution of (10). If we wish to use the TC solution, we should pick the cycles  $k_1$  and  $k_2$ , which maximizes  $|E(k)|$  subject to the constraint

that  $|k_2| \neq |k_1|$ . With this choice of cycles, we will be estimating the cyclic correlations that have the most significant energy when compared with the noise power and using these coefficients to find the channel. We use this method of cycle selection in the simulations that follow.

IV. SIMULATIONS

In this section, we examine the performance of the OC, TC channel estimates in the OFDM system. We use the root mean square error (RMSE), which is defined as  $\frac{1}{\|\mathbf{h}\|} \sqrt{\frac{1}{I(L_h+1)} \sum_{i=0}^I \|\hat{\mathbf{h}}^i - \mathbf{h}\|^2}$ , and the channel average bias  $\frac{1}{I(L_h+1)} \sum_{l=0}^{L_h} |\sum_{i=0}^I \hat{h}^i(l) - h(l)|$ , both averaged over  $I$  Monte Carlos to evaluate the channel error. To evaluate the usefulness of the channel estimator in the receiver, we employ the symbol RMSE  $\frac{1}{\sigma_s^2} \sqrt{\frac{1}{IN} \sum_{i=0}^I \sum_{j=0}^N |\hat{s}^i(l) - s(l)|^2}$  and the estimated probability of bit error.

*Experiment 1:* Here, we consider a two-ray multipath channel  $h_c(t) = e^{-j2\pi(0.15)} r_c(t - T/2, \beta) + 0.8 e^{-j2\pi(0.6)} r_c(t - 5T/4, \beta)$ , where  $r_c(t)$  is the raised cosine<sup>1</sup> with rolloff  $\beta = 0.35$  sampled at  $t = 0, T, \dots, 4T$ ; thus,  $L_h = 4$ . We used  $I = 100$ ,  $M = 15$ , and 16-QAM for modulation. In Fig. 6, with  $P = 19$ , we show the average channel estimate for SNR = 20 dB and 120 M data along with the channel error versus the number of blocks in Fig. 7. The corresponding performance for 120 M data varied over SNR is displayed in Fig. 8. From Figs. 7 and 8, we see that the estimator is consistent and that the estimates improve as the noise power decreases. In Fig. 9, for SNR = 20 dB and 120 M symbols, we consider the channel error as  $P = M + L$  and varies from  $L = 1, \dots, 15$ , whereas the channel is fixed to observe how the estimate depends on the difference  $L = P - M$ . From Fig. 9, we see that although the estimator works with  $L$  as small as 1, increasing the length of the cyclic prefix decreases the error. This is intuitively appealing because

<sup>1</sup> See [10] for information about the use of the raised-cosine pulse-shaping filter in OFDM.

we would like to trade off extra redundancy for error performance. Next, we consider the channel RMSE for SNR = 20 dB when the order  $L_h$  is overestimated and  $P = M + L_h$  in Fig. 10 and when the order is overestimated with  $P = 19$  in Fig. 11. In Fig. 10, we see the beneficial effects of having a larger prefix, whereas Fig. 11 shows the graceful degradation when the channel is overestimated.

*Experiment 2:* In this experiment, we consider the effect of the cycle chosen on the resulting channel error in estimating the two-ray channel above. Fig. 12 considers the performance of the OC approach for  $I = 100$ ,  $P = 19$ ,  $M = 15$ , SNR = 20 dB, and 120 M symbols for cycles 1 . . . 6, whereas Fig. 13 considers similarly the performance using the TC approach with cycles 1 and 2 . . . 7. Cycle selection seems to have an effect on the channel error, but asymptotic performance analysis is required to determine its precise role.

*Experiment 3:* Now, we look at the probability of bit error for an OFDM system. In Fig. 14, we plot the RMS symbol estimation error, and in Fig. 15, we plot the probability of bit error (assuming Gray coding in selection of the 16 QAM symbols) estimated over 500 Monte Carlos of 500 M data for an OFDM system with  $M = 15$  and  $P = 19$ , with and without a (15, 11) two symbol-error correcting Reed–Solomon (RS) equivalent code for the artificial channel  $h = [1, 2, 1, -1, 1]/\sqrt{8}$ . We used the standard OFDM ZF and MMSE structures [12] to equalize the  $L_h = 4$  channel above. Next, we consider the same channel and  $M = 15$  and  $P = 17$  to observe the effects of channels longer than the cyclic prefix. We estimate the channel as before but look at MMSE equalization with and without the use of impulse response shortening [9] and RS(15, 11) coding. We used an eight-tap, zero-delay shortening filter derived from the estimated channel. In Fig. 16, we plot the RMS symbol estimation error, and in Fig. 17, we plot the estimated probability of error. For comparison purposes, in Figs. 15 and 17, we plot the MMSE uncoded and coded solutions for the case when  $h(n) = \delta(n)$  as well as when there is no attempt at equalization. In Fig. 15, we see that the performance of the system using equalization with our channel estimate approaches the performance of the case where  $h(n) = \delta(n)$ . From Fig. 17, we see that impulse response shortening may be a beneficial technique when combined with our channel estimate since it reduces the error floor present in the unshortened scenario. Performance of impulse response shortening varies with the channel and may be improved by changing shortening parameters. Further improvements may be obtained using vector MMSE or vector MMSE decision feedback equalizers at the expense of further complexity [6].

#### REFERENCES

- [1] P. S. Chow, J. M. Cioffi, and J. A. C. Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Trans. Commun.*, vol. 43, pp. 773–775, Mar. 1995.
- [2] L. J. Cimini, Jr., "Performance studies for high-speed indoor wireless communications," *Wireless Pers. Commun.*, vol. 2, nos. 1–2, pp. 67–85, 1995.
- [3] M. de Courville, P. Duhamel, P. Madec, and J. Palicot, "A least mean squares blind equalization techniques for OFDM systems," *Ann. Telecommun.*, vol. 52, nos. 1–2, pp. 12–20, Jan.–Feb. 1997.
- [4] Z. Ding, "Characteristics of band-limited channels unidentifiable from second-order cyclostationary statistics," *IEEE Signal Processing Lett.*, vol. 3, pp. 150–152, May 1996.
- [5] G. B. Giannakis, "Filterbanks for blind channel identification and equalization," *IEEE Signal Processing Lett.*, vol. 4, pp. 184–187, June 1997.
- [6] R. W. Heath, Jr., "Mitigating channel distortions in wireless orthogonal frequency division multiplexing communication systems," Dept. Elect. Eng., Univ. Virginia, Charlottesville, Aug. 1997.

- [7] J. W. Lechleider, "The optimum combination of block codes and receivers for arbitrary channels," *IEEE Trans. Commun.*, vol. 38, pp. 615–621, May 1990.
- [8] B. L. Floch, M. Alard, and C. Berrou, "Coded orthogonal frequency division multiplex," *Proc. IEEE*, vol. 83, pp. 982–996, June 1995.
- [9] P. Melsa, R. C. Younce, and C. E. Rohrs, "Impulse response shortening for discrete multitone transceivers," *IEEE Trans. Commun.*, vol. 44, pp. 1662–1672, Dec. 1996.
- [10] T. Pollet and M. Moeneclaey, "The effect of carrier frequency offset on the performance of band limited single carrier and OFDM signals," in *Proc. GLOBECOM*, London, U.K., Nov. 18–22, 1996, pp. 719–723.
- [11] H. Sari, G. Karam, and I. Jeanclaude, "An analysis of orthogonal frequency-division multiplexing for mobile radio applications," in *Prof. Vehic. Technol. Conf.*, Stockholm, Sweden, June 8–10, 1994, pp. 1635–1639.
- [12] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Commun. Mag.*, pp. 100–109, Feb. 1995.
- [13] E. Serpedin and G. B. Giannakis, "Blind channel identification and equalization using modulation induced cyclostationarity," *IEEE Trans. Signal Processing*, vol. 46, pp. 3099–3104, Nov. 1998; see also *Proc. 31st Conf. Inform. Sci. Syst.*, Johns Hopkins Univ., Baltimore, MD, vol. II, Mar. 19–21, 1997, pp. 792–797.
- [14] M. K. Tsatsanis and G. B. Giannakis, "Transmitter induced cyclostationarity for blind channel equalization," *IEEE Trans. Signal Processing*, vol. 45, pp. 1785–1794, July 1997.
- [15] J.-J. van de Beek, M. Sandell, and P. O. Börjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 180–1805, July 1997.
- [16] L. Vandendorpe, "MMSE equalizers for multitone systems without guard time," in *Proc. Euro. Signal Process. Conf.*, Sept. 10–13, 1996.
- [17] E. Viterbo and K. Fazel, "How to combat long echoes in OFDM transmission schemes: Sub-channel equalization or more powerful channel coding," in *Proc. GLOBECOM*, Singapore, Nov. 14–16, 1995, pp. 2069–2074.

## On the Equivalence of Blind Equalizers Based on MRE and Subspace Intersections

David Gesbert, Alle-Jan van der Veen, and A. Paulraj

**Abstract**—Two classes of algorithms for multichannel blind equalization are the mutually referenced equalizer (MRE) method by Gesbert *et al.*, and the subspace intersection (SSI) method by van der Veen *et al.* Although these methods seem, at first sight, unrelated, we show here that certain variants of the SSI and the MRE methods both optimize a new blind criterion, which is referred to as *maximum coherence* and, thus, are equivalent.

**Index Terms**—Array signal processing, fractionally spaced equalization, mobile communications, multichannel blind equalization.

### I. INTRODUCTION

Blind equalization has been an active research area during the last few years. Two major factors appear to drive the wide interest in this topic. First, there is an increasing number of interesting and promising applications in the area of digital communications: wireless

Manuscript received February 17, 1998; revised August 4, 1998. The associate editor coordinating the review of this paper and approving it for publication was Dr. Lai C. Godara.

D. Gesbert and A. Paulraj are with the Information Systems Laboratory, Stanford University, Stanford CA 94305 USA (e-mail: gesbert@rascals.stanford.edu).

A.-J. van der Veen is with the Department of Electrical Engineering/DIMES, Delft University of Technology, Delft, The Netherlands.

Publisher Item Identifier S 1053-587X(99)01349-5.

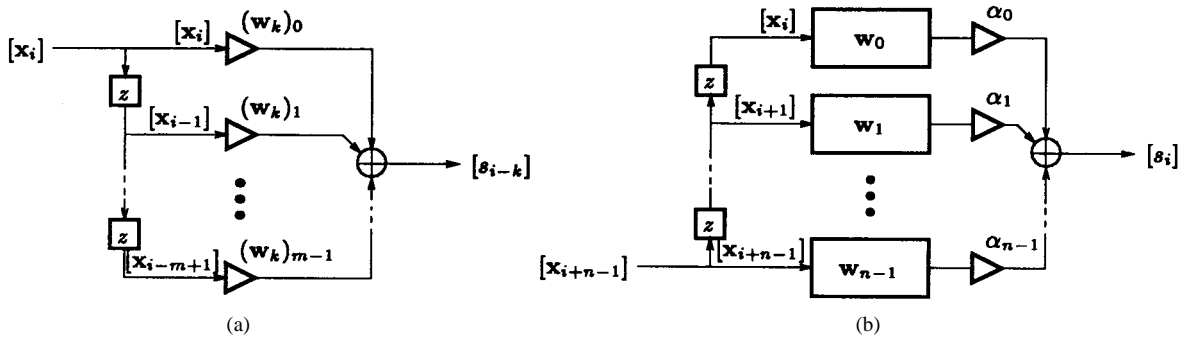


Fig. 1. (a) Equalizer with delay  $k$  and (b) superequalizer, combining the outputs of several equalizers at different delays.

or otherwise. Second, it was recognized that channel oversampling, either temporally (fractionally spaced equalizers) or in space (antenna arrays), leads to a multichannel data representation that offers several new leverages for solving the blind equalization problem and, thus, enhances its applicability.

From an algebraic perspective, oversampling leads to a low-rank model for the output vector signal. This has been extensively exploited in the so-called second-order statistics and algebraic methods for the single-input, multiple-output (SIMO) identification problem [1]. At least three classes can be identified. The first tries to estimate the channels, viz., e.g., [2]–[4], the second considers the estimation of channel inverses (equalizers) [5]–[7], and the third attempts to recover the transmitted symbols directly from a (typically small) batch of output samples without resorting to channel/equalizer estimates [8], [9].

Categories 2 and 3 have the advantage of bypassing the channel estimation step, and this can result in increased robustness. The direct symbol-estimation methods [8], [9] have sometimes been called row-span methods as they exploit the row-span information of the data matrix to find the vector of unknown symbols. Following a seemingly different strategy, MRE techniques [6] estimate a collection of channel equalizers by forcing them to produce the same (unknown) output sequence up to fixed equalization lags. The goal of this correspondence is to demonstrate that these two methods are, in fact, identical with small differences arising only due to variations in the implementation.

In this correspondence, we first provide a new perspective of the row-span method of [9] by showing that the symbol estimates produced by this technique can be regarded as the outputs of linear equalizer averaged across all equalization lags. We show that these equalizers optimize a *maximal coherence* (MC) criterion. Finally, we show the equivalence between the MC criterion and a particular member in the class of MRE criteria.

*Notation:* For a vector  $\mathbf{x}$ ,  $\mathbf{x}^t$  is its transpose,  $\mathbf{x}^*$  its conjugate-transpose, and  $\|\mathbf{x}\|$  its  $\ell_2$ -norm. A sequence (row vector) with entries  $x_i$  is denoted by  $\mathbf{x} = [x_i]$ .

## II. DATA MODEL

### A. Data Matrices

A digital symbol sequence  $[s_i]$  is transmitted through a medium and received by an array of  $M \geq 1$  sensors. The received signals are sampled  $P \geq 1$  times faster than the symbol rate, which, here, is normalized to  $T = 1$ . Hence, during each symbol period, a total of  $MP$  measurements are available, which can be stacked into  $MP$ -dimensional vectors  $\mathbf{x}_i$  as  $\mathbf{x}_i = [x_i^1, \dots, x_i^{MP}]^t$ . Assuming an FIR channel, we can model  $\mathbf{x}_i$  as the output of an  $MP$ -dimensional vector channel with impulse response  $[\mathbf{h}_0, \mathbf{h}_1, \dots, \mathbf{h}_{L-1}]$ , where  $L$  denotes

the channel length. In the noise-free case,  $\mathbf{x}_i$  is then given by

$$\mathbf{x}_i = \sum_{k=0}^{L-1} \mathbf{h}_k s_{i-k}. \quad (1)$$

Consider a finite block of data, and define the  $mMP \times N$  block-Toeplitz data matrix

$$\mathcal{X}^{(i)} = \begin{bmatrix} \mathbf{x}_i & \mathbf{x}_{i+1} & \cdots & \mathbf{x}_{i+N-1} \\ \mathbf{x}_{i-1} & \mathbf{x}_i & \cdots & \mathbf{x}_{i+N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_{i-m+1} & \mathbf{x}_{i-m+2} & \cdots & \mathbf{x}_{i+N-m} \end{bmatrix}.$$

$N$  is the block length, whereas  $m$  can be interpreted as the memory of an equalizer acting on the rows of  $\mathcal{X}^{(i)}$ . Let  $n = L + m - 1$ . From (1),  $\mathcal{X}^{(i)}$  has a factorization as  $\mathcal{X}^{(i)} = \mathcal{H}\mathcal{S}^{(i)}$ , where  $\mathcal{H}$  is an  $mMP \times n$  channel matrix, and  $\mathcal{S}^{(i)}$  is an  $L + m - 1 \times N$  signal matrix, viz.

$$\mathcal{H} = \begin{bmatrix} \mathbf{h}_0 & \cdots & \mathbf{h}_{L-1} & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{h}_0 & \cdots & \mathbf{h}_{L-1} \end{bmatrix}$$

and

$$\mathcal{S}^{(i)} = \begin{bmatrix} s_i & s_{i+1} & \cdots & s_{i+N-1} \\ \vdots & \vdots & \ddots & \vdots \\ s_{i-n+1} & s_{i-n+2} & \cdots & s_{i+N-m} \end{bmatrix}. \quad (2)$$

We will assume that  $\mathcal{H}$  is tall ( $mMP \geq L + m - 1$ ) and  $\mathcal{S}^{(i)}$  is wide ( $L + m - 1 \leq N$ ) so that this is a low-rank factorization. This requires at least  $MP \geq 2$  and a sufficiently large  $m$  and  $N$ . We assume that  $\mathcal{H}$  has full column rank; therefore, we can recover any row of  $\mathcal{S}^{(i)}$  by taking linear combinations of the rows of  $\mathcal{X}^{(i)}$ . Finally, the matrices  $\mathcal{S}^{(i)}$  are supposed to have full row rank.

### B. Equalizers

An equalizer with delay  $k$  acting on  $\mathcal{X}^{(i)}$  tries to reconstruct the  $k + 1$ st row of  $\mathcal{S}^{(i)}$

$$\mathbf{w}_k^* \mathcal{X}^{(i)} = [s_{i-k} \quad s_{i-k+1} \quad \cdots].$$

See Fig. 1(a). Since  $\mathcal{S}^{(i)}$  has  $n$  rows, there is a total of  $n$  possible delays, and hence, there are  $n$  different equalizers  $\mathbf{w}_k$  ( $k = 0, \dots, n - 1$ ). Note, in particular, that  $\mathbf{w}_0^* \mathcal{X}^{(i)} = [s_0 \quad s_1 \quad \cdots]$ , and hence

$$\mathbf{w}_i^* \mathcal{X}^{(i)} = \mathbf{w}_k^* \mathcal{X}^{(k)}, \quad i, k = 0, \dots, n - 1. \quad (3)$$

If  $m$  is large enough, then  $\mathcal{X}^{(i)}$  is rank deficient, leading to nonuniqueness for the equalizers  $\{\mathbf{w}_i\}$ . Any vector from the left null

space of  $\mathcal{X}^{(i)}$  may be added. The null space component is removed if we require the equalizer to have minimum norm. We can also define the equalizer to act on a minimal basis of the row span of  $\mathcal{X}^{(i)}$  rather than  $\mathcal{X}^{(i)}$  itself. Thus, we introduce the SVD's

$$\mathcal{X}^{(i)} = U_i \Sigma_i V^{(i)}, \quad i = 0, \dots, n-1.$$

If  $\mathcal{X}^{(i)}$  has rank  $n$ , then  $U_i$  has  $n$  orthonormal columns,  $V^{(i)}$  has  $n$  orthonormal rows, and  $\Sigma_i$  is a diagonal matrix containing the  $n$  nonzero singular values. The rows of  $V^{(i)}$  form an orthonormal basis for the row span of  $\mathcal{X}^{(i)}$ . A "normalized" equalizer acting on  $V^{(i)}$  is called  $\mathbf{t}_i$ , which is related to  $\mathbf{w}_i$  via  $\mathbf{t}_i = \Sigma_i U_i^* \mathbf{w}_i$ . Similarly to regular equalizers, we have (for  $i, k = 0, \dots, n-1$ )

$$\mathbf{t}_i^* V^{(i)} = [s_0 \quad s_1 \quad \dots]$$

and

$$\mathbf{t}_i^* V^{(i)} = \mathbf{t}_k^* V^{(k)}. \quad (4)$$

### C. Superequalizers

Define

$$X_T = \begin{bmatrix} \mathcal{X}^{(0)} \\ \vdots \\ \mathcal{X}^{(n-1)} \end{bmatrix}, \quad V_T = \begin{bmatrix} V^{(0)} \\ \vdots \\ V^{(n-1)} \end{bmatrix}. \quad (5)$$

"Superequalizers" are long vectors that collect several equalizers with different delays, each reconstructing the same sequence  $[s_0 \ s_1 \ \dots]$ . They act on the data  $X_T$  or on the normalized data  $V_T$ , respectively

$$\mathbf{w}^* = [\mathbf{w}_0^* \ \dots \ \mathbf{w}_{n-1}^*], \quad \mathbf{t}^* = [\mathbf{t}_0^* \ \dots \ \mathbf{t}_{n-1}^*].$$

It is interesting to consider the superequalizer as combining the outputs of the regular equalizers, forming an average over all admissible delays. (By itself, it can also be interpreted as an ordinary equalizer of length  $n+m-1$  at delay  $n-1$ .) See Fig. 1(b). Note that there is an issue of how to weight the outputs of each equalizer to combine them in an optimal fashion.

## III. BLIND EQUALIZATION

### A. Subspace Intersection Method

The problem of blind equalization is, for given a data matrix  $\mathcal{X}$ , to find a factorization  $\mathcal{X} = \mathcal{H}\mathcal{S}$ , where  $\mathcal{S}$  meets the required Toeplitz structure. Since a Toeplitz matrix is generated by a single vector in a linear way, this translates to finding  $\mathbf{s} = [s_0 \ s_1 \ \dots \ s_{N-1}]$  such that  $\mathbf{s}$  lies simultaneously in row( $\mathcal{X}^{(0)}$ ), row( $\mathcal{X}^{(1)}$ ),  $\dots$ , and row( $\mathcal{X}^{(n-1)}$ ), where "row( $\cdot$ )" stands for the row span. The goal of subspace intersection methods (SSI's) such as in [8] and [9] is to find the single vector  $\mathbf{s}$ , which is in the intersection of all  $n$  subspaces.

Numerically, there are several ways to compute the intersection. The algorithm proposed in [8] constructs the union of the complement of all row spans and takes the complement again. The problem with this is that the complementary spaces can be highly dimensional (order  $N$  each). The "minimum noise subspace" (MNS) technique [10] is a method to prune the dimensions of each complementary space without changing the resulting union too much, thus greatly reducing the complexity. Although it was proposed in a different context, it could be translated to apply to the current situation, but the pruning would still incur a loss in performance.

It was proven in [9] that since the rows of  $V^{(i)}$  form a minimal and "orthonormal" basis for row( $\mathcal{X}^{(i)}$ ), the exact intersection can also be obtained by constructing the matrix  $V_T$  in (5) and looking for the right singular vector corresponding to the *largest* singular value of  $V_T$ . This computation has a complexity that is much smaller than the algorithm in [8] and smaller than what the MNS technique would

give. Nonetheless, even with noise perturbations, we find exactly the same output sequence as that produced by the algorithm in [8]. The corresponding principal left singular vector of  $V_T$  can be interpreted as the superequalizer that returns this sequence.

In particular, it is proven in [9] that if  $\mathbf{t}_{ssi}$  is the principal left singular vector of  $V_T$  and  $n = L + m - 1$ , then (without noise)

$$\mathbf{t}_{ssi}^* V_T = \alpha [s_0 \quad s_1 \quad \dots \quad s_{N-1}]$$

where  $\alpha$  is some nonzero scalar that makes the output sequence have norm 1. Because of the normalization, the largest singular value of  $V_T$  is bounded by  $\sqrt{n}$ . This bound is attained when  $\mathbf{t}_{ssi}^* = [\mathbf{t}_0^* \ \dots \ \mathbf{t}_{n-1}^*]$ , where each component by itself is an equalizer on the normalized signals [viz. (4)], returning a multiple  $\alpha_i$  of  $[s_0 \ s_1 \ \dots]$ . In fact, all scaling  $\alpha_i$  will be the same.

Thus,  $\mathbf{t}_{ssi}$  is a superequalizer in the sense of Section II-C. The corresponding equalizer on unnormalized data  $X_T$  is denoted by  $\mathbf{w}_{ssi}$  and related to  $\mathbf{t}_{ssi}$  via

$$\mathbf{w}_{ssi} = [\mathbf{w}_0^* \ \dots \ \mathbf{w}_{n-1}^*]^*, \quad \mathbf{w}_i = U_i \Sigma_i^{-1} \mathbf{t}_i. \quad (6)$$

### B. Maximal Coherence Criterion

The principal left singular vector  $\mathbf{t}_{ssi}$  of  $V_T$  can also be expressed in terms of a criterion on the unnormalized received data. Indeed,  $\mathbf{t}_{ssi}$  can be written as

$$\mathbf{t}_{ssi} = \arg \max_{\|\mathbf{u}\|^2=1} \mathbf{u}^* \mathcal{R}_V \mathbf{u}$$

where  $\mathcal{R}_V = V_T V_T^*$ . Define the (empirical) correlation matrices  $R_{i,j} = \mathcal{X}^{(i)} \mathcal{X}^{(j)*}$

$$\mathcal{R}_X = X_T X_T^* = \begin{bmatrix} R_{0,0} & \dots & R_{0,n-1} \\ \vdots & & \vdots \\ R_{n-1,0} & \dots & R_{n-1,n-1} \end{bmatrix}$$

and

$$\mathcal{R}_0 = \begin{bmatrix} R_{0,0} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & R_{n-1,n-1} \end{bmatrix}.$$

Then  $\mathcal{R}_X = \mathcal{R}_0^{1/2} \mathcal{R}_V \mathcal{R}_0^{1/2*}$ , where

$$\mathcal{R}_0^{1/2} = \begin{bmatrix} R_{0,0}^{1/2} & & 0 \\ & \ddots & \\ 0 & & R_{n-1,n-1}^{1/2} \end{bmatrix}$$

and  $R_{i,i}^{1/2} := U_i \Sigma_i$ .

It follows that  $\mathbf{w}^* \mathcal{R}_X \mathbf{w} = \mathbf{u}^* \mathcal{R}_V \mathbf{u}$  for  $\mathbf{u} = \mathcal{R}_0^{1/2*} \mathbf{w}$ . Now, denote by  $\mathbf{w}_{ssi}$  the corresponding superequalizer provided by the SSI method [related to  $\mathbf{t}_{ssi}$  as in (6)]. By substitution,  $\mathbf{w}_{ssi}$  is found to optimize the constrained criterion

$$\mathbf{w}_{ssi} = \arg \max_{\mathbf{w}^* \mathcal{R}_0 \mathbf{w} = 1} \mathbf{w}^* \mathcal{R}_X \mathbf{w} = \arg \max_{\mathbf{w}^* \mathcal{R}_0 \mathbf{w} = 1} J_{ssi} \quad (7)$$

where  $J_{ssi}$  is given by

$$J_{ssi} := \left\| \sum_{i=0}^{n-1} \mathbf{w}_i^* \mathcal{X}^{(i)} \right\|^2$$

and the constraint can be written as

$$\mathbf{w}^* \mathcal{R}_0 \mathbf{w} = \sum_{i=0}^{n-1} \left\| \mathbf{w}_i^* \mathcal{X}^{(i)} \right\|^2 = 1. \quad (8)$$

Thus, the subspace intersection solution is also obtained by maximizing the power of the sum of all equalizer's outputs, subject to the constraint that the sum of the powers is kept constant. *The SSI*

$$\mathbf{w}^* \mathcal{X}_{mre} = 0 \quad \mathcal{X}_{mre} := \begin{bmatrix} \mathcal{X}^{(0)} & \mathcal{X}^{(0)} & \dots & \mathcal{X}^{(0)} \\ -\mathcal{X}^{(1)} & & & \\ & -\mathcal{X}^{(2)} & & \\ & & \ddots & \\ & & & -\mathcal{X}^{(n-1)} \end{bmatrix} \begin{bmatrix} -\mathcal{X}^{(0)} \\ \mathcal{X}^{(1)} & \mathcal{X}^{(1)} & \dots & \mathcal{X}^{(1)} \\ & -\mathcal{X}^{(2)} & & \\ & & \ddots & \\ & & & -\mathcal{X}^{(n-1)} \end{bmatrix} \begin{bmatrix} \dots \\ \dots \\ \dots \end{bmatrix} \quad (8a)$$

method maximizes the coherence of the equalizer's outputs. Indeed, in the noise-free case, all equalizers return the *same* output sequence  $[s_0 \ s_1 \ \dots]$  up to a common scaling. Note that this is true only in the case of the constraint specified in (8).

### C. The MRE Method

The idea behind the mutually referenced equalizer (MRE) method for blind equalization [6] is to exploit the relations in (3) by finding a vector of  $n$  equalizers  $\mathbf{w} = [\mathbf{w}_0^* \ \dots \ \mathbf{w}_{n-1}^*]^*$  that simultaneously minimizes all differences  $\|\mathbf{w}_i^* \mathcal{X}^{(i)} - \mathbf{w}_k^* \mathcal{X}^{(k)}\|^2$ . This can be written as a least-squares problem,<sup>1</sup> as shown in (8a) at the top of the page. To avoid trivial solutions,  $\mathbf{w}$  should be constrained, e.g., by fixing one of its entries or its norm. Another suitable constraint is one that keeps the sum of output powers to a constant  $\mathbf{w}^* \mathcal{R}_0 \mathbf{w} = 1$ . The motivation for this particular choice is that it avoids trivial null space solutions  $\mathbf{w}_i^* \mathcal{X}^{(i)} = \mathbf{0} \ \forall i$ , which is necessary in the noise-free case. Thus, we obtain

$$\begin{aligned} \mathbf{w}_{mre} &:= \arg \min_{\mathbf{w}^* \mathcal{R}_0 \mathbf{w} = 1} J_{mre} \\ J_{mre} &:= \sum_{i=0}^{n-1} \sum_{k=0}^{n-1} \left\| \mathbf{w}_i^* \mathcal{X}^{(i)} - \mathbf{w}_k^* \mathcal{X}^{(k)} \right\|^2. \end{aligned} \quad (9)$$

We elaborate and find

$$\begin{aligned} J_{mre} &= \mathbf{w}^* \mathcal{X}_{mre} \mathcal{X}_{mre}^* \mathbf{w} \\ &= 2\mathbf{w}^* \begin{bmatrix} (n-1)R_{0,0} & -R_{0,1} & \dots & -R_{0,n-1} \\ -R_{1,0} & (n-1)R_{1,1} & & \vdots \\ \vdots & & \ddots & \vdots \\ -R_{n-1,0} & \dots & & (n-1)R_{n-1,n-1} \end{bmatrix} \mathbf{w}. \end{aligned}$$

It thus follows that

$$J_{mre} + 2J_{ssi} = 2n\mathbf{w}^* \mathcal{R}_0 \mathbf{w}.$$

Under the constraint  $\mathbf{w}^* \mathcal{R}_0 \mathbf{w} = 1$ , we finally obtain

$$\min_{\mathbf{w}^* \mathcal{R}_0 \mathbf{w} = 1} J_{mre} = 2n - \max_{\mathbf{w}^* \mathcal{R}_0 \mathbf{w} = 1} J_{ssi}.$$

This means that  $\mathbf{w}_{mre} \equiv \mathbf{w}_{ssi}$ .

Hence, we conclude that the SSI method and the extended MRE method under the output power constraint are identical. Note that the MRE method can use several other constraints; however, only the one presented here guarantees the equivalence of the two methods.

### D. Remarks

The SSI method here is slightly different from the version in [9]. There, the sequence was extended with additional tail symbols, which changed the definition of  $V_T$  such that only a single matrix  $V^{(0)}$  was needed so that only a single data matrix has to be normalized, leading to computational savings. This implementation of the SSI method is asymptotically identical to the one presented here, which

<sup>1</sup>The equation is reminiscent of the cross-relation method in [4], but this connection is only optical. Here, we estimate equalizers and not the channel, as in [4]. More importantly, the CR method does not cross-relate delays of the full data matrices but rather the  $MP$  scalar subchannels so that the superscript  $(i)$  in  $\mathcal{X}^{(i)}$  has a different meaning.

was chosen for expository reasons. With noise, the SSI method on normalized data  $V_T$  and on original data  $X_T$  are slightly different. The reason is that with noise, each  $\mathcal{X}^{(i)}$  is always full rank, whereas  $V^{(i)}$  is presumably obtained from a truncated SVD, resulting in an approximate  $n$ -dimensional basis for the row span of  $\mathcal{X}^{(i)}$ . If we omit the truncation, i.e., define  $V^{(i)}$  to contain all  $mMP$  right singular vectors of  $\mathcal{X}^{(i)}$ , then the solution is exactly equal to the SSI method on  $V_T$ .

### REFERENCES

- [1] H. Liu, G. Xu, L. Tong, and T. Kailath, "Recent developments in blind channel equalization: From cyclostationarity to subspaces," *Signal Process.*, vol. 50, nos. 1/2, pp. 83–99, Apr. 1996.
- [2] 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.
- [3] 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.
- [4] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind equalization," *IEEE Trans. Signal Processing*, vol. 43, pp. 2982–2993, Dec. 1995.
- [5] D. T. M. Stock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction," in *Proc. ICASSP Conf.*, Adelaide, Australia, Apr. 1994, pp. IV.585–IV.588.
- [6] D. Gesbert, P. Duhamel, and S. Mayrargue, "On-line blind multichannel equalization based on mutually referenced filters," *IEEE Trans. Signal Processing*, vol. 45, pp. 2307–2317, Sept. 1997.
- [7] G. Giannakis and S. Halford, "Blind fractionally spaced equalization of noisy FIR channels: Direct and adaptive solutions," *IEEE Trans. Signal Processing*, vol. 45, pp. 2277–2292, Sept. 1997.
- [8] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications," *IEEE Trans. Signal Processing*, vol. 43, pp. 2714–2723, Nov. 1995.
- [9] 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.
- [10] Y. Hua, K. Abed-Meraim, and M. Wax, "Blind system identification using minimum noise subspace," *IEEE Trans. Signal Processing*, vol. 45, pp. 770–773, Mar. 1997.