

## Research Article

# A Two-Stage Approach for Improving the Convergence of Least-Mean-Square Adaptive Decision-Feedback Equalizers in the Presence of Severe Narrowband Interference

Arun Batra,<sup>1</sup> James R. Zeidler,<sup>1</sup> and A. A. (Louis) Beex<sup>2</sup>

<sup>1</sup> Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0407, USA

<sup>2</sup> Wireless@VT and the DSP Research Laboratory, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061-0111, USA

Correspondence should be addressed to Arun Batra, abatra@ucsd.edu

Received 3 January 2007; Revised 16 April 2007; Accepted 8 August 2007

Recommended by Peter Handel

It has previously been shown that a least-mean-square (LMS) decision-feedback filter can mitigate the effect of narrowband interference (L.-M. Li and L. Milstein, 1983). An adaptive implementation of the filter was shown to converge relatively quickly for mild interference. It is shown here, however, that in the case of severe narrowband interference, the LMS decision-feedback equalizer (DFE) requires a very large number of training symbols for convergence, making it unsuitable for some types of communication systems. This paper investigates the introduction of an LMS prediction-error filter (PEF) as a prefilter to the equalizer and demonstrates that it reduces the convergence time of the two-stage system by as much as two orders of magnitude. It is also shown that the steady-state bit-error rate (BER) performance of the proposed system is still approximately equal to that attained in steady-state by the LMS DFE-only. Finally, it is shown that the two-stage system can be implemented without the use of training symbols. This two-stage structure lowers the complexity of the overall system by reducing the number of filter taps that need to be adapted, while incurring a slight loss in the steady-state BER.

Copyright © 2008 Arun Batra et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

Maintaining reliable wireless communication performance is a challenging problem because of channel impairments such as fading, intersymbol interference (ISI), narrowband interference, and noise. Therefore, there is a need for innovative receivers which can mitigate these impairments rapidly, especially when the information is being transferred in small packets or short bursts.

There has been a considerable amount of work on mitigating the effects of ISI (see [1] and references therein) and fading channels (see [2] and references therein). The focus of this paper is on techniques that can quickly mitigate strong narrowband interference. Narrowband interference typically occurs because of nonlinearities in the mixer or by other communication systems radiating in the same frequency band (as occurs in many of the unlicensed bands, e.g., Bluetooth is a narrowband interferer for WLAN systems). A

strong interferer can make recovering the transmitted information quite challenging.

Several methods for suppressing narrowband interference have been discussed in the literature. A linear equalizer (LE) and a decision-feedback equalizer (DFE) were studied in [3]. It was shown that the performance of the DFE is better than that of the LE. The LE seen in both systems removes the interference, while the additional feedback taps of the DFE enable the cancellation of the post-cursor ISI that is induced by the LE. Linear prediction [4, 5] is another common technique that has been used in direct-sequence CDMA systems [6–8] when the processing gain does not provide enough immunity to the interference. When the signal of interest is wideband compared to the bandwidth of the interferer, linear prediction predicts the current value of the interference from past samples. When the structure is implemented as a prediction-error filter, the estimate of the interference is removed at the cost of some signal distortion. A further

review of interference suppression techniques can be found in [9, 10].

When the statistics of the interference are known, the weights of these systems are found by minimizing the mean-squared error [11] (or equivalently by solving the Wiener-Hopf equation). In practice, however, this type of a priori information is not available. Thus, these systems are best implemented adaptively. Of the many algorithms available, we focus on a low-complexity method, specifically the least-mean square (LMS) algorithm [11]. The LMS algorithm is also noted for its robustness and improved tracking performance [11, 12]. The drawback of this particular algorithm is its slow convergence when there is a large disparity in the eigenvalues of the input signal [11]. Slow convergence leads to the need for a large number of training symbols. These symbols do not transmit any new information, reducing the overall throughput of the system. Conventional analyses of adaptive algorithms use the mean-squared error (MSE) as the metric when investigating the convergence. However, since BER is a more definitive performance metric for analyzing communication systems, the convergence is viewed in terms of the BER with the aid of a sliding window. Convergence is defined as the number of symbols needed to attain a certain BER.

Although it has been shown that alternate adaptive algorithms, such as the recursive least squares (RLS) algorithm [11], provide improved convergence relative to the LMS algorithm in cases of high eigenvalue disparity, there are many reasons why LMS is chosen for practical communications system applications. Hassibi discusses [12] some of the fundamental differences in the performance of gradient-based estimators such as the LMS algorithm and time-averaged recursive estimators such as the RLS algorithm in the cases of modeling errors and incomplete statistical information concerning the input signal, interference, and noise parameters. Hassibi [12] examines the conditions for which LMS can be shown to be more robust to variations and uncertainties in the signaling environment than RLS. LMS has also been shown to track more accurately than RLS because it is able to base the filter updates on the instantaneous error rather than the time-averaged error [13–16]. The improved tracking performance of LMS over RLS for a linear chirp input is well established [11, 16]. In [17] it is shown that an extended RLS filter that estimates the chirp rate of the input signal can minimize the tracking errors associated with the RLS algorithm and provides performance that exceeds that of LMS. It should be noted, however, that the improved tracking performance requires a significant increase in computational complexity and knowledge that the underlying variations in the input signal can be accurately modeled by a linear FM chirp. For cases where the input is not accurately represented by the linear chirp model, performance can be expected to be significantly worse than simply using an LMS estimator, for the reasons discussed in [12]. The computational complexity of RLS, in particular for high-order systems, favors the use of LMS. The latter is also more robust in fixed-point implementations. In addition, the LMS estimator has been shown to provide nonlinear, time-varying weight dynamics that allow the LMS filter to perform significantly better than the

time-invariant Wiener filter in several cases of practical interest [18, 19]. It is further shown that the improved performance associated with these non-Wiener effects is difficult to realize for RLS estimators due to the time averaging that is inherent in the estimation process [20].

In this paper, we first demonstrate that the LMS DFE possesses an extended convergence time (greater than 10,000 symbols for the cases investigated here) when severe narrow-band interference ( $SIR < -20$  dB) is present, due to the fact that the equalizer does not have a true reference for the interference. To reduce the convergence time and the number of training symbols needed, we propose a two-stage system that uses an LMS prediction-error filter (PEF) as a prefilter to the LMS DFE-only. For strong interference the PEF generates a direct reference for the interference from past samples and mitigates it prior to equalization.

A two-stage system employing a linear predictor has been previously investigated [21, 22] in combination with the constant modulus algorithm (CMA). The prediction filter is employed to mitigate the interference and ensure that the CMA locks on to the signal of interest. The prediction filter is not used specifically for its convergence properties. The two-stage structure in this paper uses a supervised algorithm for the adaptation of the second structure and is developed with the goal of improving the convergence of the overall system.

The second contribution of this paper is to show that the two-stage system reduces the number of training symbols required to reach a BER of  $10^{-2}$  by two orders of magnitude without substantially degrading the steady-state BER performance as compared to the LMS DFE-only case. All comparisons will be made under the condition that the LMS DFE-only and the two-stage structure have the same total number of taps. The two-stage system's adaptive implementation is superior due to the fact that the prediction-error filter utilizes the narrowband nature of the interference to obtain a beneficial initialization point. On the other hand, the LMS DFE-only employs only the training symbols which have no knowledge of the statistical characteristics of the interference.

Finally, the two-stage system may be implemented in a manner that does not require any training symbols. The PEF is inherently a blind algorithm because the error signal is determined from the current sample and the past samples. A relationship between the PEF weights and the DFE feedback weights is obtained, allowing the DFE to be operated in decision-directed mode after convergence of the PEF weights. This technique outperforms the nonblind decision-directed implementation when a small number of training symbols is used. The nonblind decision-directed implementation suffers because the feedback weights lie far from their steady-state values prior to the switch to decision-directed mode. This blind method also allows for a reduction in the complexity of the system (i.e., fewer weights that need to be adapted) at the cost of a slight increase in steady-state BER.

The paper is organized as follows. Section 2 describes the system model. The LMS algorithm and its convergence properties are reviewed in Section 3. In Section 4, the previous approaches of the DFE and the PEF are discussed. The proposed two-stage system is revealed in Section 5 along with its relation to the DFE. A blind implementation for the proposed

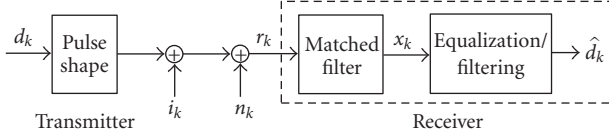


FIGURE 1: Discrete-time system model.

system is also presented in Section 5. In Section 6, the convergence and steady-state BER results are presented. Concluding remarks are given in Section 7.

## 2. SYSTEM MODEL

A complex baseband representation of a single-carrier communication system is depicted in Figure 1. The signal of interest,  $d_k$ , is composed of i.i.d. symbols, drawn from an arbitrary QAM constellation, with average power equal to  $\sigma_s^2$ . It is passed through a pulse shaping filter that is necessary for bandlimited transmission. This signal is corrupted by narrowband interference,  $i_k$ , modeled as a pure complex exponential and additive white Gaussian noise. A matched filter is employed at the receiver to maximize the signal-to-noise ratio (SNR) at the output of the filter. Note that the overall frequency response of the pulse shape and the matched filter is assumed to satisfy Nyquist's criterion for no intersymbol interference (ISI) and the filters operate at the symbol rate.

The signal at the input to the equalizer,  $x_k$ , is defined as

$$\begin{aligned} x_k &= d_k + i_k + n_k \\ &= d_k + \sqrt{J}e^{j(\Omega kT + \theta)} + n_k, \end{aligned} \quad (1)$$

where  $T$  is the symbol duration,  $J$  is the interferer power,  $\Omega$  is the angular frequency of the interferer, and  $\theta$  is a random phase that is uniformly distributed between 0 and  $2\pi$ . The additive noise,  $n_k$ , is modeled as a zero-mean Gaussian random process with variance  $\sigma_n^2$ . The signal-to-noise ratio is defined as  $\text{SNR} = \sigma_s^2/\sigma_n^2$  and the signal-to-interference ratio is defined as  $\text{SIR} = \sigma_s^2/J$ .

It is assumed that the communication signal, interferer, and noise are uncorrelated to each other. The autocorrelation function of the input,  $r_x(m)$ , is defined as

$$\begin{aligned} r_x(m) &= E[x_k x_{k-m}^*] \\ &= (\sigma_s^2 + \sigma_n^2)\delta_m + J e^{j\Omega mT}, \end{aligned} \quad (2)$$

where  $E[\cdot]$  is the expectation operator,  $(\cdot)^*$  indicates conjugation, and  $\delta_m$  is the Kronecker delta function.

## 3. LMS ALGORITHM

The LMS algorithm [11] is defined by the following three equations:

$$\begin{aligned} y_k &= \mathbf{w}_k^H \mathbf{x}_k, \\ e_k &= \begin{cases} d_k - y_k, & \text{training,} \\ \hat{d}_k - y_k, & \text{decision-directed,} \end{cases} \\ \mathbf{w}_{k+1} &= \mathbf{w}_k + \mu e_k^* \mathbf{x}_k, \end{aligned} \quad (3)$$

where  $\mathbf{x}_k$  is the input vector to the equalizer,  $\mathbf{w}_k$  is the vector of adapted tap weights,  $d_k$  is the desired signal,  $\hat{d}_k$  is the output of the decision-device when  $y_k$  is its input,  $e_k$  is the error signal,  $\mu$  is the step-size parameter, and  $(\cdot)^H$  represents conjugate (Hermitian) transpose.

Note that there are two stages associated with the adaptive algorithm. The first stage is the training phase, where known training symbols are used to push the filter in the direction of the optimal weights. After the training symbols have been exhausted, the algorithm switches to decision-directed mode. The output of the decision device is used as the desired symbol when calculating the error signal. Ideally, at the end of the training phase the output of the filter is close to the desired signal.

### 3.1. LMS convergence

In conventional analyses, convergence refers to the asymptotic progress of either the adaptive weights or the MSE toward the optimal solutions. The convergence (as well as the stability) of the system is dependent on the step-size. The step-size parameter is chosen in a manner to guarantee convergence in the mean-square sense, namely,

$$0 < \mu < \frac{1}{\lambda_{\max}}, \quad (4)$$

where  $\lambda_{\max}$  is the maximum eigenvalue of the input autocorrelation matrix.

Assuming that the adaptive weights and the input vector are independent, Shensa [23] showed that the convergence of the weight vector can be expressed as

$$\|\mathbf{w}_{\text{opt}} - E[\mathbf{w}_k]\|^2 = \sum_{i=1}^N (1 - \mu\lambda_i)^{2k} |\mathbf{v}^{iH} \mathbf{w}_{\text{opt}}|^2, \quad (5)$$

where  $\lambda_i$  are the eigenvalues and  $\mathbf{v}^i$  are the eigenvectors of the input autocorrelation matrix. The optimal Wiener solution is represented by  $\mathbf{w}_{\text{opt}}$ . A similar equation arises for the convergence of the mean-square error (MSE) [24], when gradient noise (on the order of  $N\mu E[e_{\min}^2]$ ) is neglected

$$\|E[e_k^2] - E[e_{\min}^2]\|^2 = \sum_{i=1}^N (1 - \mu\lambda_i)^{2k} \lambda_i |\mathbf{v}^{iH} \mathbf{w}_{\text{opt}}|^2. \quad (6)$$

Letting the learning curve be approximated by a single exponential allows a time constant [11] to be defined for each mode,

$$\tau_i \approx \frac{1}{2\mu\lambda_i}. \quad (7)$$

The maximum modal time constant is associated with the minimum eigenvalue,

$$\tau_{\max} \approx \frac{1}{2\mu\lambda_{\min}}. \quad (8)$$

This maximal time constant can be seen to be a conservative estimate by examining (5) more closely. The convergence will be influenced only by those eigenvalues for which

the projection of the corresponding eigenvector on the optimal weights is large. Lastly, it can be seen for the case of  $\lambda_i \ll 1$ , that it is possible for the convergence of the filter output (mean-square error) to be faster than the convergence of the filter weights. This is because there may be fewer modes controlling the MSE convergence (i.e., when  $\lambda_i |\mathbf{v}^{iH} \mathbf{w}_{\text{opt}}| < |\mathbf{v}^{iH} \mathbf{w}_{\text{opt}}|$ ).

The equations above provide excellent insight into the convergence of the LMS algorithm; however, in this paper, we are interested in the convergence in a limited time interval when the metric of interest is BER. Therefore, we define the convergence to be the average number of training symbols needed to achieve a BER of  $10^{-2}$ . This value is consistent with the notion that the BER should be less than  $10^{-1}$  when switching from training to decision-directed mode [25]. Additionally, using a convolutional code with an input BER equal to  $10^{-2}$  is equivalent to a BER of  $10^{-5}$  at the output of the decoder [26].

### 3.2. Sliding BER window

As mentioned above, the convergence of an adaptive filter is viewed by the ensemble average learning curve [11], a plot of the MSE versus iteration. Note that in this work, each iteration of the adaptive algorithm occurs at the symbol rate. To examine the convergence of the BER here, we employ a sliding window of  $N_{\text{window}}$  symbols. For example, the first BER value corresponds to the average number of bit errors over symbols 1 through 100; the second value corresponds to the average number of bit errors over symbols 2 through 101; and so forth. These values are then averaged for  $N_{\text{runs}}$  trials. A general formula for BPSK modulation can be seen as

$$\text{BER}_k = \frac{1}{N_{\text{runs}}} \sum_{n=1}^{N_{\text{runs}}} \frac{1}{N_{\text{window}}} \sum_{m=k-N_{\text{window}}+1}^k |d_m^{(n)} - \hat{d}_m^{(n)}|, \quad (9)$$

$$k \geq N_{\text{window}},$$

where  $d_m^{(n)}$  is the  $m$ th transmitted symbol of the  $n$ th packet and  $\hat{d}_m^{(n)}$  is the decision of the  $m$ th symbol of the  $n$ th packet. Note that the minimum nonzero BER value will be equal to  $1/N_{\text{runs}}N_{\text{window}}$ .

## 4. PREVIOUS APPROACHES

### 4.1. Decision-feedback equalizer

#### 4.1.1. Equalizer structure

The DFE is composed of a transversal feedforward filter with  $K + 1$ -taps (one main tap and  $K$  side taps) and a feedback filter that has  $M$ -taps. A block diagram of the DFE is shown in Figure 2. The output of the filter,  $y_{\text{DFE},k}$ , with inputs  $x_k$  and  $\hat{d}_k$  is

$$y_{\text{DFE},k} = \sum_{l=0}^K w_l^* x_{k-l} + \sum_{l=1}^M f_l^* \hat{d}_{k-l}, \quad (10)$$

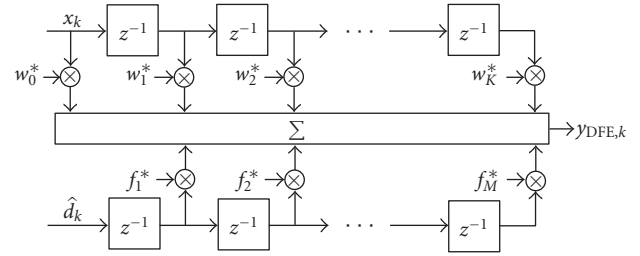


FIGURE 2: Decision-feedback equalizer block diagram.

where  $\hat{d}_k$  is the estimate of the symbol  $d_k$  out of the decision device. Note that  $w_l$  are the tap weights associated with the feedforward filter, and  $f_l$  are the tap weights associated with the feedback filter. During the training phase,  $\hat{d}_k$  in (10) equals  $d_k$ .

The feedback taps allow the equalizer to cancel out post-cursor ISI based on the estimated decisions without enhancing the noise. The BER analysis of the DFE with error propagation can be accomplished utilizing Markov chains to model the term  $[d_{k-l} - \hat{d}_{k-l}]$  as contents of a shift register and the assumption that the feedback decisions are perfect [3, 27–29]. The number of states in the Markov chain grows exponentially with the number of feedback taps.

#### 4.1.2. DFE optimal weights

The optimal weights under the minimum mean-square error (MMSE) criterion can be found using the orthogonality principle [11].  $K + M + 1$  equations are obtained, and the weights can be found using the method described in [3, 30]. The optimal DFE tap weights are given by

$$w_{\text{DFE},l} = \frac{\text{SNR}[(1 + \text{SNR})(\sigma_n^2 + MJ) + (K - M)J]}{(1 + \text{SNR})[(1 + \text{SNR})(\sigma_n^2 + MJ) + (K - M + 1)J]} \quad (11)$$

$$= C_0, \quad l = 0,$$

$$w_{\text{DFE},l} = \frac{-J \text{SNR}}{(1 + \text{SNR})(\sigma_n^2 + MJ) + (K - M + 1)J} e^{-j\Omega l T} \quad (12)$$

$$= C_1 e^{-j\Omega l T}, \quad l = 1, \dots, M,$$

$$w_{\text{DFE},l} = \frac{-J \text{SNR}}{(1 + \text{SNR})[(1 + \text{SNR})(\sigma_n^2 + MJ) + (K - M + 1)J]} e^{-j\Omega l T} \quad (13)$$

$$= \frac{C_1}{1 + \text{SNR}} e^{-j\Omega l T}, \quad l = M + 1, \dots, K,$$

$$f_{\text{DFE},l} = \frac{J \text{SNR}}{(1 + \text{SNR})(\sigma_n^2 + MJ) + (K - M + 1)J} e^{-j\Omega l T} \quad (14)$$

$$= -C_1 e^{-j\Omega l T}, \quad l = 1, \dots, M.$$

Observe that the weight of the feedback taps (14) is the negative of the feedforward side taps (12) when  $l = 1, \dots, M$ . This implies that if the data fed back is perfect, the ISI caused

by the  $M$  previous data symbols will be completely canceled. Also note that (13) is a scaled (by  $1/(1 + \text{SNR})$ ) multiple of (12). This scaling value effectively removes the influence of the associated data symbols that can not be canceled by the feedback taps. For the special case of  $K = M$ , it can be seen that if the data fed back is perfect, the ISI caused by the feedforward equalizer will be completely canceled, leaving only the symbol of interest.

#### 4.1.3. DFE SINR calculation

The signal-to-interference-plus-noise ratio (SINR) at the input to the decision device of the DFE can be found using (10) and the optimal weights given in (11)–(14) to be

$$\begin{aligned} \text{SINR} &= \left( C_0^2 + (K - M) \left( \frac{C_1}{1 + \text{SNR}} \right)^2 \right) \text{SNR} \\ &\times \left( \left( C_0 + MC_1 + (K - M) \frac{C_1}{1 + \text{SNR}} \right)^2 J / \sigma_n^2 \right. \\ &\quad \left. + C_0^2 + MC_1^2 + (K - M) \left( \frac{C_1}{1 + \text{SNR}} \right)^2 \right)^{-1}. \end{aligned} \quad (15)$$

#### 4.1.4. Autocorrelation structure

The input to the decision-feedback equalizer is the concatenation of the received input to the equalizer and the fed back decisions, given by  $\mathbf{u}_k = [\mathbf{x}_k^T, \hat{\mathbf{d}}_k^T]^T$ , where  $(\cdot)^T$  is the transpose operator. The vector,  $\hat{\mathbf{d}}_k$ , is composed of the fed back decisions that are assumed to be correct, and is thus defined as

$$\hat{\mathbf{d}}_k = \mathbf{d}_k \triangleq [d_{k-1}, d_{k-2}, \dots, d_{k-M}]^T. \quad (16)$$

The autocorrelation matrix for the  $K+1$ -tap feedforward and  $M$ -tap feedback equalizer is defined as

$$\mathbf{R}_{\text{DFE}} = E[\mathbf{u}_k \mathbf{u}_k^H] = E \begin{bmatrix} \mathbf{x}_k \mathbf{x}_k^H & \mathbf{x}_k \mathbf{d}_k^H \\ \mathbf{d}_k \mathbf{x}_k^H & \mathbf{d}_k \mathbf{d}_k^H \end{bmatrix} = \begin{bmatrix} r_x(0) & r_x(1) & r_x(2) & \cdots & r_x(K) & 0 & 0 & \cdots & 0 \\ r_x^*(1) & r_x(0) & r_x(1) & \cdots & r_x(K-1) & \sigma_s^2 & 0 & \cdots & 0 \\ r_x^*(2) & r_x^*(1) & r_x(0) & \cdots & r_x(K-2) & 0 & \sigma_s^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r_x^*(K) & r_x^*(K-1) & r_x^*(K-2) & \cdots & r_x(0) & 0 & 0 & \cdots & \sigma_s^2 \\ 0 & \sigma_s^2 & 0 & \cdots & 0 & \sigma_s^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_s^2 & \cdots & 0 & 0 & \sigma_s^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_s^2 & 0 & 0 & \cdots & \sigma_s^2 \end{bmatrix}. \quad (17)$$

The autocorrelation matrix seen in (17) is partitioned into 4 submatrices. The matrices on the diagonal are the autocorrelation matrix of the received input to the equalizer and the

autocorrelation matrix of the data symbols, respectively. The values in the upper left submatrix are given by (2). The cross-correlation matrix between the received input to the equalizer and the data symbols is located on the off-diagonal.

#### 4.1.5. Eigenvalues

There is no closed form expression for determining the eigenvalues of the correlation matrix defined in (17). A method to bound the eigenvalues of positive-definite Toeplitz matrices can be found in [31] and its application to the correlation matrix given in (17) can be found in [32]. However, for the case of  $K \geq 1$  and  $M \geq 2$ , the minimum and maximum eigenvalues are found to be

$$\lambda_{\text{DFE},\min} = \frac{2\sigma_s^2 + \sigma_n^2 - \sqrt{4(\sigma_s^2)^2 + (\sigma_n^2)^2}}{2}, \quad (18)$$

$$\lambda_{\text{DFE},\max} \approx \sigma_s^2 + (K+1)J + \sigma_n^2,$$

and the eigenvalue spread is

$$\chi(\mathbf{R}_{\text{DFE}}) = \frac{\lambda_{\text{DFE},\max}}{\lambda_{\text{DFE},\min}} = \frac{2(\sigma_s^2 + (K+1)J + \sigma_n^2)}{2\sigma_s^2 + \sigma_n^2 - \sqrt{4(\sigma_s^2)^2 + (\sigma_n^2)^2}}. \quad (19)$$

Note that the eigenvalues given in (18) are not a function of  $M$ .

#### 4.1.6. Convergence properties

The projection of any of the eigenvectors on the optimal weight vector is nonzero. This implies that the time constant (8) is inversely proportional to the minimum eigenvalue (i.e.,  $\tau_{\text{DFE}} \approx 1/\{\mu(2\sigma_s^2 + \sigma_n^2 - \sqrt{4(\sigma_s^2)^2 + (\sigma_n^2)^2})\}$ ). The delay in convergence can be attributed to the fact that the DFE does not have a direct reference for the interferer during adaptation and is thus forced to converge on the basis of the training data only. The feedback taps converge slower than the feedforward taps because the DFE is designed such that the interferer is canceled by the feedforward taps, while the feedback taps attempt to cancel out the signal distortion caused by the feedforward taps [3].

## 4.2. Prediction filter

#### 4.2.1. Predictor structure

The linear predictor (LP) is a structure that uses the correlation between past samples to form an estimate of the current sample [11, 25, 33]. A variant of this filter, the prediction-error filter (PEF), has the property that it removes the correlation between samples, thereby whitening the spectrum. A common example of this property is seen when determining the parameters of an autoregressive (AR) process. The prediction-error filter (assuming a sufficient filter order) of such an input provides both the AR parameters and a white output sequence that is equal to the innovations process.

This technique has also been used to remove narrowband interference in many applications [6–8, 29, 30]. The filter is

able to predict the interferer due to its narrowband properties. A block diagram of the prediction-error filter is shown in Figure 3. The PEF is a transversal filter with  $L$  taps. The decorrelation delay ( $\Delta$ ) ensures that the signal of interest at the current sample is decorrelated from the samples in the filter when calculating the error term. Because the data is i.i.d.,  $\Delta = 1$  is a sufficient choice, giving the one-step predictor. The linear combination of the weighted input samples,  $x_k$ , forms an estimate of the interferer, given by

$$y_{LP,k} = \sum_{l=0}^{L-1} c_l^* x_{k-\Delta-l}, \quad (20)$$

where  $c_l$  are the tap weights of the predictor. The output of the PEF,  $y_{PEF,k}$ , is defined as the subtraction of the estimate of the interference given in (20) from the current input sample

$$y_{PEF,k} = x_k - y_{LP,k} = x_k - \sum_{l=0}^{L-1} c_l^* x_{k-\Delta-l}. \quad (21)$$

Note that  $y_{PEF,k}$  is also the error term of the structure. This implies that the PEF is in fact a blind algorithm. It does not require any training symbols when calculating the error term.

#### 4.2.2. Predictor optimal weights

The optimal tap weights can be found in a way similar to those for the equalizer above [3, 30]. Using the orthogonality principle,  $L$  equations are obtained and the weights of the PEF are given by

$$\mathbf{c}_{PEF} = \left[ 1, \underbrace{0, \dots, 0}_{\Delta-1}, -Ae^{-j\Omega\Delta T}, \dots, -Ae^{-j\Omega(L-1+\Delta)T} \right], \quad (22)$$

where  $A$  is equal to

$$A = \frac{J}{\sigma_s^2 + \sigma_n^2 + LJ}. \quad (23)$$

For the scenario of interest in this paper, the interference power is much larger than both the signal power and the noise power. Therefore, the SIR and the noise-to-interference ratio (NIR) can be assumed to be very small (i.e.,  $\text{SIR} \ll 0$  dB,  $\text{NIR} \ll 0$  dB [3]) and  $A$  can be approximated as

$$A \cong \frac{1}{L}. \quad (24)$$

#### 4.2.3. Sensitivity to additive noise

The PEF has been shown to be sensitive to additive noise when used for channel estimation [34, 35]. An algorithm was proposed in [36] to provide adaptive estimation of unbiased linear predictors with the goal of obtaining a consistent estimate of an ISI single-input multiple-output (SIMO) channel. To examine the effect of the additive noise on the PEF

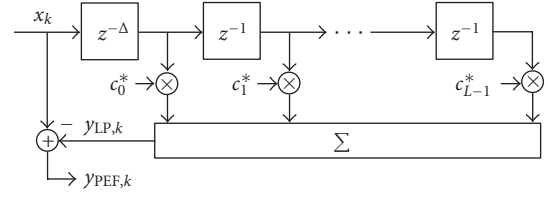


FIGURE 3: Prediction-error filter block diagram.

for this problem, we are interested in the noise-free predictor weights, given by

$$\mathbf{c}_{PEF,\text{no.noise}} = \left[ 1, \underbrace{0, \dots, 0}_{\Delta-1}, -\tilde{A}e^{-j\Omega\Delta T}, \dots, -\tilde{A}e^{-j\Omega(L-1+\Delta)T} \right], \quad (25)$$

where  $\tilde{A}$  is equal to

$$\tilde{A} = \frac{J}{\sigma_s^2 + LJ}. \quad (26)$$

We compare (25) with the biased predictor weights given in (22) and look at the norm of the difference (bias),

$$\|\mathbf{c}_{PEF,\text{no.noise}} - \mathbf{c}_{PEF}\| = \frac{L\sigma_n^2 J}{(\sigma_s^2 + \sigma_n^2 + LJ)(\sigma_s^2 + LJ)}. \quad (27)$$

This bias can be approximated using the assumptions that the SIR and NIR are very small to give

$$\|\mathbf{c}_{PEF,\text{no.noise}} - \mathbf{c}_{PEF}\| \approx \frac{(\sigma_n^2/J)}{\sqrt{L}} = \frac{\text{NIR}}{\sqrt{L}}. \quad (28)$$

The value in (28) is quite small due to the assumption that the NIR is small. Thus, in this work, the bias in the linear predictor does not substantially affect the system's performance.

#### 4.2.4. Autocorrelation structure

The  $L \times L$  input autocorrelation matrix for the PEF is defined as

$$\mathbf{R}_{PEF_i} = E[\mathbf{x}_k \mathbf{x}_k^H] = \begin{bmatrix} r_x(0) & r_x(1) & r_x(2) & \cdots & r_x(L-1) \\ r_x^*(1) & r_x(0) & r_x(1) & \cdots & r_x(L-2) \\ r_x^*(2) & r_x^*(1) & r_x(0) & \cdots & r_x(L-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_x^*(L-1) & r_x^*(L-2) & r_x^*(L-3) & \cdots & r_x(0) \end{bmatrix}, \quad (29)$$

where the components of the matrix are given by (2).

#### 4.2.5. Eigenvalues

The eigenvalues for the correlation matrix given by (2) and (29) can be found [7, 23, 37] to be equal to

$$\lambda_{PEF} = \begin{cases} \sigma_s^2 + \sigma_n^2 + LJ, & \text{order } 1, \\ \sigma_s^2 + \sigma_n^2, & \text{order } L-1. \end{cases} \quad (30)$$

The eigenvalue spread is defined [11] as

$$\chi(\mathbf{R}_{\text{PEF},i}) = \frac{\lambda_{\text{PEF},\max}}{\lambda_{\text{PEF},\min}} = 1 + \frac{LJ}{\sigma_s^2 + \sigma_n^2}. \quad (31)$$

#### 4.2.6. Convergence properties

In this case the  $L - 1$  eigenvectors corresponding to the minimum eigenvalues are orthogonal to the optimal weight vector, hence these eigenvalues do not affect the convergence [23]. Thus the time constant is dependent only upon the maximum eigenvalue (i.e.,  $\tau_{\text{PEF}} \approx 1/\{2\mu(\sigma_s^2 + \sigma_n^2 + LJ)\}$ ).

#### 4.2.7. Output Autocorrelation

The whitening property of the PEF can be seen more clearly through the autocorrelation function of the output of the PEF, which is derived to be

$$\begin{aligned} r_{\text{PEF},o}(m) &= E[y_{\text{PEF},k} y_{\text{PEF},k-m}^*] = (1 - AL)J e^{j\Omega m T} \\ &+ (\sigma_s^2 + \sigma_n^2) \begin{cases} (1 + A^2 L), & m = 0, \\ A^2(L - |m|) e^{j\Omega m T}, & |m| = 1, \dots, \Delta - 1, \\ A(A(L - |m|) - 1) e^{j\Omega m T}, & |m| = \Delta, \dots, L - 1, \\ -A e^{j\Omega m T}, & |m| = L, \dots, L + \Delta - 1. \end{cases} \end{aligned} \quad (32)$$

An approximation for the output autocorrelation function in (32) can be found using the approximation given in (24),

$$\begin{aligned} r_{\text{PEF},o}(m) &\cong (\sigma_s^2 + \sigma_n^2) \begin{cases} 1 + \frac{1}{L}, & m = 0, \\ \left(\frac{1}{L} - \frac{|m|}{L^2}\right) e^{j\Omega m T}, & |m| = 1, \dots, \Delta - 1, \\ -\frac{|m|}{L} e^{j\Omega m T}, & |m| = \Delta, \dots, L - 1, \\ -\frac{1}{L} e^{j\Omega m T}, & |m| = L, \dots, L + \Delta - 1. \end{cases} \end{aligned} \quad (33)$$

Finally, letting the filter order increase toward infinity shows that the output spectrum is approximately white,

$$\lim_{L \rightarrow \infty} r_{\text{PEF},o}(m) \cong (\sigma_s^2 + \sigma_n^2) \delta_m. \quad (34)$$

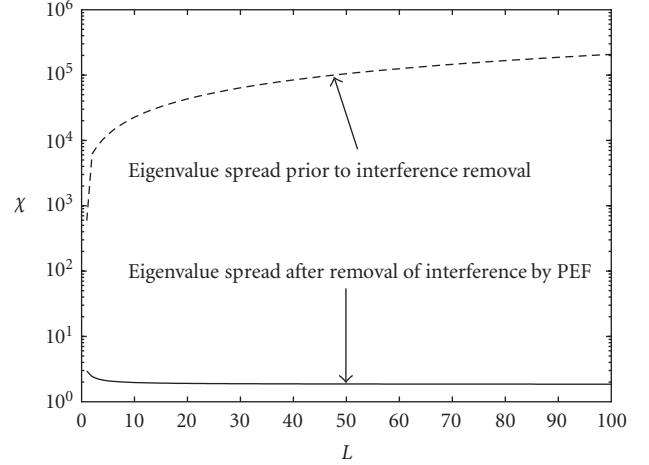


FIGURE 4: Eigenvalue spread of input to DFE-only and output of PEF for SNR = 10 dB, SIR = -20 dB, and  $\Omega = \pi/6$ .

#### 4.2.8. Eigenvalue spread

The effect of the PEF is that the interference is removed, which then results in the reduction of the eigenvalue spread. This can be seen in Figure 4 for SNR = 10 dB, SIR = -20 dB, and  $\Omega = \pi/6$ . Also in the plot is the eigenvalue spread of the received data given by (31). Note that it is assumed that  $L = K$ . It is clearly seen that the spread has been reduced, and the modes of this input to the LMS DFE will converge in similar amounts of time.

## 5. TWO-STAGE SYSTEM

As discussed in Section 4.2.7, the PEF provides an approximately white output spectrum when an infinite number of filter taps is used. Each additional tap provides an increase in spectral resolution when notching out the narrowband interference. However, the implementation of a large number of taps is not generally feasible and some distortion in the form of postcursor ISI will be present. To combat the distortion induced by the PEF, the DFE is a simple structure that removes the ISI without enhancing the noise. This leads to a simple two-stage structure that uses the PEF for rapid convergence and the DFE for removing postcursor ISI as a system to mitigate narrowband interference.

A similar approach is discussed in [25, pages 364-365] when deriving the zero-forcing decision-feedback equalizer. Barry et al. demonstrate that the optimal DFE precursor equalizer is related to optimal linear prediction. Consider transmitting data through a channel that induces ISI. This distortion can be removed by employing a linear zero-forcing equalizer, while causing the noise samples at the output of the equalizer to be correlated. This correlation can be subsequently removed with a PEF, at the expense of postcursor ISI. Finally, a zero-forcing feedback postcursor equalizer removes the ISI without enhancing the noise.

We now consider the performance of the PEF followed by the DFE, which will be abbreviated as PEF + DFE. A block diagram of the two-stage structure is shown in Figure 5. The

PEF is tasked with whitening the spectrum by removing the interference, but due to its limited length it will introduce postcursor ISI; this ISI is then removed by the DFE. The DFE is designed to have a one tap feedforward section and an  $M$ -tap feedback section. In general, there is no need for a feedforward section, because the input is distorted with only postcursor ISI that can be resolved by the feedback equalizer portion. We have chosen to include the one tap to compensate for any phase shifts that might exist because of phase errors, and/or gain mismatch between the transmitter and receiver.

### 5.1. Feedback filter order estimation

We can estimate the optimal feedback filter order by looking at the output of the DFE. Assuming that the feedforward filter weight is equal to 1 and the decisions fed back are perfect, let the output be defined as

$$\begin{aligned} y_{\text{PEF+DFE},k} &= \sum_{n=0}^0 w_{\text{PEF+DFE},n}^* y_{\text{PEF},k-n} + \sum_{n=1}^M f_{\text{PEF+DFE},n}^* d_{k-n} \\ &= y_{\text{PEF},k} + \sum_{n=1}^M f_{\text{PEF+DFE},n}^* d_{k-n}. \end{aligned} \quad (35)$$

We would like to find the weights that minimize the error,

$$\begin{aligned} f_{\text{PEF+DFE},l} &= \arg \min_{f_l} E \left[ \left\| d_k - \left( y_{\text{PEF},k} + \sum_{l=1}^M f_l^* d_{k-l} \right) \right\|^2 \right] \\ &= \arg \min_{f_l} E \left[ \left\| d_k - \left( x_k - A e^{j\Omega\Delta T} \sum_{n=0}^{L-1} x_{k-\Delta-n} e^{j\Omega n T} \right. \right. \right. \\ &\quad \left. \left. \left. + \sum_{l=1}^M f_l^* d_{k-l} \right) \right\|^2 \right]. \end{aligned} \quad (36)$$

Taking the derivative of the expected value term and setting this result to zero, the optimal weights are given by

$$f_{\text{PEF+DFE},l} = \begin{cases} A e^{-j\Omega l T}, & l = \Delta, \dots, \min(M, L), \\ 0, & l = \{1, \dots, \Delta - 1\} \cup \{L + \Delta, \dots, M\}. \end{cases} \quad (37)$$

When  $\Delta = 1$ , the optimal choice for the feedback filter order is  $M = L$ . This ensures that the ISI caused by the PEF is removed. With these choices and the assumption that the interference is canceled by the PEF, the output of the DFE is given by

$$y_{\text{PEF+DFE},k} = d_k + n_k - A \sum_{n=1}^L n_{k-n} e^{j\Omega n T}. \quad (38)$$

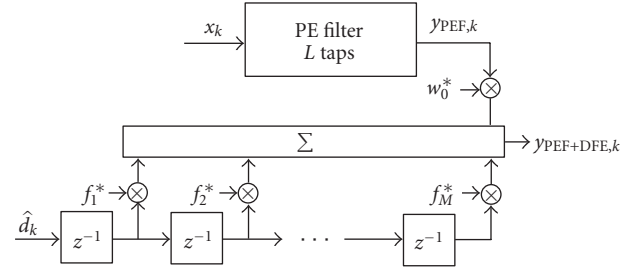


FIGURE 5: Two-stage structure (PEF + DFE) block diagram.

### 5.2. Optimal equalizer weights after prediction-error filtering

The DFE possesses a 1-tap feedforward section and an  $M$ -tap feedback section. The optimal weights for the DFE are found by solving the Wiener-Hopf equations [11, 19]. The feedforward weight is equal to  $w_{\text{PEF+DFE}} = (\mathbf{R}_{\text{PEF},o} - \mathbf{Q}^H \mathbf{Q} / \sigma_s^2)^{-1} \mathbf{p}$ . The output autocorrelation matrix  $\mathbf{R}_{\text{PEF},o}$  reduces to a scalar value due to the 1-tap feedforward filter and is defined as

$$\mathbf{R}_{\text{PEF},o} = r_{\text{PEF},o}(0). \quad (39)$$

The latter term is given in (32).  $\mathbf{Q}$  is defined as

$$\mathbf{Q} = E[\mathbf{d}_k \mathbf{y}_{\text{PEF},k}^*], \quad (40)$$

where the components of  $\mathbf{Q}$  are given by

$$\begin{aligned} E[d_{k-m} y_{\text{PEF},k}^*] &= -A \sigma_s^2 e^{-j\Omega m T}, \\ m &= \{\Delta, \dots, \Delta + L - 1\} \cap \{1, \dots, M\}. \end{aligned} \quad (41)$$

Finally,  $\mathbf{p}$  is defined as

$$\mathbf{p} = E[y_{\text{PEF},k} d_k^*] = \sigma_s^2. \quad (42)$$

The feedback weights are defined as  $\mathbf{f}_{\text{PEF+DFE}} = -\mathbf{Q} w_{\text{PEF+DFE}} / \sigma_s^2$ .

### 5.3. Steady-state equivalence

The two-stage structure can be viewed in a different manner when operating in steady-state. Based on linear system theory, two linear time-invariant (LTI) systems can be combined into one LTI structure [38, pages 107-108]. For example, the PEF weights given in (22) and the feedforward weight of the subsequent DFE ( $w_{\text{PEF+DFE}}$ ) can be combined to form an extended feedforward filter ( $\mathbf{w}_{\text{ext}}$ ) of a DFE with one main tap and  $K = L + \Delta - 1$  side taps. This is accomplished by

$$\mathbf{w}_{\text{ext}} = \mathbf{c}_{\text{PEF}} * w_{\text{PEF+DFE}} = w_{\text{PEF+DFE}} \times \mathbf{c}_{\text{PEF}}, \quad (43)$$

where “\*” represents linear convolution. The feedback taps remain the same, that is  $\mathbf{f}_{\text{ext}} = \mathbf{f}_{\text{PEF+DFE}}$ . Observe that  $\mathbf{w}_{\text{ext}}$  and  $\mathbf{f}_{\text{ext}}$  are the weights of a DFE operating in steady state. The case of interest is when  $\Delta = 1$  and  $L = M$  (as postulated in Section 5.1).



Solving,  $w_{\text{PEF+DFE}} = (\mathbf{R}_{\text{PEF},0} - \mathbf{Q}^H \mathbf{Q} / \sigma_s^2)^{-1} \mathbf{p}$  and  $\mathbf{f}_{\text{PEF+DFE}} = -\mathbf{Q} w_{\text{PEF+DFE}} / \sigma_s^2$  for the weights gives

$$w_{\text{PEF+DFE}} = \frac{\text{SNR}}{\text{SNR} + (A^2 M + 1) + (1 - AM)^2 J / \sigma_n^2}, \quad (44)$$

$$f_{\text{PEF+DFE},l} = \frac{A \text{SNR}}{\text{SNR} + (A^2 M + 1) + (1 - AM)^2 J / \sigma_n^2} e^{-j\Omega l T}, \quad l = 1, \dots, M. \quad (45)$$

The extended feedforward filter weights can be found according to (43),

$$w_{\text{ext},0} = \frac{\text{SNR}}{\text{SNR} + (A^2 M + 1) + (1 - AM)^2 J / \sigma_n^2}, \quad (46)$$

$$w_{\text{ext},l} = \frac{-A \text{SNR}}{\text{SNR} + (A^2 M + 1) + (1 - AM)^2 J / \sigma_n^2} e^{-j\Omega l T}, \quad l = 1, \dots, M, \quad (47)$$

$$f_{\text{ext},l} = \frac{A \text{SNR}}{\text{SNR} + (A^2 M + 1) + (1 - AM)^2 J / \sigma_n^2} e^{-j\Omega l T}, \quad l = 1, \dots, M. \quad (48)$$

Note that the feedback weights remain the same, namely (45) is equal to (48).

As mentioned previously in Section 4.2.2, the scenario of interest occurs when the interference dominates the signal of interest and the noise. Equations (46)–(48) can be approximated in this region using (24) to give

$$\begin{aligned} w_{\text{ext},0} &\cong \frac{\text{SNR}}{(1 + \text{SNR}) + 1/M}, \\ w_{\text{ext},l} &\cong -\frac{\text{SNR}}{(1 + \text{SNR})M + 1} e^{-j\Omega l T}, \quad l = 1, \dots, M, \\ f_{\text{ext},l} &\cong \frac{\text{SNR}}{(1 + \text{SNR})M + 1} e^{-j\Omega l T}, \quad l = 1, \dots, M. \end{aligned} \quad (49)$$

As a comparison to (49), the DFE-only weights described by (11)–(14) need to be approximated for the assumption of small SIR and NIR as well. Letting  $K = M$ , so that there are  $M + 1$  taps in the feedforward section and  $M$  taps in the feedback section, the DFE-only weights are approximated as

$$\begin{aligned} w_{\text{DFE},0} &\cong \frac{\text{SNR}}{(1 + \text{SNR}) + 1/M}, \\ w_{\text{DFE},l} &\cong -\frac{\text{SNR}}{(1 + \text{SNR})M + 1} e^{-j\Omega l T}, \quad l = 1, \dots, M, \\ f_{\text{DFE},l} &\cong \frac{\text{SNR}}{(1 + \text{SNR})M + 1} e^{-j\Omega l T}, \quad l = 1, \dots, M. \end{aligned} \quad (50)$$

Comparing (49) and (50), it can be seen that combining the two-stage weights approximates the weights of the DFE-only.

## 5.4. Blind implementation

The previous sections established a relationship between the PEF weights, the feedforward weight, and the feedback weights. Note that in Section 5.1 the feedback weights are equal to the PEF weights associated with past data symbols scaled by the feedforward tap weighting. Also, recall that the weights of the PEF rapidly converge and the structure does not require knowledge of training symbols. With  $\Delta = 1$  and  $L = M$ , the two-stage system can be implemented in a manner where the feedback tap weights are not adapted. After the PEF weights have converged, the multiplication of the PEF weights and the feedforward weight is used as the feedback weights. The feedforward tap is initialized to unity and is adapted in decision-directed mode. Thus, no explicit training symbols are required during the adaptation process. This method also reduces the complexity of the system; only  $M + 1$  of the total  $2M + 1$  tap weights are adapted. In the scenario where there is a phase and/or gain error, the system requires the use of either training symbols to adapt the feedforward weight or a phase locked loop (PLL) and automatic gain control (AGC). Observe that these two components can be implemented in a decision-directed manner with no need for training symbols.

## 6. RESULTS

### 6.1. Simulation parameters

In the simulation results to follow, a QPSK constellation is utilized and the SNR = 9 dB. For convergence results, a 100-symbol window was used and the BER values are averaged over 1000 runs. The interferer frequency is located at DC ( $\Omega = 0$ ). All of the data were considered as training data, unless specified otherwise. The step-sizes are chosen to ensure convergence toward the steady-state BER. The DFE steady-state BER results in the convergence plots are given by  $Q(\sqrt{\text{SINR}})$ , where  $Q(\cdot)$  is the Q-function [29, page 40] and the SINR is given in (15). The simulation results demonstrating complete agreement with this theory-based result are omitted to avoid unnecessary clutter in the figures to follow.

The DFE adapted with the RLS algorithm [11] is also simulated as a benchmark for the LMS DFE and the LMS PEF + DFE. The forgetting factor and the regularization factor were found through trial and error and set to  $\lambda = 0.99$ ,  $\delta = 0.001$ , respectively, for all simulations.

The adaptive weights are initialized such that the main tap is set to one, resulting in the desired symbol being part of the output of the equalizer. The remaining taps are set to zero.

### 6.2. Convergence results

In previous works [3, 39], the convergence has been viewed through the adaptive weights, even though they may not be unique [18]. As mentioned above in Section 3.1, the convergence of the weights may lag behind the MSE convergence if the eigenvalues are small. Similarly, the weight convergence does not provide an indication of how the BER behaves during the transient period. Thus, the convergence results

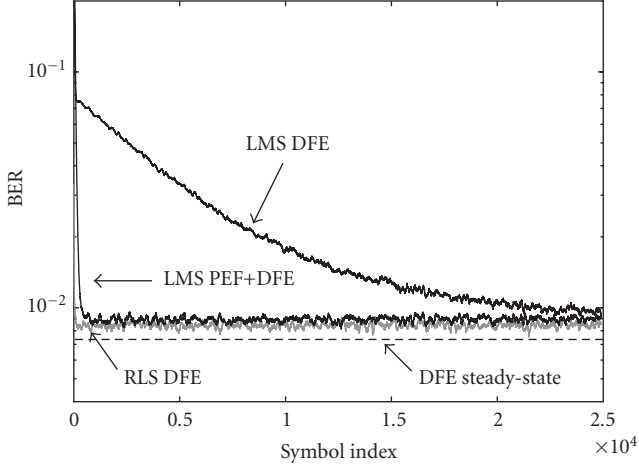


FIGURE 6: Convergence comparison of the LMS DFE, the LMS PEF + DFE, and the RLS DFE for SNR = 9 dB, SIR = -20 dB,  $K = L = M = 3$ ,  $\Omega = 0$ ,  $\mu_{\text{DFE}} = 0.0001$ ,  $\mu_{\text{PEF}} = 0.0001$ ,  $\mu_{\text{PEF+DFE}} = 0.01$ ,  $\lambda = 0.99$ ,  $\delta = 0.001$ .

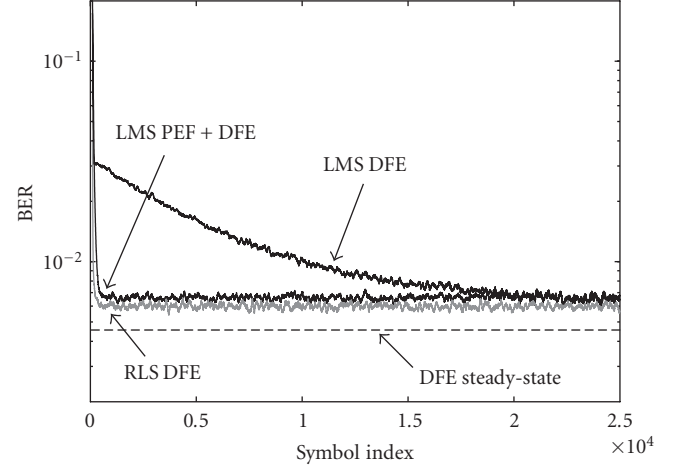


FIGURE 8: Convergence comparison of the LMS DFE, the LMS PEF + DFE, and the RLS DFE for SNR = 9 dB, SIR = -20 dB,  $K = L = M = 6$ ,  $\Omega = 0$ ,  $\mu_{\text{DFE}} = 0.0001$ ,  $\mu_{\text{PEF}} = 0.00005$ ,  $\mu_{\text{PEF+DFE}} = 0.01$ ,  $\lambda = 0.99$ ,  $\delta = 0.001$ .

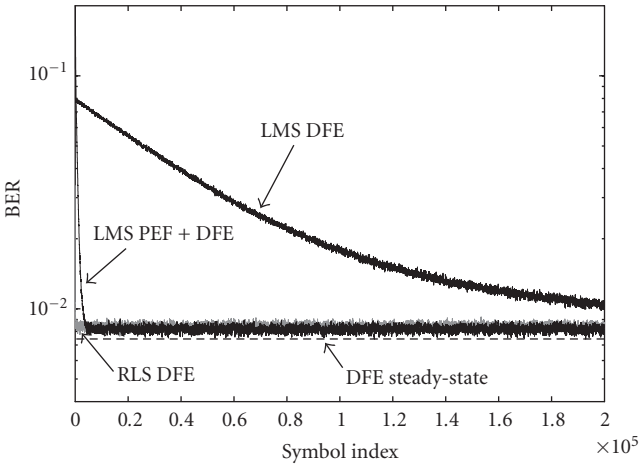


FIGURE 7: Convergence comparison of the LMS DFE, the LMS PEF + DFE, and the RLS DFE for SNR = 9 dB, SIR = -30 dB,  $K = L = M = 3$ ,  $\Omega = 0$ ,  $\mu_{\text{DFE}} = 0.00001$ ,  $\mu_{\text{PEF}} = 0.00001$ ,  $\mu_{\text{PEF+DFE}} = 0.001$ ,  $\lambda = 0.99$ ,  $\delta = 0.001$ .

are shown in terms of a sliding BER window, discussed in Section 3.2.

Figure 6 demonstrates the convergence of the LMS DFE, the LMS PEF + DFE, and the RLS DFE in relation to the steady-state BER for SIR = -20 dB. The number of taps is set such that  $K = L = M = 3$ , and the step-sizes for each structure are  $\mu_{\text{DFE}} = 0.0001$ ,  $\mu_{\text{PEF}} = 0.0001$ ,  $\mu_{\text{PEF+DFE}} = 0.01$ . The LMS PEF + DFE is seen to converge significantly faster than the LMS DFE. Specifically, the LMS PEF + DFE converges to a BER of  $10^{-2}$  in approximately 450 symbols (or iterations, as adaptation takes place at the symbol rate), while the LMS DFE converges in approximately 20 000 symbols. An improvement of two orders of magnitude is obtained by implementing the LMS PEF + DFE structure instead of the LMS DFE structure for this particular scenario. In the case of the

RLS DFE, convergence to a BER of  $10^{-2}$  occurs in 150 symbols. As expected, RLS provides faster convergence because it whitens the input by using the inverse correlation matrix. This improved convergence comes at the cost of higher complexity. For example, in the context of echo cancellation, it has been shown that the implementation of RLS in floating point on the 32 bit, 16 MIPS, 1 serial port, TMS320C31 requires 20 times the number of machine cycles that LMS does [40].

Figure 7 is a plot of the convergence for the above scenario when the SIR = -30 dB. The step-sizes for this case are  $\mu_{\text{DFE}} = 0.00001$ ,  $\mu_{\text{PEF}} = 0.00001$ ,  $\mu_{\text{PEF+DFE}} = 0.001$ . Again, the time required for convergence of the LMS PEF + DFE is dramatically less than for the convergence of the LMS DFE. The LMS PEF + DFE converges in 3000 symbols, while the LMS DFE requires 200 000 symbols. The RLS DFE requires 160 symbols to converge for this case.

Finally, Figure 8 shows the convergence of the two systems when the number of filter coefficients for each stage is doubled, namely,  $K = L = M = 6$  and SIR = -20 dB. The step-sizes for this scenario are  $\mu_{\text{DFE}} = 0.0001$ ,  $\mu_{\text{PEF}} = 0.00005$ ,  $\mu_{\text{PEF+DFE}} = 0.01$ . The LMS PEF + DFE converges in 300 symbols and the LMS DFE converges in 10 000 symbols. The RLS DFE converges in 130 symbols. Doubling the complexity reduces the convergence time of the LMS DFE and the LMS PEF + DFE more than that of the RLS DFE. Note that increasing the order will eventually lead to a degradation in the performance due to the increase of gradient noise. This degradation is observed when increasing the number of taps from  $K = L = M = 3$  (in Figure 6) to  $K = L = M = 6$  (in Figure 8).

### 6.2.1. Blind implementation

In this section, we examine the convergence of the blind implementation discussed in Section 5.4. This algorithm allows the LMS PEF to converge before the LMS DFE that follows

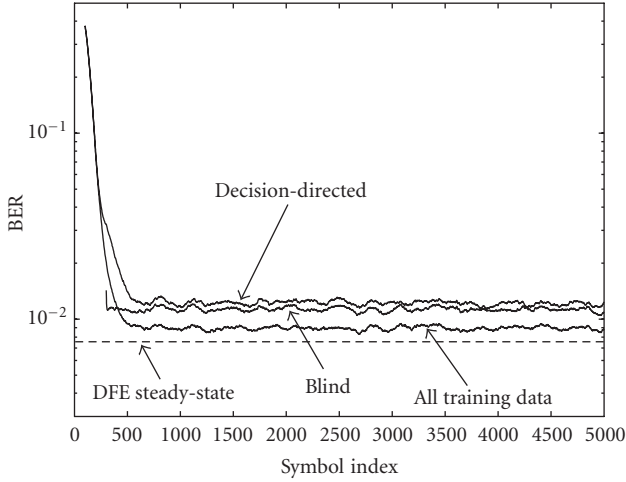


FIGURE 9: Convergence comparison of the different LMS PEF + DFE implementations for SNR = 9 dB, SIR = -20 dB,  $K = L = M = 3$ ,  $\Omega = 0$ ,  $\mu_{\text{PEF}} = 0.0001$ ,  $\mu_{\text{PEF+DFE}} = 0.01$ ,  $N_{\text{off}} = 200$ .

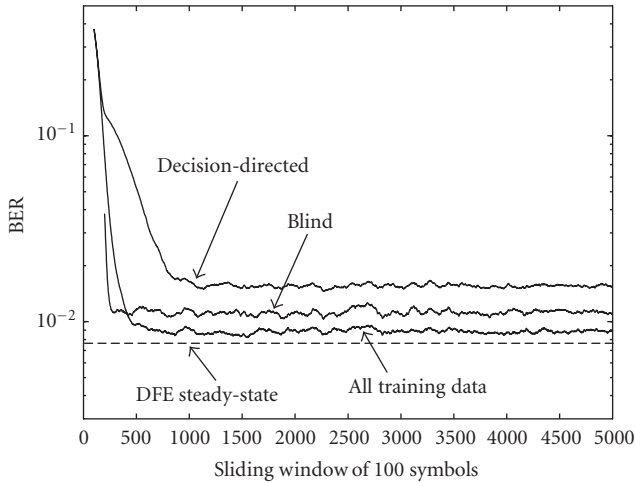


FIGURE 10: Convergence comparison of the different LMS PEF + DFE implementations for SNR = 9 dB, SIR = -20 dB,  $K = L = M = 3$ ,  $\Omega = 0$ ,  $\mu_{\text{PEF}} = 0.0001$ ,  $\mu_{\text{PEF+DFE}} = 0.01$ ,  $N_{\text{off}} = 100$ .

it is turned on. Let  $N_{\text{off}}$  represent the number of symbols that is allocated to allow for PEF convergence. This system is compared to two other cases. The first is the scenario where all the transmitted symbols are considered as training data (similar to the results shown above). The second scenario demonstrates the convergence when a subset of the symbols is used for training, while the adaptive algorithm operates in decision-directed mode for the remaining symbols. We refer to this case as the decision-directed algorithm. The number of training symbols used for this case will also be equal to  $N_{\text{off}}$ .

Figure 9 demonstrates the BER convergence of the three discussed cases in relation to the steady-state BER for SIR = -20 dB and  $N_{\text{off}} = 200$  symbols. The number of taps is set such that  $K = L = M = 3$ , and the step-sizes for each structure are  $\mu_{\text{PEF}} = 0.0001$  and  $\mu_{\text{PEF+DFE}} = 0.01$ . The

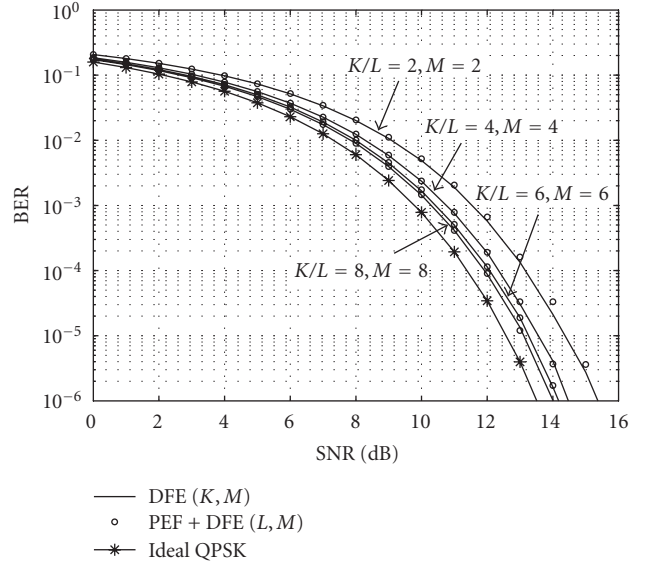


FIGURE 11: Steady-state BER results of the DFE and the PEF + DFE for SIR = -20 dB and  $\Omega = 0$ . DFE results obtained using optimal weights given in (11)–(14), PEF + DFE results obtained using optimal weights given in (22), (44), and (45).

performance of both the blind algorithm and the decision-directed algorithm deviates from the case of using all training data. This is due to propagation of feedback errors that cause more errors. Observe that the blind algorithm produces faster convergence and better BER performance than the decision-directed algorithm.

Figure 10 demonstrates the BER convergence of the three discussed cases in relation to the steady-state BER for SIR = -20 dB, however now  $N_{\text{off}} = 100$  symbols. The blind algorithm outperforms the decision-directed algorithm in terms of both convergence and BER. The degradation of the decision-directed algorithm arises from the fact that the number of training symbols used does not allow the feedback weights to approach their steady-state values before switching to decision-directed mode.

### 6.3. BER results

Figure 11 is a plot of the steady-state BER results for the DFE and PEF + DFE for SIR = -20 dB and varying filter orders. The performance of ideal QPSK is plotted as a reference. The performance of the PEF + DFE is seen to be approximately the same as the performance of the DFE when both structures are operating in steady state. This validates the analysis performed in Section 5.3. It is also seen that the performance of the systems improves as the number of filter taps is increased, approaching the performance of QPSK. The improvement results from the increased spectral resolution provided by the larger number of taps in the feedforward section of each system.

Figure 12 demonstrates the BER results of the LMS PEF + DFE blind implementation in comparison to the steady-state PEF + DFE results. For the blind implementation, the

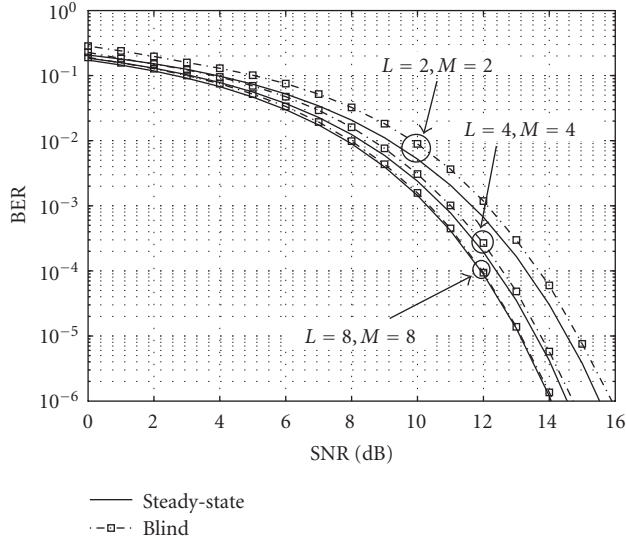


FIGURE 12: Steady-state BER results of PEF + DFE and the BER for the LMS blind implementation for  $SIR = -20$  dB and  $\Omega = 0$ . PEF + DFE steady-state results obtained using optimal weights given in (22), (44), and (45).

TABLE 1: Step-sizes and Convergence (at SNR = 10 dB) for LMS PEF + DFE Blind Implementation.

$L = M$	2	4	8
$\mu_{PEF}$	$1e-4$	$5e-5$	$1e-5$
$\mu_{PEF + DFE}$	0.01	0.01	0.01
No. of symbols to $BER = 10^{-2}$	354	250	555

DFE is turned on after  $N_{off} = 250$  symbols and the BER is calculated over the last 2500 symbols. The step-sizes are chosen for convergence to the steady-state BER and are noted in Table 1. This table also gives the average number of symbols required to obtain a BER of  $10^{-2}$  for the blind implementation when SNR = 10 dB. A convergence value equal to  $N_{off}$  indicates that the blind algorithm has converged to the target BER after the first windowed calculation. It is clear that there is a small degradation in the BER when implementing the blind version of the PEF + DFE algorithm. This degradation is attributed to the combination of the misadjustment of the adaptive algorithm and the presence of uncanceled interference that causes feedback errors. Note that this degradation in BER becomes smaller as the number of parameters is increased. This occurs because a larger number of taps allows for more of the interference to be canceled, thereby reducing the number of feedback errors.

## 7. CONCLUSION

We investigated the response of the LMS DFE in the presence of severe narrowband interference. Due to the absence of a reference for the interference, the convergence time for this equalizer may be unacceptably slow for use in some realistic applications. The proposed system of an LMS PEF as a prefilter to the equalizer is shown to provide a solution to

this problem. This two-stage system was shown to reduce the convergence time, in terms of reaching a BER of  $10^{-2}$ , by two orders of magnitude. An added benefit is that the steady-state BER for the two-stage system approximates that of the LMS DFE-only. Thus, it is possible to improve the convergence results of the LMS DFE, by splitting the system into an LMS prediction-error filter and a separate LMS DFE while not significantly degrading the steady-state BER results. The convergence results were also benchmarked against the DFE adapted with the RLS algorithm, which demonstrated faster convergence at the cost of higher complexity. A blind implementation (i.e., no training symbols are needed) that reduces complexity at the cost of a small degradation in the steady-state BER is also discussed.

## ACKNOWLEDGMENTS

A portion of the material in this paper was presented at the European Signal Processing Conference (EUSIPCO), Florence, Italy, September, 2006. This work was supported by the Office of Naval Research, Code 313, SPAWAR Systems Center, San Diego, and the UCSD Center for Wireless Communications (UCDG Grant no. Com 06-10216). The authors would like to thank the two anonymous reviewers whose comments and suggestions greatly improved the presentation of the material in this paper.

## REFERENCES

- [1] S. U. H. Qureshi, "Adaptive equalization," *Proceedings of the IEEE*, vol. 73, no. 9, pp. 1349–1387, 1985.
- [2] B. Sklar, "Rayleigh fading channels in mobile digital communication systems—I: Characterization," *IEEE Communications Magazine*, vol. 35, no. 7, pp. 90–100, 1997.
- [3] L.-M. Li and L. Milstein, "Rejection of CW interference in QPSK systems using decision-feedback filters," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 473–483, 1983.
- [4] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.
- [5] J. R. Zeidler, "Performance analysis of LMS adaptive prediction filters," *Proceedings of the IEEE*, vol. 78, no. 12, pp. 1781–1806, 1990.
- [6] L.-M. Li and L. B. Milstein, "Rejection of narrow-band interference in PN spread-spectrum systems using transversal filters," *IEEE Transactions on Communications*, vol. 30, no. 5, pp. 925–928, 1982.
- [7] M. J. Reed and B. Liu, "An analysis of LMS adaptive two-sided transversal filters," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '91)*, vol. 3, pp. 2145–2148, Toronto, Ont, Canada, April 1991.
- [8] F. Symons Jr., "Narrow-band interference rejection using the complex linear prediction filter," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 94–98, 1978.
- [9] L. B. Milstein, "Interference rejection techniques in spread spectrum communications," *Proceedings of the IEEE*, vol. 76, no. 6, pp. 657–671, 1988.
- [10] J. D. Laster and J. H. Reed, "Interference rejection in digital wireless communications," *IEEE Signal Processing Magazine*, vol. 14, no. 3, pp. 37–62, 1997.
- [11] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, NJ, USA, 4th edition, 2002.

- [12] B. Hassibi, "On the robustness of LMS filters," in *Least-Mean-Square Adaptive Filters*, S. Haykin and B. Widrow, Eds., chapter 4, pp. 105–144, John Wiley & Sons, Hoboken, NJ, USA, 2003.
- [13] O. M. Macchi and N. J. Bershad, "Adaptive recovery of a chirped sinusoid in noise—I: Performance of the RLS algorithm," *IEEE Transactions on Signal Processing*, vol. 39, no. 3, pp. 583–594, 1991.
- [14] N. J. Bershad and O. M. Macchi, "Adaptive recovery of a chirped sinusoid in noise—II: Performance of the LMS algorithm," *IEEE Transactions on Signal Processing*, vol. 39, no. 3, pp. 595–602, 1991.
- [15] O. M. Macchi, N. J. Bershad, and M. Mboup, "Steady-state superiority of LMS over LS for time-varying line enhancer in noisy environment," *IEE Proceedings, Part F: Radar and Signal Processing*, vol. 138, no. 4, pp. 354–360, 1991.
- [16] P. C. Wei, J. Han, J. R. Zeidler, and W. H. Ku, "Comparative tracking performance of the LMS and RLS algorithms for chirped narrowband signal recovery," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1602–1609, 2002.
- [17] S. Haykin, A. H. Sayed, J. R. Zeidler, P. Yee, and P. C. Wei, "Adaptive tracking of linear time-variant systems by extended RLS algorithms," *IEEE Transactions on Signal Processing*, vol. 45, no. 5, pp. 1118–1128, 1997.
- [18] A. A. Beex and J. R. Zeidler, "Steady-state dynamic weight behavior in (N)LMS adaptive filters," in *Least-Mean-Square Adaptive Filters*, S. Haykin and B. Widrow, Eds., chapter 9, pp. 335–443, John Wiley & Sons, Hoboken, NJ, USA, 2003.
- [19] M. Reuter and J. R. Zeidler, "Nonlinear effects in LMS adaptive equalizers," *IEEE Transactions on Signal Processing*, vol. 47, no. 6, pp. 1570–1579, 1999.
- [20] A. A. Beex and J. R. Zeidler, "Non-Wiener effects in recursive least squares adaptation," in *Proceedings of the 17th International Symposium on Signal Processing and Its Applications (ISSPA '03)*, vol. 2, pp. 595–598, Paris, France, July 2003.
- [21] E. H. Satorius, S. Krishnan, X.-L. Yu, L. J. Griffiths, I. S. Reed, and T.-K. Truong, "Suppression of narrowband interference via single channel adaptive preprocessing," in *Proceedings of the 22nd Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 270–273, Pacific Grove, Calif, USA, October 1988.
- [22] R. P. Gooch and B. Daellenbach, "Prevention of interference capture in a blind (CMA-based) adaptive receive filter," in *Proceedings of the 23rd Annual Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 898–902, Pacific Grove, Calif, USA, October 1989.
- [23] M. J. Shensa, "Time constants and learning curves of LMS adaptive filters," Tech. Rep. 312, Naval Ocean Systems Center, San Diego, Calif, USA, 1978.
- [24] M. J. Shensa, "The spectral dynamics of evolving LMS adaptive filters," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '79)*, vol. 4, pp. 950–953, Washington, DC, USA, April 1979.
- [25] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communication*, Kluwer Academic Press, Boston, Mass, USA, 3rd edition, 2004.
- [26] J. P. Odenwalder, "Error control coding handbook," Tech. Rep., Linkabit, San Diego, Calif, USA, July 1976.
- [27] P. Monsen, "Adaptive equalization of the slow fading channel," *IEEE Transactions on Communications*, vol. 22, no. 8, pp. 1064–1075, 1974.
- [28] J. E. Smee and N. C. Beaulieu, "Error-rate evaluation of linear equalization and decision feedback equalization with error propagation," *IEEE Transactions on Communications*, vol. 46, no. 5, pp. 656–665, 1998.
- [29] J. G. Proakis, *Digital Communications*, McGraw-Hill, Boston, Mass, USA, 4th edition, 2001.
- [30] J. R. Zeidler, E. Satorius, D. Chabries, and H. Wexler, "Adaptive enhancement of multiple sinusoids in uncorrelated noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 3, pp. 240–254, 1978.
- [31] A. Dembo, "Bounds on the extreme eigenvalues of positive-definite Toeplitz matrices," *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 352–355, 1988.
- [32] F. Riera-Palou, J. M. Noras, and D. G. M. Cruickshank, "Analysis of the decision delay effect on the convergence of gradient recursive decision feedback equalizers," in *Proceedings of the IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP '02)*, vol. 3, pp. 2665–2668, Orlando, Fla, USA, May 2002.
- [33] J. G. Proakis and D. G. Manolakis, *Introduction to Digital Signal Processing*, MacMillian, New York, NY, USA, 1998.
- [34] D. T. M. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '94)*, vol. 4, pp. 585–588, Adelaide, Australia, April 1994.
- [35] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 694–705, 1997.
- [36] D. Gesbert and P. Duhamel, "Unbiased blind adaptive channel identification and equalization," *IEEE Transactions on Signal Processing*, vol. 48, no. 1, pp. 148–158, 2000.
- [37] J. R. Treichler, "Transient and convergent behavior of the adaptive line enhancer," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 1, pp. 53–62, 1979.
- [38] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals & Systems*, Prentice-Hall, Upper Saddle River, NJ, USA, 2nd edition, 1997.
- [39] A. Batra, J. R. Zeidler, and A. A. Beex, "Mitigation of narrowband interference using adaptive equalizers," in *Proceedings of the 14th European Signal Processing Conference (EUSIPCO '06)*, Florence, Italy, September 2006.
- [40] S. G. Sankaran, "Implementation and evaluation of echo cancellation algorithms," M.S. thesis, Virginia Tech, Blacksburg, VA, USA, December 1996.