

Stable Pole–Zero Modeling of Long FIR Filters with Application to the MMSE–DFE

Naofal Al-Dhahir, Ali H. Sayed, and John M. Cioffi

Abstract—The problem of approximating a long FIR filter by a reduced-parameter stable pole–zero filter is addressed in this paper. We derive a computationally efficient order-recursive algorithm that achieves this task with high accuracy. Our main emphasis is on applying this algorithm to reduce the implementation complexity of the decision feedback equalizer’s long FIR feedforward and feedback filters encountered in high-speed data transmission on digital subscriber loops.

Index Terms—Decision-feedback equalizer, digital subscriber lines, pole–zero models.

I. INTRODUCTION

IN A VARIETY of applications, the engineer is faced with the task of implementing a long finite-impulse response (FIR) filter on a digital signal processor. Although in some situations the underlying true response is of infinite length (IIR), assuming it to be FIR (e.g., by truncating up to the most significant N samples) has some advantages. Among those advantages are ease of computation (e.g., through the use of efficient time-domain algorithms such as the Levinson algorithm or frequency-domain algorithms such as the FFT algorithm) and a lower sensitivity to finite-precision effects. However, to achieve satisfactory performance, a large number of FIR filter taps is usually needed. This can lead to a prohibitive implementation cost in terms of the increased memory needed to store the filter taps and the large record of previous input samples, in addition to the high processing power required to compute the filter output samples through sum-of-products calculations. This cost even multiplies for high-speed applications and time-varying environments where the filter taps are frequently updated.

In this paper, we study the generic problem of approximating a long FIR filter by a pole–zero filter with a much smaller total (numerator and denominator) number of coefficients, with a special focus on the decision feedback equalization application. Although this problem was also studied in [4], our approach differs significantly from that of [4] in several

Paper approved by E. Eleftheriou, the Editor for Equalization and Coding of the IEEE Communications Society. Manuscript received October 13, 1995; revised June 21, 1996 and November 22, 1996. This work was supported by the National Science Foundation under Contract NCR-9203131, Joint Services Electronic Program under Contract DAAL 03-91-C-0010, and NASA under Contract NAG2-842. This paper was presented in part at the Allerton Conference on Communication, Control, and Computing (Allerton ’93), IL, September 1993.

N. Al-Dhahir is with GE Corporate R&D Center, Schenectady, NY 12309 USA.

A. H. Sayed is with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA.

J. M. Cioffi is with the Information Systems Laboratory, Stanford University, Stanford, CA 94305 USA.

Publisher Item Identifier S 0090-6778(97)03727-6.

aspects. First, the tail of a long FIR filter was assumed in [4] to be accurately modeled by two poles only. This assumption is specific to the high-bit-rate digital subscriber loop (HDSL) environment considered in [4]. Although we shall also present simulation results from the HDSL environment, our algorithm is quite general and does not make any such assumptions. Second, precursor ISI was assumed to be negligible in [4], and hence the feedforward filter was taken to be a short FIR filter. We do not make this assumption either since, for higher data rates and less benign channel characteristics, as in the asymmetric digital subscriber loop (ADSL) environment, the feedforward filter must be very long to achieve satisfactory performance. Therefore, we shall attempt to approximate both the feedforward and feedback filters by pole–zero models. Finally, the DFE coefficients were computed in [4] using adaptive IIR algorithms. Again, the environment-specific assumption of a two-pole model made stability monitoring a simple task, which would not be the case in situations where a two-pole model is not adequate (e.g., in echo cancellation). Instead, we shall compute the DFE coefficients directly from the available channel and noise estimates using the efficient algorithms of [1]. In case of environment changes, straightforward adaptation is performed on the long FIR filter, which is then converted to a pole–zero filter for a reduced-complexity implementation.

II. THE GENERALIZED ARMA–LEVINSON ALGORITHM

We shall start by deriving a *new* algorithm for approximating a long FIR filter by a pole–zero *stable* filter with many fewer taps. This algorithm is a generalization of the ARMA–Levinson algorithm derived in [9] using the “embedding” technique of [7]. The novelty in our algorithm is *its ability to relax the restriction of equal numbers of poles and zeros* that was imposed in [9], [7]. This flexibility will prove to be useful in obtaining better fits, as will be demonstrated by the simulation results of Section III.

The output samples of the long FIR filter are given by $y_k = \sum_{i=0}^L h_i x_{k-i}$. We want to approximate this input–output relationship by that of an autoregressive moving-average (ARMA) model with p poles and q zeros, denoted in this paper by ARMA(p, q), whose output samples \hat{y}_k are given by

$$\hat{y}_k = -a_1 \hat{y}_{k-1} - a_2 \hat{y}_{k-2} - \cdots - a_p \hat{y}_{k-p} + n_0 x_k + n_1 x_{k-1} + \cdots + n_q x_{k-q}. \quad (1)$$

For brevity, we shall assume that $p \geq q$.

Our objective is to estimate the $(p + q + 1)$ ARMA parameters $\{a_1, \dots, a_p, n_0, \dots, n_q\}$ based on knowledge of the second-order statistics of the input and output sequences

$\{y_k, x_k\}$. If we denote these estimates at the j th recursion ($1 \leq j \leq p$) by $\{a_1^j, \dots, a_p^j, n_0^j, \dots, n_{j-\delta}^j\}$ where $\delta \triangleq p - q \geq 0$, then we can relate the output samples of an ARMA($j, j - \delta$) filter to those of the original FIR filter as follows:

$$y_k \stackrel{\text{def}}{=} \hat{y}_k + e_k^j \\ = -a_1^j \hat{y}_{k-1} - a_2^j \hat{y}_{k-2} - \dots - a_p^j \hat{y}_{k-j} + n_0^j x_k \\ + n_1^j x_{k-1} + \dots + n_{j-\delta}^j x_{k-j+\delta} + e_k^j \quad (2)$$

where e_k^j is the j th-order residual error sequence and $n_i^j = 0$ for $i < 0$. In words, at the j th recursion, we determine an estimate of an ARMA model in which the difference between the number of poles and zeros is also δ , i.e., equal to the desired difference $p - q$.

Assuming $\{x_k\}$ to be white and that $\hat{y}_{k-i} = y_{k-i}$ for $1 \leq i \leq p$ (i.e., that previous estimates have been accurate) and defining the augmented vector $\mathbf{z}_k \stackrel{\text{def}}{=} \begin{bmatrix} y_k \\ x_{k+\delta} \end{bmatrix}$, then the ARMA model of (2) can be converted to the two-channel autoregressive (AR) model:

$$\mathbf{z}_k = -\Theta_1^j \mathbf{z}_{k-1} - \dots - \Theta_{\delta-1}^j \mathbf{z}_{k-\delta+1} \\ - \Theta_\delta^j \mathbf{z}_{k-\delta} - \dots - \Theta_j^j \mathbf{z}_{k-j} + \begin{bmatrix} e_k^j \\ x_{k+\delta} \end{bmatrix} \\ \stackrel{\text{def}}{=} \hat{\mathbf{z}}_k + \mathbf{e}_k$$

where we have defined the (matrix) prediction coefficients of the linear predictor $\hat{\mathbf{z}}_k$ as follows:

$$-\Theta_i^j \stackrel{\text{def}}{=} \begin{cases} \begin{bmatrix} -a_i^j & 0 \\ 0 & 0 \end{bmatrix}, & 1 \leq i \leq \delta - 1 \\ \begin{bmatrix} -a_i^j & n_{i-\delta}^j \\ 0 & 0 \end{bmatrix}, & \delta \leq i \leq j. \end{cases} \quad (3)$$

Using the *orthogonality principle* which states that $E[(\mathbf{z}_k - \hat{\mathbf{z}}_k) \mathbf{z}_{k-i}^*] = \mathbf{0}$ where $i = 1, 2, \dots, j$ and $(\cdot)^*$ denotes the complex-conjugate transpose, we get

$$E \left[\left(\mathbf{z}_k + \sum_{m=1}^j \Theta_m^j \mathbf{z}_{k-m} \right) \mathbf{z}_{k-i}^* \right] = \mathbf{0} \quad (4)$$

or, equivalently,

$$\mathbf{R}(i) + \Theta_1^j \mathbf{R}(i-1) + \dots + \Theta_j^j \mathbf{R}(i-j) = \mathbf{0}, \quad 1 \leq i \leq j \quad (5)$$

where we have defined

$$\mathbf{R}(i) \stackrel{\text{def}}{=} E[\mathbf{z}_k \mathbf{z}_{k-i}^*] = \begin{bmatrix} R_{yy}(i) & R_{yx}(i-\delta) \\ R_{xy}(i+\delta) & R_{xx}(i) \end{bmatrix} = \mathbf{R}^*(-i). \quad (6)$$

For $i = 0$, we have

$$E[(\mathbf{z}_k - \hat{\mathbf{z}}_k) \mathbf{z}_k^*] = E \left[\left(\mathbf{z}_k + \sum_{m=1}^j \Theta_m^j \mathbf{z}_{k-m} \right) \mathbf{z}_k^* \right] \\ = \mathbf{R}(0) + \sum_{m=1}^j \Theta_m^j \mathbf{R}(-m) \\ = E[\mathbf{e}_k (\mathbf{z}_k - \hat{\mathbf{z}}_k)^*] = E[\mathbf{e}_k \mathbf{e}_k^*] \stackrel{\text{def}}{=} \Sigma_j^f.$$

Therefore, we have

$$\mathbf{R}(0) + \Theta_1^j \mathbf{R}(-1) + \dots + \Theta_j^j \mathbf{R}(-j) = \Sigma_j^f, \quad i = 0. \quad (7)$$

Alternatively, (5) and (7) can be written in matrix form as follows:

$$\begin{bmatrix} \Theta_j^j & \dots & \Theta_1^j & \mathbf{I} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{R}(0) & \dots & \mathbf{R}(-(j-1)) & \mathbf{R}(-j) \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{R}(j-1) & \dots & \mathbf{R}(0) & \mathbf{R}(-1) \\ \mathbf{R}(j) & \dots & \mathbf{R}(1) & \mathbf{R}(0) \end{bmatrix} \\ = [\mathbf{0} \quad \dots \quad \mathbf{0} \quad \Sigma_j^f], \quad (8)$$

or, more compactly

$$\Theta^j \mathbf{R}^j = [\mathbf{0} \quad \dots \quad \mathbf{0} \quad \Sigma_j^f].$$

Equation (8) describes a j th-order AR model whose vector parameters Θ_i^j can be calculated by solving a block-Toeplitz Hermitian system of linear equations. This can be done efficiently using the following multichannel form of the famed Levinson algorithm, sometimes known as the Levinson-Wiggins-Robinson (LWR) algorithm [5], [6].

Algorithm (Generalized ARMA-Levinson): Given $\{\mathbf{R}(0), \dots, \mathbf{R}(p)\}$,

Initial Conditions: $\Theta_1^1 = \mathbf{K}_1^f = -\mathbf{R}(1)\mathbf{R}^{-1}(0)$; $\Phi_1^1 = \mathbf{K}_1^b = -\mathbf{R}(-1)\mathbf{R}^{-1}(0)$; $\Sigma_1^f = (\mathbf{I} - \mathbf{K}_1^f \mathbf{K}_1^b) \mathbf{R}(0)$; $\Sigma_1^b = (\mathbf{I} - \mathbf{K}_1^b \mathbf{K}_1^f) \mathbf{R}(0)$.

Recursions: For $1 \leq j \leq p-1$, see (9) at the bottom of the page, where the backward prediction vector

$$\Phi^j \stackrel{\text{def}}{=} [\mathbf{I} \quad \Phi_1^j \quad \dots \quad \Phi_j^j]$$

satisfies the following auxiliary block-Toeplitz system of equations:

$$\Phi^j \mathbf{R}^j = [\Sigma_j^b \quad \mathbf{0} \quad \dots \quad \mathbf{0}].$$

Assuming the input sequence to be *white* (i.e., $R_{xx}(l) = S_x \delta_l$), then the output autocorrelation and input-output cross-correlation sequences needed to form (6) can be computed using knowledge of the FIR filter taps as follows:

$$R_{yy}(l) = S_x \sum_{m=0}^{\nu} h_m h_{m-l}^*; \quad R_{yx}(l) = S_x h_l = R_{xy}^*(-l).$$

The parameters $\{a_1^p, \dots, a_p^p, n_0^p, n_1^p, \dots, n_q^p\}$ are read off directly from Θ_i^p ($1 \leq i \leq p$).

Remarks:

- 1) Provided that $\{x_k\}$ is white, we can show that the Θ_i^j 's generated by the recursions of (9), indeed follow the form of (3), i.e., they have a second row of all zeros, using *induction* as follows. First, we show that the initial condition Θ_1^1 has the desired form. Then, we assume it for Θ_i^j and show that it holds for Θ_i^{j+1} (where $1 \leq i \leq j$).

Since $\Theta_1^1 = -\mathbf{R}(1)\mathbf{R}(0)^{-1}$, it suffices to show that the second row of $\mathbf{R}(1)$ consists of all zeros. This follows directly from (6) since $R_{xx}(1) = 0$ and $R_{xy}(1 + \delta) = S_x h_{-(1+\delta)}^* = 0$ (since $\delta \geq 0$ and $\{h_k\}$ is causal).

Now, assume that Θ_i^j has a second row of all zeros; then, using the recursion

$$\begin{aligned}\Theta_i^{j+1} &= \Theta_i^j + \mathbf{K}_{j+1}^f \Phi_{j-i+1}^j \\ &= \Theta_i^j - \Delta_{j+1}^f (\Sigma_j^b)^{-1} \Phi_{j-i+1}^j\end{aligned}$$

we need only show that Δ_{j+1}^f has an all-zero second row. This follows from the recursion

$$\Delta_{j+1}^f = \mathbf{R}(j+1) + \sum_{m=1}^j \mathbf{R}(m) \Theta_{j-m+1}^j$$

and the fact that $\mathbf{R}(i)$ has a second row of all zeros for $i \geq 1$ [cf. (6)]. Note that since $\Theta_{j+1}^{j+1} = \mathbf{K}_{j+1}^f$, it also has this same special structure.

- 2) The diagonal nature of Σ_j^f for $\delta > 0$ can be shown from (2) by using the assumed whiteness of $\{x_k\}$ and causality of $\{h_k\}$.
- 3) The algorithm uses both forward and backward reflection coefficient matrices. By defining a suitable *normalization*, it is possible to perform the order update using a single reflection coefficient matrix [5].
- 4) The generated pole-zero models are *guaranteed* to be stable. Next, we shall outline a proof of this fact. Define the j th-order predictor polynomial matrix shown in the equation at the bottom of the page. It is a well-known result (see [3, Theorem 5.1]) that, since $\Theta^j(D)$ is generated by the multichannel Levinson algorithm, its determinant, which is equal to $a(D)$, is guaranteed to be stable.
- 5) The algorithm does not restrict the generated pole-zero approximations to be minimum phase (i.e., the zeros of the numerator polynomial could lie inside the unit circle in the $D \stackrel{\text{def}}{=} z^{-1}$ domain). This result can be explained as follows. The numerator coefficients are read off directly from the matrix coefficients Θ_i^j computed using the LWR algorithm. The only constraint on Θ_i^j is that $\det(\Theta^j(D))$ is stable. This determinant is *independent* of the numerator polynomial $n(D)$ because of the upper triangular nature of $\Theta^j(D)$ (see Remarks 1) and 4) above).

- 6) The proposed algorithm is a two-channel Levinson algorithm, which is known to have a computational complexity of $O(2(\max(p, q))^2)$ operations [6], [5]. Furthermore, the algorithm can be implemented in a *normalized* form. This alternative form involves *orthogonal rotations* only, which have desirable numerical and computational properties.
- 7) A multichannel least squares algorithm with a different number of parameters in each channel was previously derived in [8]. However, the algorithm of [8] is distinct from ours in that it is of lattice type, time-recursive, and has a computational complexity on $O(36 \max(p, q))$ which is higher than the complexity of our algorithm when $\max(p, q) \leq 18$. Finally, the algorithm in [8] was not derived using the “embedding” approach that we follow here. In fact, the use of “embedding” to approximate long FIR filters by ARMA filters with unequal numbers of poles and zeros is new, to the best of our knowledge.
- 8) In choosing p and q , we assume a maximum allowable implementation complexity, which in turn sets an upper bound on their values. Therefore, we seek to find the best, in terms of low normalized norm tap error NNTTE $\stackrel{\text{def}}{=} (\sum_{i=0}^{\infty} |h_i - \hat{h}_i|^2 / \sum_{i=0}^{N-1} |h_i|^2)$, pole-zero approximation subject to this complexity constraint. It is worth emphasizing that increasing p and/or q could result in a worse approximation, depending on the FIR filter response.

III. SIMULATION RESULTS

The FIR MMSE-DFE consists of two filters: a feedforward filter $w(D)$ that combats precursor ISI and noise, and a strictly causal feedback filter $b(D)$ that suppresses postcursor ISI. The channel impulse response is assumed to be a linear time-invariant FIR filter with memory ν . For wide-band transmission on twisted copper lines, ν is very large, which entails the use of very long feedforward and feedback filters to achieve satisfactory performance. We shall apply the algorithm of Section II to convert these long FIR filters to pole-zero filters with many fewer parameters, without losing stability, and while still maintaining satisfactory performance. Other applications of the algorithm are discussed in [2].

$$\begin{aligned}\Delta_{j+1}^f &= \mathbf{R}(j+1) + \mathbf{R}(j) \Theta_1^j + \dots + \mathbf{R}(1) \Theta_j^j \\ \mathbf{K}_{j+1}^f &= -\Delta_{j+1}^f (\Sigma_j^b)^{-1}: \text{forward reflection coefficient matrix} \\ \mathbf{K}_{j+1}^b &= -\Delta_{j+1}^{*f} (\Sigma_j^f)^{-1}: \text{backward reflection coefficient matrix} \\ \Sigma_{j+1}^f &= (\mathbf{I} - \mathbf{K}_{j+1}^f \mathbf{K}_{j+1}^b) \Sigma_j^f: \text{forward prediction residual error matrix} \\ \Sigma_{j+1}^b &= (\mathbf{I} - \mathbf{K}_{j+1}^b \mathbf{K}_{j+1}^f) \Sigma_j^b: \text{backward prediction residual error matrix} \\ \Theta_i^{j+1} &= \Theta_i^j + \mathbf{K}_{j+1}^f \Phi_{j-i+1}^j, \quad 1 \leq i \leq j \\ \Theta_{j+1}^{j+1} &= \mathbf{K}_{j+1}^f \\ \Phi_i^{j+1} &= \Phi_i^j + \mathbf{K}_{j+1}^b \Theta_{j-i+1}^j, \quad 1 \leq i \leq j \\ \Phi_{j+1}^{j+1} &= \mathbf{K}_{j+1}^b\end{aligned}\tag{9}$$

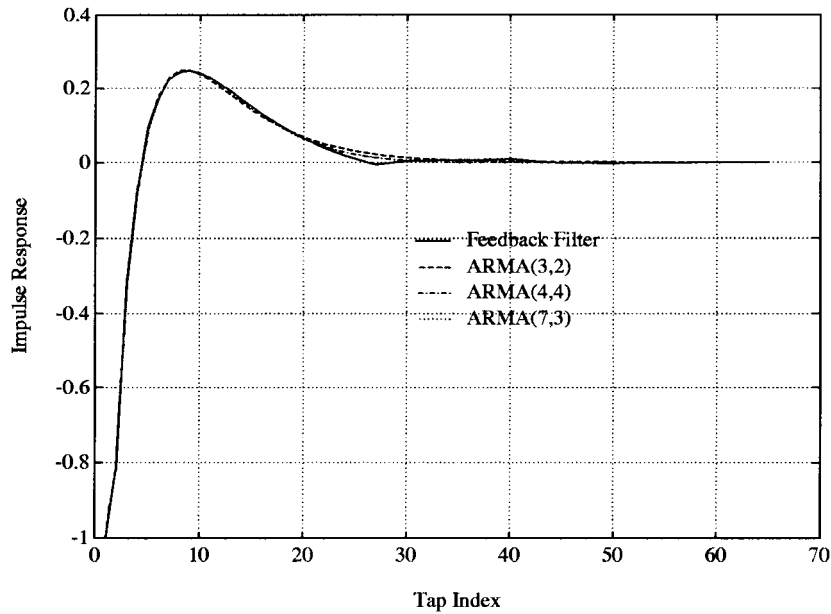


Fig. 1. Pole-zero approximations of the feedback filter.

 TABLE I
 ARMA APPROXIMATIONS FOR $\hat{b}(D)$ AND THEIR ACHIEVABLE NNTE'S

Model Order	$\hat{b}(D)$	NNTE(dB)
ARMA(3, 2)	$\frac{-1+0.5354D+0.5056D^2}{1-1.3501D+0.2757D^2+.1275D^3}$	-27.2946
ARMA(4, 4)	$\frac{-1+0.6731D+0.7541D^2-.2917D^3-.1145D^4}{1-1.4879D+0.1395D^2+0.5757D^3-.1944D^4}$	-31.1634
ARMA(7, 3)	$\frac{-1+.7325D+.572D^2-.2791D^3}{1-1.5472D+.3699D^2+0.3943D^3-.2299D^4+.0636D^5-.0157D^6+.0048D^7}$	-31.384

 TABLE II
 ARMA APPROXIMATIONS FOR $\hat{w}_1(D)$ AND THEIR ACHIEVABLE NNTE'S

Model Order	$\hat{w}_1(D)$	NNTE(dB)
ARMA(3, 2)	$\frac{47.1549-6.4777D-18.9055D^2}{1-1.0134D-.127D^2+.207D^3}$	-36.5522
ARMA(5, 2)	$\frac{47.1549-4.3601D-24.7598D^2}{1-.9685D-.2903D^2+.3231D^3-.0263D^4+.0166D^5}$	-41.9531
ARMA(6, 2)	$\frac{47.1549-6.5874D-24.4024D^2}{1-1.0157D-.2416D^2+.3091D^3-.0193D^4+.0131D^5+.0028D^6}$	-42.2905

We have found through extensive computer simulations that a direct reduced-parameter pole-zero approximation of the feedforward filter is very difficult to obtain. This is due to the fact that the initial part of the impulse response (IR) prior to the peak could be very long. Pole-zero models can more easily model the decaying tail of the IR following the peak. With this observation in mind, we proposed in [2] to “split” the IR of the feedforward filter at its peak sample into two components, and to approximate each component separately by a pole-zero model.

For the HDSL environment, we shall assume the worst case 9 kft 26 gauge DSL. The input power level is 17 dBm, evenly distributed over the transmission bandwidth, and the two-sided AWGN power spectral density (psd) is taken to be -113 dBm/Hz. The standard $|H_x(f)|^2 = k_{\text{NEXT}}f^{3/2}$ near-end crosstalk (NEXT) model is adopted with $k_{\text{NEXT}} = 10^{-13}$. The target bit rate is set at 800 kbits/s, and the input signal constellation is 16-QAM, which is near optimum in this case. A fixed probability of error $P_e = 10^{-7}$ is assumed and a 4.2 coding gain is included. The finite-length MMSE-DFE is

$$\begin{aligned} \Theta^j(D) &\stackrel{\text{def}}{=} I + \Theta_1^j D + \dots + \Theta_j^j D^j \\ &= \begin{bmatrix} (1 + a_1^j D + \dots + a_j^j D^j) & -D^\delta (n_0^j + n_1^j D + \dots + n_{j-\delta}^j D^{j-\delta}) \\ 0 & 1 \end{bmatrix} \\ &\stackrel{\text{def}}{=} \begin{bmatrix} a(D) & -n(D) \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

TABLE III
ARMA APPROXIMATIONS FOR $w_2(D)$ AND THEIR ACHIEVABLE NNTE'S

Model Order	$\hat{w}_2(D)$	NNTE(dB)
ARMA(2,1)	$\frac{-6.0337+4348D}{1-.5621D-.3772D^2}$	-26.0219
ARMA(3,3)	$\frac{-6.0337+2.7584D+.6662D^2-.3245D^3}{1-.9472D-.2917D^2+.2708D^3}$	-26.4864
ARMA(6,6)	$\frac{-6.0337+2.5235D+.6947D^2+.7904D^3+.7084D^4+.2405D^5-.1915D^6}{1-.9082D-.3155D^2+.0725D^3-.0224D^4+.0965D^5+.0914D^6}$	-30.4245

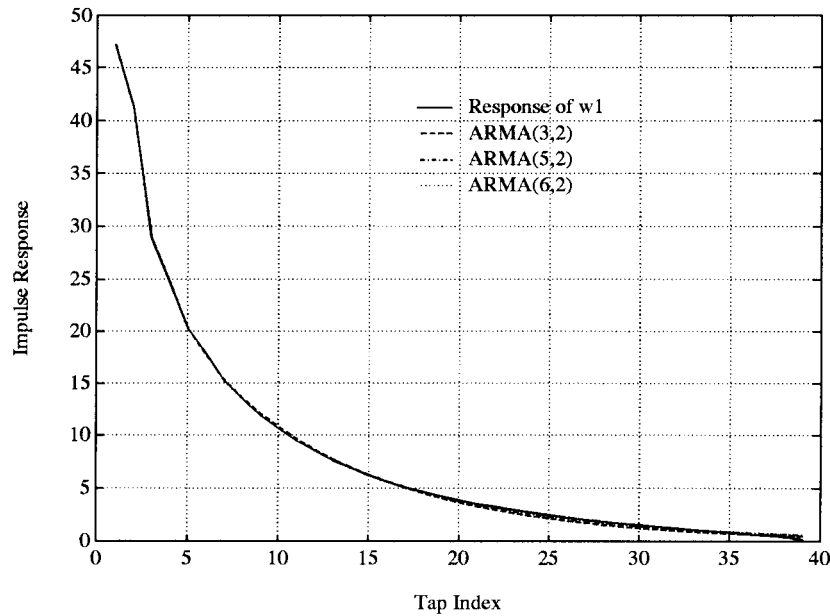


Fig. 2. Pole-zero approximations for $w_1(D)$ (component of $w(D)$ after the peak sample). Note that the first sample of $w_1(D)$ is equal to 58th sample of $w(D)$.

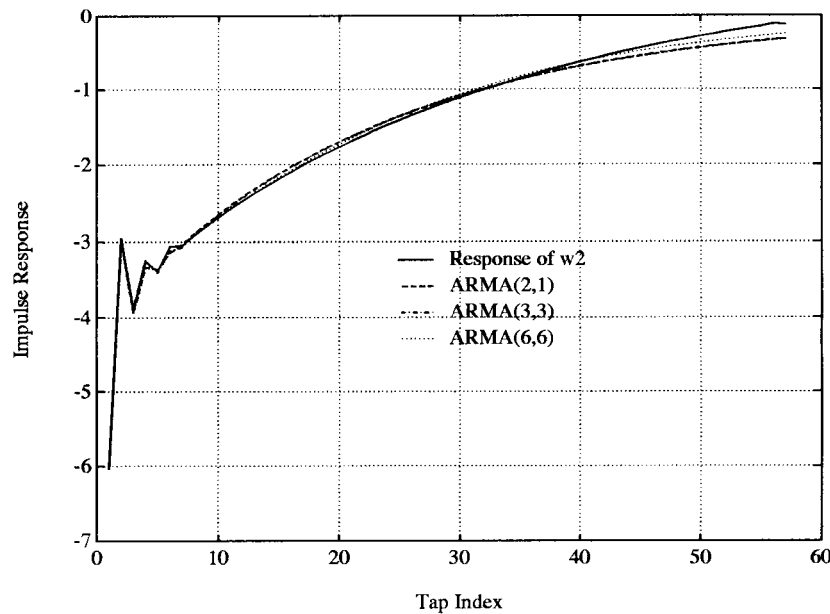


Fig. 3. Pole-zero approximations for $w_2(D)$ (component of $w(D)$ before the peak sample and time reversed). Note that the first sample of $w_2(D)$ is equal to the 57th sample of $w(D)$.

assumed to have 96 feedforward taps and 64 feedback taps. This choice results in an operational margin of around 4.9 dB.

In Tables I-III, we present some of the best pole-zero approximations of the 64-tap feedback filter and the two components of the 96-tap feedforward filter and their cor-

responding NNTE's. The impulse responses of those approximations together with that of the desired response are given in Figs. 1–3. It is evident that the proposed algorithm generates fairly accurate pole–zero approximations with the added advantages of significant reductions in the number of filter coefficients, fast computation, and guaranteed stability.

In concluding this section, it is worth mentioning that we have investigated the effect of pole–zero modeling on the performance of the MMSE–DFE, as measured by the operational margin at $P_e = 10^{-7}$. For example, we found that using the four-pole, four-zero model given in Table I for the 64-tap feedback filter, together with the five-pole, two-zero and the two-pole, one-zero models given in Tables II and III for the two components of the 96-tap feedforward filter results in a negligible margin reduction of 0.05 dB. This further confirms the accuracy of the pole–zero approximations to the original FIR filters.

IV. CONCLUSION

In this paper, we derived a computationally efficient algorithm that accurately approximates long FIR filters by stable pole–zero filters with far fewer coefficients. The algorithm was successfully applied to the problem of reducing the implementation complexity of the MMSE–DFE's long FIR feedforward and feedback filters on digital subscriber lines.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their thoughtful comments and suggestions.

REFERENCES

- [1] N. Al-Dhahir and J. M. Cioffi, "Fast computation of channel-estimate based equalizers in packet data transmission," *IEEE Trans. Signal Processing*, vol. 43, pp. 2462–2473, Nov. 1995.
- [2] N. Al-Dhahir, A. H. Sayed, and J. M. Cioffi, "A high-performance cost-effective pole–zero MMSE-DFE," in *Proc. Allerton Conf. Commun., Contr., Computing*, Sept. 1993, pp. 1166–1175.
- [3] D. Alpay and I. Gohberg, "On orthogonal matrix polynomials," *Oper. Theory: Adv. and Appl.*, vol. 34, pp. 25–46, 1988.
- [4] P. Crespo and M. Honig, "Pole–zero decision feedback equalization with a rapidly converging adaptive IIR algorithm," *IEEE J. Select. Areas Commun.*, vol. 9, pp. 817–829, Aug. 1991.
- [5] T. Kailath, "Linear estimation for stationary and near-stationary processes," in *Lecture Notes for Arab School on Science and Technology*. (Summer School on Signal Processing). Hemisphere Publishing, 1983.
- [6] S. Kay, *Modern Spectral Estimation: Theory and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
- [7] D. Lee, B. Friedlander, and M. Morf, "Recursive ladder algorithms for ARMA modeling," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 753–764, Aug. 1982.
- [8] F. Ling and J. Proakis, "A generalized multichannel least squares lattice algorithm based on sequential processing stages," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 381–389, Apr. 1984.
- [9] J. Tu, "Theory, design, and application of multi-channel modulation for digital communications", Ph.D. dissertation, Stanford Univ., Stanford, CA, June 1991.