Impulse Response Shortening for Discrete Multitone Transceivers

Peter J. W. Melsa, *Member, IEEE*, Richard C. Younce, *Member, IEEE*, and Charles E. Rohrs, *Senior Member, IEEE*

***Abstract*—In discrete multitone (DMT) transceivers an intelligent guard time sequence, called a cyclic prefix (CP), is inserted between symbols to ensure that samples from one symbol do not interfere with the samples of another symbol. The length of the CP is determined by the length of the impulse response of the effective physical channel. Using a long CP reduces the throughput of the transceiver. To avoid using a long CP, a short time-domain finite impulse response (FIR) filter is used to shorten the effective channels impulse response.**

This paper explores various methods of determining the coefficients for this time-domain filter. An optimal shortening and a least-squares (LS) approach are developed for shortening the channel's impulse response. To provide a computationally efficient algorithm a variation of the LS approach is explored. In full-duplex transceivers the length of the effective echo path impacts the computational requirements of the transceiver. A new paradigm of joint shortening is introduced and three methods are developed to jointly shorten the channel and the echo impulse responses in order to reduce the length of the CP and reduce computational requirements for the echo canceller.

I. Introduction

In discrete multitone (DMT) transceivers each symbol is comprised of samples to be transmitted to the remote receiver plus a cyclic prefix (CP) of length ν [1]. The CP is simply the last ν samples of the original N samples to be transmitted. The CP length is determined by the length of the channel's impulse response, and is chosen to minimize intersymbol interference (ISI). At the receiver the CP is discarded, the remaining N samples are then processed by the receiver. If the impulse response of the channel is of length $\nu + 1$ or shorter then a CP of length ν is sufficient to eliminate ISI. Since the efficiency of the transceiver is reduced by a factor of $N/N + \nu$ it is either desirable to make ν as small as possible or utilize a large N. Increasing N increases the computational complexity, system delay, and memory requirements of the transceiver. Additionally, the length of the channel's impulse response is typically not under the control of the designer and varies from channel to channel, leaving the designer to choose both a large ν and large N in order to achieve reasonable efficiency. To alleviate these problems a short time-domain FIR filter, referred to here as a shortened impulse response filter (SIRF), is typically placed in the receiver immediately following the analog-to-digital (A/D) converter. The purpose of this filter is to shorten the impulse response of the effective channel. The effective channel is the convolution of transmit filters, physical channel, receive filters, and the SIRF. The impulse response of the effective channel needs to be shorter than the length of the CP. The length of the SIRF and CP are usually fixed *a priori* and not changed from channel to channel.

This paper addresses the calculation of the coefficients for the SIRF. It was shown in [2] that the greatest throughput, for a given computational complexity, of a DMT transceiver can be accomplished using an equalizer that shortens the impulse response of the channel and a CP. This approach was demonstrated for a DMT based high-speed digital subscriber line (HDSL) transceiver in [1]. In [3], an efficient means for determining the coefficients of the SIRF was developed. However, the algorithms in [3] are hampered by stability and convergence problems, and [2] fails to present an efficient algorithm. This paper develops computationally efficient algorithms that can achieve impulse response shortening unhindered by stability problems and are realizable in real-time using off-the-shelf digital signal processing (DSP) hardware.

When the transceiver is used in a full-duplex environment the portion of the receive signal arising from echo must be reduced through the use of echo cancellation beyond the loss provided by the hybrid. Echo canceller structures for DMT have been discussed previously [5]–[10], but the important point to note is that the complexity is directly related to the length of the echo path impulse response [6], [9]. If the length of the SIRF is chosen appropriately, the echo path impulse response can be explicitly shortened without reducing the effectiveness of the channel shortening.

The algorithms presented in this paper approach the problem of equalization from the perspective of impulse response shortening. It may be possible that this shortening measure is not the best measure of equalization since ultimately the channel SNR in each subsymbol of the multitone system is what is important. However, in practice and theory this approach provides adequate performance for most applications as we will demonstrate in Section IV.

There are other issues such as tracking slow channel variations, misadjustment resulting from finite training time, quantization effects, etc. which affect the real world application of these algorithms. We do not address these issues here but instead create a theoretical benchmark for optimal shortening

Paper approved by E. Eleftheriou, the Editor for Equalization and Coding of the IEEE Communications Society. Manuscript received June 6, 1995; revised April 2, 1996.

P. J. W. Melsa and C. E. Rohrs are with the Tellabs Research Center, Mishawaka, IN 46545 USA.

R. C. Younce is with Tellabs Operations Inc., Bolingbrook, IL 60440 USA.

Publisher Item Identifier S 0090-6778(96)09024-1.

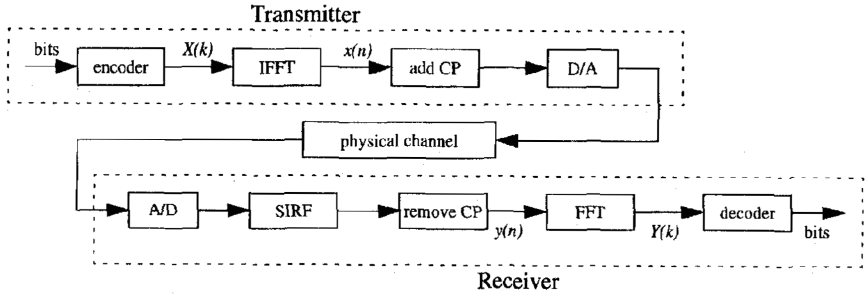

Fig. 1. Simplified DMT transceiver.

and introduce a new paradigm and method for joint channel and echo path shortening.

Section II-A develops the motivation for channel impulse response shortening for DMT transceivers. Section II-B develops a new algorithm based on optimal shortening of an impulse response. Given the original impulse response, an SIRF length, and a CP length the optimal algorithm finds the coefficients which generate the shortest impulse response. Section II-C uses a more traditional least-squares (LS) approach to fit a pole-zero model to the impulse response. By choosing an appropriate numbers of poles and an appropriate numbers of zeros, the resulting poles can be used as the SIRF coefficients to shorten the effective impulse response Section II-D restates ideas from [4] and [12] as a more efficient means of realizing the algorithm developed in Section II-C. Section II-E presents some observations on applying the algorithms presented in Sections II-B, II-C, and II-D. Section III presents a new concept where both the channel and echo impulse responses are simultaneously shortened. The algorithms presented in Section II are expanded and reformulated for this new "joint shortening" model. Both an optimal and a computationally efficient algorithm are derived that achieve significant echo impulse response shortening while maintaining channel shortening performance. Finally, some simulation results are reported in Section IV, and conclusions are drawn in Section V.

II. Channel Shortening

A. Background

A simplified DMT transceiver (without echo cancellation) is shown in Fig. 1. The encoder maps incoming bits onto $N/2+1$ carrier frequencies forming the complex frequency domain subsymbols $X(k)$ where $k = 0, 1, \cdots, N/2$. The $N/2+1$ subsymbols taken together form a frequency domain symbol. A conjugate symmetry condition is then imposed creating a length N frequency domain signal. This signal is inverse Fourier transformed to generate a real, length N time domain symbol, $x(n)$, which is prefixed with a CP of length ν and transmitted to the far end receiver. The CP consists of the last ν samples of $x(n)$. At the receiver, the receive signal is passed through a short time-domain FIR filter, and then the CP is removed. This signal, $y(n)$ is then Fourier transformed back to the frequency domain. Each of the frequency domain subsymbols $Y(k)$ can then be decoded to an outgoing bit stream.

In Fig. 1, the physical channel is considered to be the combination of the actual cable and any filtering done in the transmit or receive analog circuitry. Denoting the impulse response of the physical channel by $h(n)$, the output of the SIRF can be expressed as

$$y(n) = (h(n) * w(n)) * x(n) = h_{\text{eff}}(n) * x(n) \qquad (1)$$

where $w(n)$ is the impulse response of the SIRF and $*$ denotes the convolution operator. It can be shown that if $x(n)$ is transmitted with a CP of length ν and the effective channel, $h_{\text{eff}}(n)$, is at most of length $\nu+1$, each $y(n)$ in the current symbol will only depend upon $x(n)$'s currently being transmitted. If the effective channel is not constrained to a length of $\nu+1$, then $y(n)$ will depend upon the current symbol's time samples $x(n)$ as well as the previous symbol's time samples $x_p(n)$. The previous transmit symbol $x_p(n)$ will then contribute ISI which presents itself as noise to the slicer and so decreases the performance of the transceiver.

Regardless of the choice of $w(n)$ it is generally not possible to shorten the impulse response perfectly. Some energy will lie outside the largest $\nu+1$ consecutive samples of $h_{\text{eff}}(n)$. As a measure of the ISI we can measure the shortening SNR (SSNR) which is the ratio of the energy in the largest consecutive $\nu+1$ samples to the energy in the remaining samples. The largest $\nu+1$ samples will not necessarily start with the first sample. This delay, d, is normally compensated for at the receiver by delaying the start of the receive symbol. Since $h(n)$ is not under the control of the designer, we use the coefficients of the SIRF, $w(n)$, to limit the length of the effective channel's impulse response, thereby minimizing ISI. The shape of the resulting impulse response of the effective channel, $h_{\text{eff}}(n)$, is usually unimportant, what is important is that the SSNR be maximized.

A necessary step in any algorithm is estimating the channel and echo impulse responses. Typically, DMT transceivers use known training sequences for purposes of adapting various parameters associated with the transmitter and receiver. Since DMT transceivers have the frequency response naturally available, the training sequences can be used to calculate the frequency responses of the channel and echo path. The frequency responses can then be inverse transformed into the time domain, generating the time-domain impulse responses. Care should be taken to reduce the variance of the noise in the estimate, as well as compensate for any other known effects such as aliasing.

Our goal in this paper is to develop algorithms for impulse response shortening that result in sufficient shortening while being realizable in off-the-shelf DSP chips.

B. Optimal Shortening

In this section, we develop a new algorithm for the optimal shortening of an impulse response. The algorithm utilizes eigenvalues and eigenvectors to generate the coefficients of an optimal shortening filter, given the original impulse response, the CP length and the SIRF length.

We can rewrite (1) using such matrixes as (2) shown on the bottom of this page, and where M is the length of the physical channel's impulse response and t is the length of the SIRF. We wish to force as much of the effective channel's impulse response to lie in $\nu+1$ consecutive samples and as a consequence minimize ISI. Let $\boldsymbol{h}_{\text{win}}$ represent a window of $\nu+1$ consecutive samples of $\boldsymbol{h}_{\text{eff}}$ starting with sample d, and let wall $\boldsymbol{h}_{\text{wall}}$ represent the remaining $M+t-\nu-2$ samples of $\boldsymbol{h}_{\text{eff}}$. With these definitions

$$\boldsymbol{h}_{\text{win}} = \begin{bmatrix} h_{\text{eff}}(d) \\ h_{\text{eff}}(d+1) \\ \vdots \\ h_{\text{eff}}(d+\nu) \end{bmatrix} = \begin{bmatrix} h(d) & h(d-1) & \cdots & h(d-t+1) \\ h(d+1) & h(d) & \cdots & h(d-t+2) \\ \vdots & & \ddots & \vdots \\ h(d+\nu) & h(d+\nu-1) & \cdots & h(d+\nu-t+1) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \\ \cdots \\ w(t-1) \end{bmatrix} \equiv \boldsymbol{H}_{\text{win}}\boldsymbol{w} \tag{3}$$

$$\boldsymbol{h}_{\text{wall}} = \begin{bmatrix} h_{\text{eff}}(0) \\ \vdots \\ h_{\text{eff}}(d-1) \\ h_{\text{eff}}(d+\nu+1) \\ \vdots \\ h_{\text{eff}}(M+t-2) \end{bmatrix} = \begin{bmatrix} h(0) & 0 & \cdots & 0 \\ \vdots & \ddots & & \\ h(d-1) & h(d-2) & \cdots & h(d-t) \\ h(d+\nu+1) & h(d+\nu) & \cdots & h(d+\nu-t+2) \\ \vdots & \ddots & & \\ 0 & \cdots & 0 & h(M-1) \end{bmatrix} \cdot \begin{bmatrix} w(0) \\ w(1) \\ \cdots \\ w(t-1) \end{bmatrix} \equiv \boldsymbol{H}_{\text{wall}}\boldsymbol{w}. \tag{4}$$

Optimal shortening can be expressed as choosing w to minimize $\boldsymbol{h}_{\text{wall}}^T\boldsymbol{h}_{\text{wall}}$ while satisfying the constraint: $\boldsymbol{h}_{\text{win}}^T\boldsymbol{h}_{\text{win}} = 1$. Constraining the energy in the window ensures that the trivial solution, $\boldsymbol{w} = [0\cdots 0]^T$, is disallowed. The value of one is picked arbitrarily since $\boldsymbol{w}$ can be scaled arbitrarily without changing the energy ratio. The expressions for the energy outside and inside the window can be written as

$$\boldsymbol{h}_{\text{wall}}^T\boldsymbol{h}_{\text{wall}} = \boldsymbol{w}^T\boldsymbol{H}_{\text{wall}}^T\boldsymbol{H}_{\text{wall}}\boldsymbol{w} = \boldsymbol{w}^T\boldsymbol{A}\boldsymbol{w} \tag{5}$$

$$\boldsymbol{h}_{\text{win}}^T\boldsymbol{h}_{\text{win}} = \boldsymbol{w}^T\boldsymbol{H}_{\text{win}}^T\boldsymbol{H}_{\text{win}}\boldsymbol{w} = \boldsymbol{w}^T\boldsymbol{B}\boldsymbol{w} \tag{6}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric and positive semidefinite matrixes. Optimal shortening can be considered as choosing $\boldsymbol{w}$ to minimize $\boldsymbol{w}^T\boldsymbol{A}\boldsymbol{w}$ while satisfying a constraint of $\boldsymbol{w}^T\boldsymbol{B}\boldsymbol{w} = 1$.

The following development assumes that $\boldsymbol{B}$ is invertible. The case where $\boldsymbol{B}$ is singular is more complex and is handled in the Appendix. Typical DMT transceivers use a SIRF length which is shorter than the length of the CP (i.e. $t<\nu$). The rows of $\boldsymbol{H}_{\text{win}}$ consist of shifted windows of the channel impulse

$$\boldsymbol{h}_{\text{eff}} = \boldsymbol{H}\boldsymbol{w} = \begin{bmatrix} h_{\text{eff}}(0) \\ h_{\text{eff}}(1) \\ \vdots \\ h_{\text{eff}}(M-1) \\ h_{\text{eff}}(M) \\ \vdots \\ h_{\text{eff}}(M+t-2) \end{bmatrix} = \begin{bmatrix} h(0) & 0 & \cdots & & 0 \\ h(1) & h(0) & \ddots & & \\ \vdots & \vdots & & & \\ h(M-1) & h(M-2) & \cdots & h(M-t+1) & h(M-t) \\ 0 & h(M-1) & \cdots & & h(M-t+1) \\ \vdots & \ddots & & & \vdots \\ 0 & \cdots & & 0 & h(M-1) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \\ \vdots \\ w(t-1) \end{bmatrix} \tag{2}$$

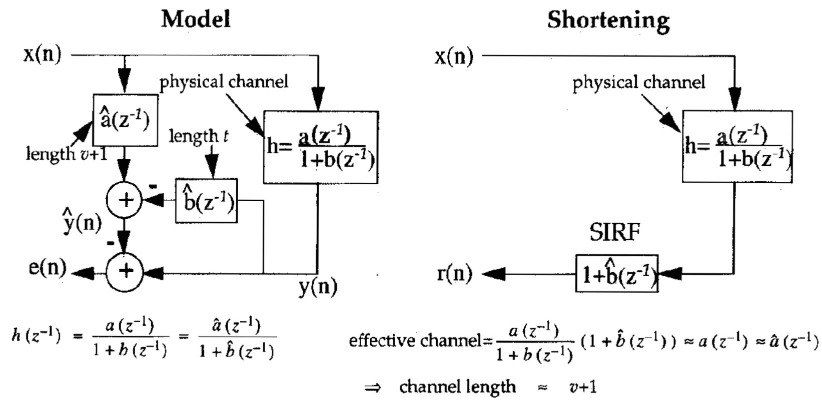

Fig. 2. LS approach to impulse shortening.

response and it has been found in practical applications (e.g. twisted-pair copper cables) that the rank $(\boldsymbol{H}_{\text{win}}) = t$. As a consequence, the matrix $\boldsymbol{B}$ of dimension $t \times t$ will be positive definite and can be decomposed using Cholesky Decomposition into

$$\begin{aligned}\boldsymbol{B} &= \boldsymbol{Q}\Lambda\boldsymbol{Q}^T = (\boldsymbol{Q}\sqrt{\Lambda})(\sqrt{\Lambda}\boldsymbol{Q}^T)\\ &= (\boldsymbol{Q}\sqrt{\Lambda})(\boldsymbol{Q}\sqrt{\Lambda})^T = \sqrt{\boldsymbol{B}}\sqrt{\boldsymbol{B}}^T \end{aligned} \tag{7}$$

where Λ is a diagonal matrix formed from the eigenvalues of $\boldsymbol{B}$, and the columns of $\boldsymbol{Q}$ are the orthonormal eigenvectors. Note that, since $\boldsymbol{B}$ is of full rank, the matrix $(\sqrt{\boldsymbol{B}})^{-1}$ exists.

In order to satisfy the constraint on $\boldsymbol{w}^T\boldsymbol{B}\boldsymbol{w}$ define

$$\boldsymbol{y} = \sqrt{\boldsymbol{B}}^T\boldsymbol{w} \tag{8}$$

so that

$$\boldsymbol{y}^T\boldsymbol{y} = \boldsymbol{w}^T\sqrt{\boldsymbol{B}}\sqrt{\boldsymbol{B}}^T\boldsymbol{w} = \boldsymbol{w}^T\boldsymbol{B}\boldsymbol{w} = 1. \tag{9}$$

Solving in (8) for $\boldsymbol{w}$

$$\boldsymbol{w} = (\sqrt{\boldsymbol{B}}^T)^{-1}\boldsymbol{y} \tag{10}$$

we find that

$$\boldsymbol{w}^T\boldsymbol{A}\boldsymbol{w} = \boldsymbol{y}^T(\sqrt{\boldsymbol{B}})^{-1}\boldsymbol{A}(\sqrt{\boldsymbol{B}}^T)^{-1}\boldsymbol{y} = \boldsymbol{y}^T\boldsymbol{C}\boldsymbol{y} \tag{11}$$

where $\boldsymbol{C}$ is appropriately defined. Optimal shortening can thus be considered as choosing y to minimize $\boldsymbol{y}^T\boldsymbol{C}\boldsymbol{y}$ while constraining $\boldsymbol{y}^T\boldsymbol{y} = 1$. The solution to this problem occurs for $\boldsymbol{y} = \boldsymbol{l}_{\min}$ where $\boldsymbol{l}_{\min}$ is the unit-length eigenvector corresponding to the minimum eigenvalue $\lambda_{\min}$ of $\boldsymbol{C}$. The resulting SIRF coefficients are thus

$$\boldsymbol{w}_{\text{opt}} = (\sqrt{\boldsymbol{B}}^T)^{-1}\boldsymbol{l}_{\min}. \tag{12}$$

Substituting (7) into (11) we find

$$\boldsymbol{C} = (\boldsymbol{Q}\sqrt{\Lambda})^{-1}\boldsymbol{A}(\sqrt{\Lambda}\boldsymbol{Q}^T)^{-1}. \tag{13}$$

The shortening SNR can be expressed as

$$\begin{aligned}\text{SSNR}_{\text{opt}} &= 10\log\left(\frac{\boldsymbol{w}_{\text{opt}}^T\boldsymbol{B}\boldsymbol{w}_{\text{opt}}}{\boldsymbol{w}_{\text{opt}}^T\boldsymbol{A}\boldsymbol{w}_{\text{opt}}}\right) = 10\log\left(\frac{1}{\lambda_{\min}}\right)\\ &= -10\log(\lambda_{\min}).\end{aligned} \tag{14}$$

In summary, the optimal algorithm uses a composite matrix, $\boldsymbol{C}$, whose elements are calculated using the wall and window impulse response matrixes, $\boldsymbol{H}_{\text{wall}}$ and $\boldsymbol{H}_{\text{win}}$. The optimal SSNR is directly related to the minimum eigenvalue of this composite matrix and the optimal SIRF coefficients are a linearly transformed version of the unit-length eigenvector associated with the minimum eigenvalue.

While this algorithm provides the shortest possible effective channel, it requires the computation of eigenvalues and eigenvectors which can be difficult to implement in real-time off-the-shelf DSP chips. It is more desirable to have an algorithm which approaches the shortening of the optimal algorithm while providing a more computationally efficient algorithm.

C. LS Shortening

Fig. 2 shows the LS approach to impulse response shortening by modeling the impulse response of the channel with a pole-zero model. Assume that the original impulse response of the channel is represented as a transfer function $h(z^{-1}) = a(z^{-1})/1 + b(z^{-1})$. The LS approach finds the best pole-zero model with transfer function $\hat{a}(z^{-1})/1 + \hat{b}(z^{-1})$. If the poles of the model, $1 + \hat{b}(z^{-1})$, are used for the coefficients of the SIRF, then the effective channel will have a transfer function of

$$\left(\frac{a(z^{-1})}{1+b(z^{-1})}\right)(1+\hat{b}(z^{-1})) \approx a(z^{-1}) \approx \hat{a}(z^{-1}). \tag{15}$$

The SIRF cancels the poles of the physical channel, leaving the zeros of the model [1]. If the number of zeros in the model is chosen to be $\nu + 1$ then the resulting effective channel impulse response will be approximately of length $\nu+1$. Throughout the remainder of this paper it is assumed that $x(n)$ is zero-mean and white, and that the channel is stationary and causal.

Define a parameter vector as

$$\Theta = [\hat{a}_0 \quad \hat{a}_1 \quad \cdots \quad \hat{a}_\nu - \hat{b}_1 \quad -\hat{b}_2 \quad \cdots \quad -\hat{b}_t]^T \tag{16}$$

and a regressor vector as

$$\begin{aligned}\boldsymbol{\Phi}(n) = [x(n) \quad x(n-1) \quad \cdots \quad x(n-\nu) \quad y(n-1)\\ y(n-2) \quad \cdots \quad y(n-t)]^T.\end{aligned} \tag{17}$$

Using (16) and (17) we can express the estimate of $y(n)$ as

$$\hat{y}(n) = \Theta^T \boldsymbol{\Phi}(n) \tag{18}$$

and the error as

$$e(n) = y(n) - \hat{y}(n). \tag{19}$$

We desire to minimize the squared-error. From the well known LS results and assuming that $x(n)$ is sufficiently exciting we know that the parameter vector which minimizes the squared-error is

$$\Theta_{LS} = \boldsymbol{R}^{-1}\boldsymbol{r} \tag{20}$$

where

$$\boldsymbol{R} = E\{\boldsymbol{\Phi}(n)\boldsymbol{\Phi}^T(n)\} \tag{21}$$

$$\boldsymbol{r} = E\{y(n)\boldsymbol{\Phi}(n)\}. \tag{22}$$

In this paper we use the operator $E\{\ \}$ as a generalized averaging function. If $x(n)$ is viewed as a deterministic signal then $E\{x(n)\}$ is interpreted as the time average $1/n\ \Sigma_{k=0}^{n-1}\ x(k)$. If $x(n)$ is interpreted as a stationary stochastic process then $E\{x(n)\}$ is interpreted as the usual ensemble average. In the latter case the concept of least squared error is replaced by the concept of minimum mean squared error. The reader may pick which ever interpretation makes her feel more comfortable.

In typical LS applications, actual time domain samples of $x(n)$ and $y(n)$ are directly fed into the LS estimation process. DMT transceivers naturally work in the frequency domain so channel identification is usually done by generating an estimate of the channel's frequency response which is then inverse DFT'd to directly produce the estimated impulse response of the channel, $h(n)$, where $y(n) = h(n) * x(n)$. The LS method is then used to fit a $\hat{a}(z^{-1})/1 + \hat{b}(z^{-1})$ pole-zero model to the estimated impulse response. Having computed the impulse response and assuming

$$R_{xx}(k) = E\{x(n)x(n+k)\} = S_x\delta(k) \tag{23}$$

we can directly compute the components of $\boldsymbol{R}$

$$R_{xy}(k) = R_{yx}(-k) = S_x h(-k) \tag{24}$$

$$R_{yy}(k) = S_x \sum_{i=0}^{M-1} h(i)h(i-k). \tag{25}$$

In the sequel, for the joint LS and two/multichannel **AR** models, it is assumed that the computed impulse response(s) is used rather than actual time domain samples of the signals.

The discussion above places the window of $\nu + 1$ samples at the beginning of the shortened effective impulse response. As mentioned in Section II-B, this is not always the optimal choice. Potentially, modeling the impulse response from the beginning wastes some of the $\nu + 1$ zeros available on flat delay or small, insignificant samples of the impulse response. To avoid this problem, we can adjust the location of the zeros to take advantage of the flat delay of the channel.

The LS approach developed above requires the calculation and inversion of the autocorrelation matrix $\boldsymbol{R}$. The inversion makes the algorithm unsuitable for real-time system implementation. We desire a simpler approach which provides more realistic computational algorithms.

D. Two-Channel Autoregressive (AR) Modeling

A computationally practical algorithm for performing channel only shortening appears in [4] and [12]. A synopsis of those results is given here for two reasons: First, in Section IV, we will compare the results of this algorithm with the new optimal results derived in Section II-B for channel only shortening. Second, in Section III-C, we derive a new practical algorithm for joint channel and echo response shortening which utilizes many of the ideas from this algorithm.

Our pole-zero model of the channel results in a non-Toeplitz autocorrelation matrix for which there are no computationally efficient means to solve for the model parameters. If, however, the model consists only of poles, the matrix takes on a Toeplitz form for which computational efficient solutions exist. It is desirable then to reformulate the general pole-zero model into an all pole model. This method is a generalization of the approach of embedding an auto-regressive moving average (ARMA) model into a two-channel AR model first discussed in [11]. Embedding was first applied to impulse response modeling in [12] for an equal number of zeros and poles. It was generalized for a different number of zeros and poles in [4] in the context of efficiently realizing long FIR filters. Here, we utilize the approach to model the impulse response of the channel as a pole-zero model. The poles can then be used as SIRF coefficients to cancel the poles of the channel. The model used is the same used above for the LS, what differs is how the parameters are computed.

The same (t, ν) pole-zero model given in (18) can be rewritten as

$$\hat{y}_k = -\hat{b}_1 y_{k-1} - \cdots - \hat{b}_t y_{k-t} + \hat{a}_0 x_k + \cdots + \hat{a}_\nu x_{k-\nu} \tag{26}$$

where we assume here that $t \geq \nu$. The case for $t < \nu$ can easily be developed and is discussed in [4]. As is typical with AR modeling, the algorithm computes the $t + \nu + 1$ parameters recursively. In the jth recursion $(1 \leq j \leq t)$ the $(j, j - \delta)$ pole-zero model is generated from the $(j - 1, j - \delta - 1)$ pole-zero model, where $\delta = t - \nu \geq 0$. For the jth recursion the $(j; -\delta)$ pole-zero model can be written as

$$\begin{aligned} y_k = &-\hat{b}_1^j y_{k-1} - \cdots - \hat{b}_j^j y_{k-j} + \hat{a}_0^j x_k \\ &+ \cdots + \hat{a}_{j-\delta}^j x_{k-j+\delta} + e_k^j. \end{aligned} \tag{27}$$

Defining $u_k = [y_k x_{k+\delta}]^T$, (27) can be rewritten into the form of a two-channel AR model

$$u_k = -\Theta_1^j u_{k-1} - \cdots - \Theta_j^j u_{k-j} + \begin{bmatrix} e_k^j \\ x_{k+\delta} \end{bmatrix} \tag{28}$$

where

$$-\Theta_i^j = \begin{cases} \begin{bmatrix} -\hat{b}_i^j & 0 \\ 0 & 0 \end{bmatrix} & 1 \leq i \leq \delta - 1 \\ \begin{bmatrix} -\hat{b}_i^j & \hat{a}_{i-\delta}^j \\ 0 & 0 \end{bmatrix} & \delta \leq i \leq j. \end{cases} \tag{29}$$

Multiplying both sides of (28) by u_{k-i}^T and taking expectations yields

$$R(0) + \Theta_1^j R(-1) + \cdots + \Theta_j^j R(-j) = \Sigma_j^f \quad \text{for} \quad i = 0 \tag{30}$$

$$R(i)+\Theta_1^j R(i-1)+\cdots+\Theta_j^j R(i-j)=0 \quad \text{for} \quad 1 \le i \le j \tag{31}$$

where

$$R(i)=E\{u_k u_{k-i}^T\} \tag{32}$$

and

$$\Sigma_j^f = E\left\{\begin{bmatrix} e_k^j \\ x_{k+\delta}\end{bmatrix} u_k^T\right\}. \tag{33}$$

Equations (30)–(33) represent a jth order two-channel AR model. This system of linear equations can be solved efficiently using the multichannel version of the Levinson Algorithm [4], [13]. After the $j=t$th recursion, the (t,ν) pole-zero model coefficients can be read from $\Theta_i^t (1 \le i \le t)$.

The advantage of this algorithm is that computing the pole-zero model only requires the inversion of 2×2 matrixes and 2×2 matrix operations for t recursions, where the LS approach requires one recursion with the inversion of an $(t+\nu+1) \times (t+\nu+1)$ matrix. Additionally, a normalized version of the Levinson Algorithm can be used which bounds the variables in ranges more suitable for fixed-point arithmetic.

The above derivation places the zeros and poles together at the beginning of the impulse response to be modeled. It may desirable, especially when the numbers of poles and zeros are not equal, to offset one with respect to the other.

E. Observations

All of the algorithms presented above require the choice of a window location on the impulse response prior to calculating the SIRF coefficients. This can be accomplished by trying different window locations, computing the SIRF coefficients, calculating the SSNR, and choosing the window location and SIRF coefficients which yielded the best SSNR. This can be very computationally expensive and wasteful.

Often the channel has an impulse response which is very short prior to the peak (precursor) and very long after the peak (post-cursor). As such, the impulse response can often be considered as a few zeros creating the pre-cursor response, ν_{pre}, and poles and zeros, t and ν_{post}, respectively, creating the post-cursor response. Taking advantage of this fact can reduce the computational complexity of coefficient calculation. For example, one could only model the post-cursor portion of the impulse response using fewer zeros, $\nu_{\text{post}} < \nu$, and the full number of poles. When the original impulse response is convolved with the SIRF the length of the effective channel is approximately the number of unmodeled zeros in the pre-cursor, ν_{pre}, plus the number of zeros in the model, ν_{post}, where ν_{pre} and ν_{post} have been chosen such that $\nu \approx \nu_{\text{pre}} + \nu_{\text{post}}$. Additionally, since the complexity of the coefficient calculation is dependent upon the number of zeros and poles in the model, reducing the number of zeros in the model during coefficient computation in conjunction with intelligent window placement can yield solutions with lower complexity. There are other such *ad hoc* approaches which can be used to further reduce the amount of computations required to calculate the SIRF coefficients.

Another issue to be considered is model order selection. The resulting SSNR is affected if the number of poles and zeros is chosen poorly. Since simply increasing the number of poles and zeros can decrease the SSNR careful selection [4] is important. We will not discuss model order selection in this paper. The reader is referred elsewhere for information on selection of model order [14].

III. Joint Channel and Echo Shortening

So far we have only discussed the shortening of a single impulse response. When a DMT transceiver is used for bidirectional communication over a single cable there are two signal paths to consider: channel and echo. The channel is associated with communication between two remote transceivers and carries the actual data from one end to the other end. An echo path also exists between a local transmitter and receiver due to an impedance mismatch in the hybrid circuit. The receive signal is the combination of the far-end signal through the physical channel and the near-end signal through the echo path. The motivation for impulse response shortening provided in Section II-A related to the channel path. Impulse response shortening of the echo path is desirable to reduce computational complexity in the echo canceller. As the length of the impulse response of the echo path increases the computational complexity of the echo canceller increases [5], [9]. The reader is referred to [5]–[10] for discussions of echo canceller structures for DMT transceivers. The important concept for the discussion here is that it is necessary to shorten the impulse response of the channel path, and advantageous to shorten the impulse response of the echo path to reduce transceiver complexity.

A. Joint Optimal Shortening

Using similar techniques to those used in Section II-B for the optimal shortening algorithm a type of joint optimal algorithm can be developed. We can write the effective channel and echo impulse responses as e

$$\boldsymbol{h}_{\text{eff}} = \boldsymbol{H}\boldsymbol{w} \quad \text{and} \quad \boldsymbol{h}_{\text{eff},e} = \boldsymbol{H}_e \boldsymbol{w} \tag{34}$$

where $\boldsymbol{H}$ and $\boldsymbol{H}_e$ take on forms similar to that shown in (2). Let $\boldsymbol{h}_{\text{win}}$ represent a window of $\nu+1$ consecutive samples of $\boldsymbol{h}_{\text{eff}}$ starting at sample d, and let $\boldsymbol{h}_{\text{wall}}$ represent the remaining samples of $\boldsymbol{h}_{\text{eff}}$. Additionally, let $\boldsymbol{h}_{\text{win},e}$ represent a window of ν_e+1 consecutive samples of $\boldsymbol{h}_{\text{eff},e}$ starting at sample d_e, and let $\boldsymbol{h}_{\text{wall},e}$ represent the remaining samples of $\boldsymbol{h}_{\text{eff},e}$. These definitions can be expressed as

$$\boldsymbol{h}_{\text{win}} = \boldsymbol{H}_{\text{win}}\boldsymbol{w}, \quad \boldsymbol{h}_{\text{wall}} = \boldsymbol{H}_{\text{wall}}\boldsymbol{w}, \quad \boldsymbol{h}_{\text{win},e} = \boldsymbol{H}_{\text{win},e}\boldsymbol{w} \quad \text{and} \quad \boldsymbol{h}_{\text{wall},e} = \boldsymbol{H}_{\text{wall},e}\boldsymbol{w}. \tag{35}$$

One way to formulate joint optimal shortening is to choose $\boldsymbol{w}$ to minimize

$$\boldsymbol{w}^T(\alpha \boldsymbol{H}_{\text{wall}}^T \boldsymbol{H}_{\text{wall}} + (1-\alpha)\boldsymbol{H}_{\text{wall},e}^T \boldsymbol{H}_{\text{wall},e})\boldsymbol{w} = \boldsymbol{w}^T \boldsymbol{A}(\alpha)\boldsymbol{w} \tag{36}$$

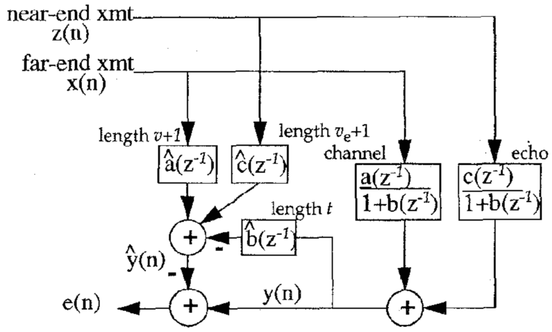

Fig. 3. Pole-Zero-Zero model for joint LS.

such that

$$w^T(\beta H_{\text{win}}^T H_{\text{win}} + (1-\beta)H_{\text{win},e}^T H_{\text{win},e})w$$
$$w^T B(\beta) w = 1. \tag{37}$$

The parameters α and β balance the energy relationship between the energy in and out of the window for both the channel and echo. If it is assumed that $B(\beta)$ is a symmetric positive definite matrix, the solution presented in Section II-B can be used to solve for the SIRF coefficients. The solution for a singular B is found in the Appendix.

Note that in (37) the weighted sum of the channel and echo windowed energy is constrained to unity. It is possible that the effective channels impulse response could have minimal or even zero energy. If this were to happen it would be necessary to increase the weight applied to the channel impulse response. The joint optimal algorithm does not appear very sensitive to α and β when ν is big enough [15]. In general, this algorithm has been found to produce the best channel and echo shortening of any of the joint shortening algorithms presented in this paper, and so serves as a benchmark in comparing the various methods.

B. Joint LS Shortening

Fig. 3 shows the model for the joint LS approach. The joint least-square approach seeks to fit a pole-zero-zero model to the echo and the channel impulse responses. Assume that the original impulse of the channel is represented as a transfer function $h(z^{-1}) = a(z^{-1})/1 + b(z^{-1})$ and the echo as $h_e(z^{-1}) = c(z^{-1})/1 + b(z^{-1})$. The joint LS approach finds the best pole-zero-zero model with transfer function $\hat{a}(z^{-1})/1 + \hat{b}(z^{-1})$ for the channel and $\hat{c}(z^{-1})/1 + \hat{b}(z^{-1})$ for the echo. If the poles of the model, $1 + \hat{b}(z^{-1})$, are used for the coefficients of the SIRF, then the effective channel will have a transfer function of $\hat{a}(z^{-1})$ and the effective echo will have a transfer function of $\hat{c}(z^{-1})$. The SIRF cancels the poles of both the physical channel and echo path, leaving the zeros of the model. While it is not obvious that a single set of poles can be used we will demonstrate in Section IV that for practical cases of interest this assumption is valid. If the number of zeros in the model for the channel is chosen to be $\nu + 1$ then the resulting effective channel impulse response will be approximately of length $\nu + 1$. The number of zeros in the model for the echo is chosen based upon the desired computational complexity of the echo canceller. It is worth noting that one can not make the echo response arbitrarily short with a limited number of poles in the model without adversely affecting the SSNR of the effective channel response.

Throughout the remainder of this paper it is assumed that $z(n)$ is zero-mean, white and independent of $x(n)$, and that the echo is stationary and causal.

Since the channel response must be shortened to remove ISI, while echo shortening is beneficial to reduce echo canceller complexity, the number of zeros used for the two responses do not have to be equal. Let $\nu_e + 1$ be the desired length of the effective echo response.

Define a new parameter vector to be

$$\Theta = [\hat{a}_0 \quad \hat{a}_1 \quad \cdots \quad \hat{a}_\nu \quad -\hat{b}_1 \quad -\hat{b}_2 \quad \cdots \quad -\hat{b}_t \quad \hat{c}_0 \quad \hat{c}_1 \quad \cdots \hat{c}_{\nu_e}]^T \tag{38}$$

and a new regressor vector

$$\begin{aligned}\Phi(n) &= [x(n)x(n-1)\cdots x(n-\nu)y(n-1)y(n-2) \\ &\quad \cdots y(n-t) \quad z(n) \quad z(n-1) \quad \cdots z(n-\nu_e)]^T \\ &= [x^T \quad y^T \quad z^T]^T.\end{aligned} \tag{39}$$

Assuming that $x(n)$ and $z(n)$ are sufficiently exciting, the parameter vector which minimizes the squared-error of the estimate is

$$\Theta_{LS} = R^{-1} r \tag{40}$$

where the autocorrelation matrix and cross correlation vector are given in (21) and (22). The elements or R and r will be derived later in the discussion of the multichannel AR model.

The same extensions that were made to the LS approach in Section II-C regarding window placement can be readily applied here. Now, however, a window can be placed independently for the channel and echo responses.

C. Multichannel AR Modeling

Since the joint LS approach requires the inversion of a large matrix, the approach is infeasible for real-time system implementation. To avoid these problems we will extend the algorithm described in Section II-D to the joint shortening case. The approach here is to embed the pole-zero-zero model into a multichannel AR model. The resulting algorithm creates a novel and computationally efficient algorithm for joint shortening.

For simplicity of derivation, assume that the number of desired zeros modeled for both the channel and echo is $\nu + 1$. It is straight forward to generalize the algorithm to the more generic case. As in Section II-D it is assumed that the length of the SIRF, t, is greater than or equal to ν. Again, the equations can be modified for the more generic case. The following derivation also places all the poles and zeros starting at the beginning of the impulse responses. The more generic case can be derived in a similar fashion to that presented below by modifying (41).

The pole-zero-zero model described in (38) and (39) can be rewritten as

$$\begin{aligned}\hat{y}_k &= -\hat{b}_1 y_{k-1} - \cdots - \hat{b}_t y_{k-t} + \hat{a}_0 x_k \\ &\quad + \cdots + \hat{a}_\nu x_{k-\nu} + \hat{c}_0 z_k + \cdots + \hat{c}_\nu z_{k-\nu}.\end{aligned} \tag{41}$$

The algorithm computes the $t+2\nu+2$ parameters recursively. In the jth recursion $(1 \leq j \leq t)$ of the algorithm the $(j, j-\delta, j-\delta)$ pole-zero-zero model is generated from the $(j-1, j-\delta-1, j-\delta-1)$ pole-zero-zero model, where $\delta = t-\nu \geq 0$. The $(j, j-\delta, j-\delta)$ pole-zero-zero model can be written as

$$y_k = -\hat{b}_1^j y_{k-1} - \cdots - \hat{b}_j^j y_{k-j} + \hat{a}_0^j x_k + \cdots + \hat{a}_{j-\delta}^j x_{k-j+\delta} + \hat{c}_0^j z_k + \cdots + \hat{c}_{j-\delta}^j z_{k-j+\delta} + e_k^j. \quad (42)$$

Defining $u_k = [y_k \; x_{k+\delta} \; z_{k+\delta}]^T$, (42) can be rewritten into the from of a multichannel AR model

$$\begin{aligned} u_k = &\begin{bmatrix} -\hat{b}_1^j & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{k-1} \\ x_{k+\delta-1} \\ z_{k+\delta-1} \end{bmatrix} + \cdots \\ &+ \begin{bmatrix} -\hat{b}_\delta^j & \hat{a}_0^j & \hat{c}_0^j \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{k-\delta} \\ x_k \\ z_k \end{bmatrix} \\ &+ \begin{bmatrix} -\hat{b}_{\delta=1}^j & \hat{a}_1^j & \hat{c}_1^j \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{k-\delta-1} \\ x_{k-1} \\ z_{k-1} \end{bmatrix} + \cdots \\ &+ \begin{bmatrix} -\hat{b}_j^j & \hat{a}_{j-\delta}^j & \hat{c}_{j-\delta}^j \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{k-j} \\ x_{k-j+\delta} \\ z_{k-j+\delta} \end{bmatrix} + \begin{bmatrix} e_k^j \\ x_{k+\delta} \\ z_{k+\delta} \end{bmatrix} \end{aligned} \quad (43)$$

$$u_k = -\Theta_1^j u_{k-1} - \cdots - \Theta_j^j u_{k-j} + \begin{bmatrix} e_k^j \\ x_{k+\delta} \\ z_{k+\delta} \end{bmatrix} \quad (44)$$

where

$$-\Theta_i^j = \begin{cases} \begin{bmatrix} -\hat{b}_i^j & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} & 1 \leq i \leq \delta - 1 \\ \begin{bmatrix} -\hat{b}_i^j & \hat{a}_{i-\delta}^j & \hat{c}_{i-\delta}^j \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} & \delta \leq i \leq j. \end{cases} \quad (45)$$

Multiplying both sides of (44) by u_{k-i}^T and taking expectations and noting that the error, e_k^j is orthogonal to $x_{k+\delta}^j$ and $z_{k+\delta}^j$ at each iteration yields

$$R(0) + \Theta_1^j R(-1) + \cdots + \Theta_j^j R(-j) = \Sigma_j^f \quad \text{for} \quad i = 0 \quad (46)$$

$$R(i) + \Theta_1^j R(i-1) + \cdots + \Theta_j^j R(i-j) = 0 \quad \text{for} \quad 1 \leq i \leq j \quad (47)$$

where

$$\begin{aligned} R(i) &= E\{u_k u_{k-i}^T\} \\ &= \begin{bmatrix} R_{yy}(i) & R_{yx}(i-\delta) & R_{yz}(i-\delta) \\ R_{xy}(i+\delta) & R_{xx}(i) & 0 \\ R_{zy}(i+\delta) & 0 & R_{zz}(i) \end{bmatrix} \end{aligned} \quad (48)$$

$$R(-i) = R^T(i) \quad (49)$$

and

$$\Sigma_j^f = \begin{bmatrix} R_{ee}(0) & 0 & 0 \\ R_{xy}(\delta) & R_{xx}(0) & 0 \\ R_{zy}(\delta) & 0 & R_{zz}(0) \end{bmatrix}. \quad (50)$$

Assuming that

$$y(n) = h(n) * x(n) + h_e(n) * z(n) \quad (51)$$

then

$$R_{xx}(k) = S_x \delta(k) \quad (52)$$

$$R_{zz}(k) = S_z \delta(k) \quad (53)$$

$$R_{xy}(k) = R_{yx}(-k) = S_x h(-k) \quad (54)$$

$$R_{zy}(k) = R_{yz}(-k) = S_z h_e(-k) \quad (55)$$

and

$$R_{yy}(k) = S_x \sum_{i=0}^{\infty} h(i)h(i-k) + S_z \sum_{i=0}^{\infty} h_e(i)h_e(i-k) \quad (56)$$

where S_x and S_z are the power in $x(n)$ and $z(n)$, respectively.

Equations (46)–(50) represent a jth order multichannel AR model. This system of linear equations can be solved efficiently using the multichannel version of the Levinson Algorithm [4], [13]. After the $j = t$ recursion, the (t, ν, ν) pole-zero-zero model coefficients can then be read from Θ_i^t $(1 \leq i \leq t)$.

As with the two-channel AR algorithm, the advantage of this algorithm over the joint LS algorithm is that computing the pole-zero-zero model only requires the inversion of 3×3 matrixes and 3×3 matrix operations for recursions, where the joint LS approach requires one recursion with the inversion of an $(t + 2\nu + 2) \times (t + 2\nu + 2)$ matrix. The same issues regarding window placement that have been discussed earlier also apply to the multichannel AR algorithm. Intelligent ad-hoc simplifications can easily be made to reduce computations while maintaining near optimum shortening performance.

IV. Simulation Results

To demonstrate the capabilities of the various algorithms presented above, the configuration shown in Fig. 4 is used. The physical cable used is 9 ft of 26 gauge unshielded twisted pair, and the bandwidth used is 20–900 kHz. The echo path is a transmit path of 20–138 kHz through a hybrid. This configuration is representative of that encountered by the async. terminal unit remote terminal (ATU-RT) in the asymmetrical digital subscriber lines (ADSL) environment [16]. The impulse responses for this configuration were provided for us by W.Y. Chen of Bellcore [17]. For all of the simulations we used a SIRF length of 16 $(t = 16)$, a CP of length 32 $(\nu = 32)$, and a target effective echo response length of 33 $(\nu_e = 32)$. Full search algorithms have also been utilized to optimize the location of the modeling windows for both the channel and echo responses. No *ad hoc* window placement algorithms have been used to ensure fair comparison of the algorithms.

Fig. 5 shows the original impulse response of the channel and the effective shortened channel if the channel only optimal algorithm presented in Section II-B is used. A SSNR = 64.1 dB was achieved. The joint optimal result is not shown since it is visually indistinguishable from the channel only optimal result. Fig. 6 shows the original echo impulse response and the effective shortened echo for both the channel only optimal and

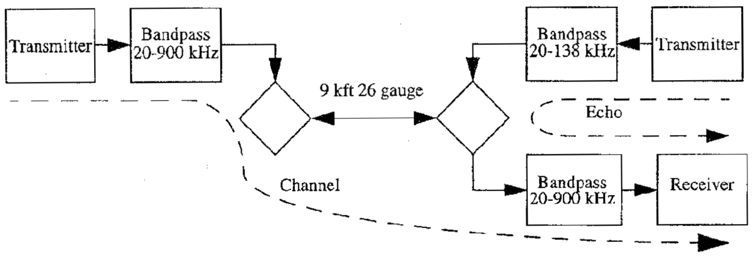

Fig. 4. Configuration for simulation examples.

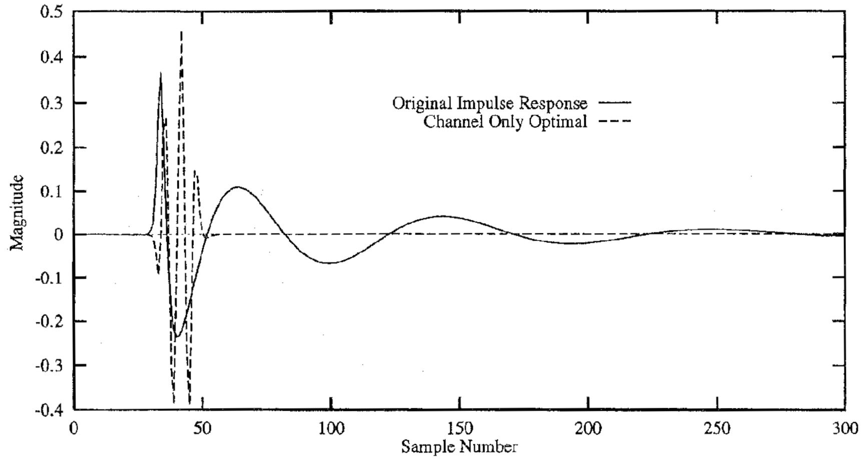

Fig. 5. Original channel response & channel only optimally shortened effective response.

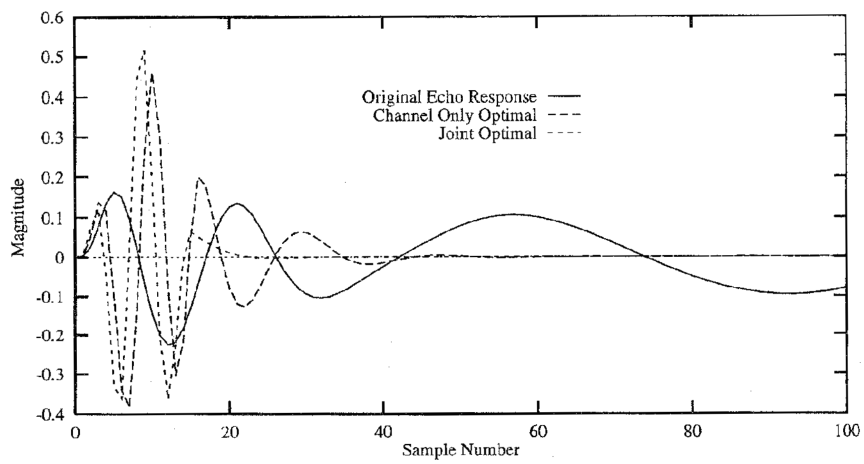

Fig. 6. Original echo impulse response and effective response resulting from channel only optimal and joint optimal shortening.

the joint optimal algorithm with $\alpha = \beta = 0.5$. As shown, both have shortened the echo response, however, the joint optimal algorithm which explicitly works to shorten the echo has done a significantly better job shortening the echo.

Table I shows the resulting channel and echo SSNR values for each of the algorithms. As shown all of the algorithms result in nearly identical channel shortening with the optimal algorithms performing the best. The joint optimal algorithm gives up very little in channel shortening and produces a large advantage in echo response shortening. As expected the algorithms which explicitly work to shorten the echo impulse response achieve the best echo SSNR. As shown in the table, the computationally efficient two-channel AR

TABLE I
SHORTENING SNR VALUES FOR THE DIFFERENT ALGORITHMS

Algorithm	Shortening Goal	SSNR	Echo SSNR
Channel Only Optimal	channel	64.1dB	23.9 dB
Least-Squares	channel	59.9 dB	22.3 dB
Two-Channel AR	channel	58.8 dB	16.2 dB
Joint Optimal	channel&echo	62.9 dB	53.2 dB
Joint Least-Squares	channel&echo	60.0 dB	45.0 dB
Multi-Channel AR	channel&echo	56.2 dB	43.1 dB

and multichannel AR algorithms perform nearly as well as the ordinary or joint LS approaches. It should be expected that the two-channel and multichannel AR algorithms perform similarly to the two LS approaches since they use the same modeling architectures. They differ for the most part in the computational approach.

When the joint shortening algorithms are utilized they take advantage of freedom remaining in the filter coefficients. It may have been that fewer coefficients could have been used if only channel shortening was the goal. It is necessary to trade-off the additional complexity in the SIRF with decreased complexity in the echo canceller. A joint shortening algorithm allows the user to utilize the SIRF to its fullest potential. In the event that a joint shortening algorithm does not provide the necessary channel shortening it is possible to increase the shortening of the channel response by increasing α and β in (36) and (37). All simulation results shown used $\alpha = \beta = 0.5$. As stated earlier, the joint optimal algorithm appears relatively insensitive to α and β when ν is large enough. A more complete study of these sensitivities is given in [15].

V. CONCLUSION

In this paper, we have developed a number of algorithms for computing the coefficients of the shortening impulse response filter (SIRF). New algorithms have been provided for shortening either the channel or the channel and echo jointly. The goal of the derivations presented was to generate computationally efficient means to calculate the SIRF coefficients in real-time using off-the-shelf DSP chips. Additionally, optimal algorithms were derived that can be used as a benchmark for future work on impulse response shortening. Simulation results were provided to demonstrate the capabilities of the various algorithms, and verify that the more computationally efficient algorithms perform adequately.

The two-channel AR algorithm provides an efficient way to shorten only the channel response, while the multichannel AR algorithm provides a realizable way to jointly shorten the echo and channel responses.

It may be possible the SSNR measure is not the best measure of equalization, since ultimately the channel SNR in each subsymbol of the multitone system is what is important. Future work may seek to develop SIRF coefficient algorithms which maximize the capacity of the multitone transceiver based upon the channel SNR. Additionally, it may be possible to apply the methods presented in this paper to communication systems that use modulation techniques other than multitone.

APPENDIX

As in Section II-B we desire to choose w to minimize $w^T A w$ given $w^T B w = 1$. Of particular interest here is the case where B is not invertible. Since B is positive semidefinite it can be written as

$$B = [U|N] \left[\begin{array}{c|c} \Sigma^2 & 0 \\ \hline 0 & 0 \end{array}\right] \left[\begin{array}{c} U^T \\ \hline N^T \end{array}\right] \tag{A-1}$$

where Σ^2 is a diagonal matrix of the positive eigenvalues of B. The columns of U are the orthonormal eigenvectors associated with Σ^2, and the columns of N are an orthonormal basis of the null space of B. The columns of N are orthogonal to the columns of U. Assume that w is of length t and that B is of rank r (i.e. $\dim(\Sigma^2) = r$). Every vector w can be written as

$$w = U\Sigma^{-1}y + Nz \tag{A-2}$$

where y is an arbitrary vector of length r and z is an arbitrary vector of length $(t-r)$. If $y^T y = 1$ then using (A-1) and (A-2)

$$\begin{aligned} w^T B w &= (U\Sigma^{-1}y + Nz)^T U\Sigma^2\, U^T(U\Sigma^{-1}y + Nz) \\ &= y^T y = 1. \end{aligned} \tag{A-3}$$

Using (A-2) the problem is now to choose y and z to minimize

$$(y^T\Sigma^{-1}\, U^T + z^T N^T)A(U\Sigma^{-1}y + Nz) \tag{A-4}$$

given $y^T y = 1$. First minimize with respect to z. Equation (A-4) can be rewritten as

$$y^T\Sigma^{-1}\, U^T AU\Sigma^{-1}y + 2y^T\Sigma^{-1}\, U^T ANz + z^T N^T ANz. \tag{A-5}$$

Setting the derivative of (A-5) with respect to z equal to zero, we get

$$z = -(N^T AN)^{-1} N^T AU\Sigma^{-1}y. \tag{A-6}$$

Substituting (A-6) into (A-4), the quantity that is left to be minimized is

$$\begin{aligned} &(y^T\Sigma^{-1}\, U^T - y^T\Sigma^{-1}\, U^T AN((N^T AN)^{-1})^T)A \\ &\quad (U\Sigma^{-1}y - N(N^T AN)^{-1}N^T AU\Sigma^{-1}y) = y^T C y \end{aligned} \tag{A-7}$$

where C is appropriately define. Using (A-2) and (A-6) the resulting optimal SIRF coefficients are

$$w_{\text{opt}} = (I - N(N^T AN)^{-1}N^T A)U\Sigma^{-1}l_{\min} \tag{A-8}$$

where $l_{\min}$ is the unit-length eigenvector associated with the minimum eigenvalue of C given in (A-7).

REFERENCES

[1] J. S. Chow, J. C Tu, and J. M. Cioffi, "A discrete multitone transceiver system for HDSL applications," *IEEE J. Select. Areas Commun.*, vol. 9, no. 6, Aug. 1991.
[2] J. S. Chow and J. M Cioffi, "A cost-effective maximum likelihood receiver for multicarrier systems," *Proc. ICC*, 1992.
[3] J. S. Chow, J. M Cioffi, and J. A. C. Bingham, "Equalizer training algorithms for multicarrier modulation systems," in *Int. Conf. Commun.*, Geneva, Switzerland, May 1993, pp. 761–765.
[4] N. Al-Dhahir, A. Sayed, and J. M. Cioffi, "A high-performance cost-effective pole-zero MMSE-DFE," in *Proc. Allerton Conf. Commun., Control Computing*, Sept. 1993.
[5] J. Cioffi and J. A. C. Bingham, "A data-driven multitone echo canceller," *Globecom 1991*, Jan. 1991.
[6] R. C. Younce, P. J. W. Melsa, and S. Kapoor, "Echo cancellation for asymmetrical digital subscriber lines," *Proc. ICC 1994*, May 1994.
[7] J. Yang, S. Roy, and N. Lewis, "Data driven echo cancellation for a multi-tone modulation system," *IEEE Trans. Commun.*, vol. 42, May 1994.
[8] J. M. Cioffi and J. A. C. Bingham, "A data-driven multitone echo canceller," *IEEE Trans. Commun.*, vol. 42, No. 10, Oct. 1994.
[9] M. Ho, J. M. Cioffi, and J. A. C. Bingham, "Discrete multitone echo cancellation," Submitted for publication.
[10] M. Ho, "Multicarrier echo cancellation and multichannel equalization," Ph.D. Thesis, Stanford University, June 1995.

[11] D. T. Lee, B. Friedlander, and M. Morf, "Recursive ladder algorithms for ARMA modeling," *IEEE Trans. Automat. Contr.*, vol. AC-27, no. 4, Aug. 1982.
[12] J. Tu, "Theory, design, and application of multi-channel modulation for digital communications," Ph.D. dissertation, Stanford University, June 1991.
[13] S. Kay, *Modern Spectral Estimation: Theory and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[14] H. Akaike, "A new look at statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, no. 6, Dec. 1974.
[15] P. Melsa, C. Rohrs, and R. Younce, "Joint optimal impulse response shortening," in *Proc. ICASSP '96*, May 1996.
[16] "Draft asymmetric digital subscriber line (ADSL) metallic interface specification," T1E1.4/95-007, Committee T1-Telecommunications, Jan. 1995.
[17] W. Y. Chen, Bellcore, private communication, July 1993.

Peter J. W. Melsa (S'85–M'89) was born in Tucson, AZ, on April 17, 1965. He received the B.S. degree in electrical engineering (with highest honors) from the University of Notre Dame, South Bend, IN, in 1987 and the M.S. degree in electrical engineering from North Carolina State University, Raleigh, in 1988.

Since January 1989, he has been working for Tellabs in their Research Center located in Indiana. His research interests include signal processing and communications.

Mr. Melsa was the recipient of a National Science Foundation Graduate Studies Fellowship in 1987.

Richard C. Younce (S'82–M'91) graduated with highest honors from the University of Notre Dame, South Bend, IN, in 1984, received the M.S. degree from the University of Illinois, Urbana-Champaign, in 1985, and the Ph.D. degree in 1990 from the University of Notre Dame, all in electrical engineering.

Since 1985, he has been at the Tellabs Research Center and beginning in 1990 has held the position of Adjunct Associate Professor of Electrical Engineering at the University of Notre Dame, South Bend, IN. His research interests include adaptive systems, signal processing, and their application to echo cancellation and other communications areas.

Charles E. Rohrs (M'82–SM'88) was born in Rockville Centre, NY, on March 16, 1954. He graduated summa cum laude in 1976 from the University of Notre Dame, South Bend, IN, received the M.S. degree in 1978, and the Ph.D. degree in 1982 from MIT, Cambridge, MA, all in electrical engineering.

From 1982 to 1985, he was an Assistant Professor of Electrical Engineering at the University of Notre Dame. In 1985, he joined Tellabs, Inc., while retaining an Adjunct Position at Notre Dame. He spent ten years as Director of Tellabs Research Center in Mishawaka, IN. In 1995, he became a Tellabs Fellow. He is currently an Adjunct Associate Professor at the University of Notre Dame where he has developed a course in Business Principles for Engineers. He also spent one year as a Visiting Scientist at MIT where he now holds an appointment as a Research Affiliate. He is the principle author of a textbook, *Linear Control System* and has published papers in the areas of control theory, communication theory, communication network theory, and digital signal processing. His current research areas of interest include broadband communication network and digital modulation and coding techniques. He has served as an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL.

Dr. Rohrs is a member of Tau Beta Pi and Eta Kappa Nu.