Blind, Adaptive Channel Shortening by Sum-Squared Auto-Correlation Minimization (SAM)

Jaiganesh Balakrishnan, *Member, IEEE*, Richard K. Martin, *Student Member, IEEE*, and C. Richard Johnson, Jr., *Fellow, IEEE*

***Abstract*—We propose a new blind, adaptive channel shortening algorithm for updating the coefficients of a time-domain equalizer in a system employing multicarrier modulation. The technique attempts to minimize the sum-squared auto-correlation terms of the effective channel impulse response outside a window of desired length. The proposed algorithm, known as "sum-squared auto-correlation minimization" (SAM), requires the source sequence to be zero-mean, white, and wide-sense stationary, and it is implemented as a stochastic gradient descent algorithm. Simulation results have been provided, demonstrating the success of the SAM algorithm in an AUTHOR: What does ADSL stand for? (ADSL) system.**

***Index Terms*—Adaptive, blind, channel shortening, DMT, equalization, multicarrier, OFDM.**

I. Introduction

Multicarrier modulation (MCM) techniques like orthogonal frequency division multiplexing (OFDM) and discrete multitone (DMT) have been gaining in popularity in recent years. One reason for this surge in popularity is the ease with which MCM can combat channel dispersion, provided the channel delay spread is not greater than the length of the cyclic prefix (CP). However, if the CP is not long enough, the orthogonality of the sub-carriers is lost, and this causes both inter-carrier interference (ICI) and inter-symbol interference (ISI). The inadequacy of the CP in digital subscriber loop (xDSL) systems can be seen by considering the standard carrier serving area (CSA) test loops [1].

A well-known technique to combat the ICI/ISI caused by the inadequate CP length is the use of a time-domain equalizer (TEQ) at the receiver front-end. The TEQ is a filter that shortens the channel so that the delay spread of the effective channel impulse response is no larger than the length of the CP. The TEQ design problem has been extensively studied in the literature. In [2], Falconer and Magee proposed a minimum mean squared error (MMSE) channel shortening method, which was designed for maximum likelihood sequence estimation. More recently, Melsa *et al.* [3] proposed the maximum shortening signal-to-noise ratio (MSSNR) method, which attempts to minimize the energy outside the window of interest while holding the energy inside fixed. However, in a point-to-point system such as DSL, the true performance metric to optimize is the maximum bit allocation that does not cause the error probability to exceed a threshold, and in broadcast systems, the true performance metric is the bit error rate (BER) for a fixed bit allocation. Optimizing the MSE or SSNR does not necessarily optimize the bit rate [4] or error probability [5]. Recent work [4], [6], [7] has addressed the problem of maximizing the bit rate in xDSL systems.

Besides the TEQ structure, there are several alternative structures that can be used for equalization in multicarrier systems. Per tone equalization moves the TEQ to the far side of the fast Fourier transform (FFT) in the receiver, allowing separate equalization for each tone [8], [9]. Alternatively, a decision feedback multi-input multi-output (MIMO) equalizer can be used to cancel the interference by estimating the transmitted symbols, filtering them, and subtracting the result from the received signal [10]. Yet another approach was devised by Trautmann and Fliege [11], in which a post-FFT block-equalizer structure is used. This is similar to the per tone structure, but it is shown in [11] that the unused tones can be exploited to remove the ICI from the remaining tones with high performance. However, each of these approaches requires matrix processing of the received signal rather than simple filtering. This paper focuses on the TEQ structure for equalization since it has a low-complexity implementation, and via the proposed approach, it can easily be made to adapt blindly.

All of the TEQ design techniques described above are nonadaptive (except [2]), and all require training (usually to estimate the channel). The MMSE solution [2] can be implemented adaptively (using training), but it is cited as converging very slowly [12]. Chow's algorithm [13] converges more quickly, but it usually converges to a distinctly suboptimal setting [12]. Lashkarian and Kiaei [14] have developed an iterative implementation of Al-Dhahir's approximate maximum bit rate method [6], but as cited in [4], the method in [6] makes several inaccurate assumptions and is not really optimal. Furthermore, the method in [14] is not truly adaptive in the sense that it assumes knowledge of large matrices that depend upon the channel; hence, it is not able to track a time-varying channel. In [15] adaptive channel shortening is discussed, but the focus is on the performance metric, and no adaptive algorithm is explicitly given.

In the context of multicarrier modulation, we use the term "blind" to refer to an algorithm that does not require knowledge

Manuscript received July 22, 2003; revised April 14, 2003. This work was supported in part by the Natioanl Science Foundation under Grant ECS-9811297 and NxtWave Communications. The associate editor coordinating the review of this paper and approving it for publication was Dr. Sergios Theodoridis.

J. Balakrishnan was with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA. He is now with Texas Instruments, Dallas, TX 75243 USA (e-mail: jai@ti.com).

R. K. Martin is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: frodo@ece.cornell.edu).

C. R. Johnson, Jr. is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: johnson@ece.cornell.edu).

Digital Object Identifier 10.1109/TSP.2003.818892

of the exact values of the transmitted signal, although other definitions could be chosen. An algorithm that exploits the transmission of known signals (including zeros) would not be considered blind by this definition, but an algorithm that exploits structural properties of the signal (such as a constant modulus signal on each tone, or the cyclo-stationarity introduced by the cyclic prefix) would be considered blind by this definition. The goal of this paper is to develop a blind, adaptive channel shortening algorithm. The problem of adapting the 1-tap frequency-domain equalizer (FEQ) [13] per tone is not considered, but this is usually done through the use of frequency-domain training or decision-directed least mean squares.

De Courville *et al.* have proposed a blind, adaptive equalizer for a multicarrier receiver [16], but it performs equalization to a single spike rather than channel shortening. The algorithm assumes that there is oversampling in the transmitter, which has the effect of zero-padding the IFFT input. The equalizer is adapted in order to restore the zeros on the corresponding FFT outputs. The transmission of zeros on certain carriers could be thought of as training signals consisting of zeros; therefore, the use of the term "blind" for this algorithm is debatable. However, [16] is the first algorithm in the literature that performs adaptive equalization (to a single spike) for a multicarrier receiver.

Martin *et al.* [17] have proposed a low-complexity, blind, adaptive TEQ algorithm known as MERRY, but it only updates once per symbol. The MERRY algorithm is based on restoring the redundancy introduced by the CP, and the cost function is the mean squared error between the data in the CP and the corresponding data in the signal. In contrast, the sum-squared auto-correlation minimization (SAM) algorithm proposed in this paper adapts in order to suppress the received signal's autocorrelation outside of a CP-length window. SAM converges much faster than the MERRY algorithm but at the expense of significantly higher complexity. SAM has the added advantage of not requiring an estimate of the symbol synchronization (i.e., the location of the start of each data block). MERRY requires that the channel not vary significantly over each symbol (since it only updates once per symbol), but the SAM algorithm can track time variations within a symbol (since it can update once per sample).

The adjectives "blind" and "adaptive" need some motivation. In the DSL case, the TEQ is expected to converge completely by the end of the initialization period, which consists entirely of training symbols. Thus, one can argue that in that situation, a blind algorithm is unnecessary. However, if there are any further variations in the channel, for example, due to temperature variations, then a blind algorithm can track those variations. In a wireless environment, one wishes to adapt continually, even between training frames, since the channel is constantly changing. Finally, even when training is available, a blind algorithm does not require knowledge of where the training symbol lies in the data. In particular, SAM does not even need to know where the symbol boundaries are.

Beyond being necessary in a time-varying environment, adaptive realizations can also lead to reduced complexity algorithms. Nonadaptive algorithms such as the minimum MSE solution and maximum SNR solution require matrix inversions and eigen decompositions, which are very costly, whereas adaptive algorithms (such as SAM) are usually vector update rules and can be thought of as just iterative approximation algorithms.

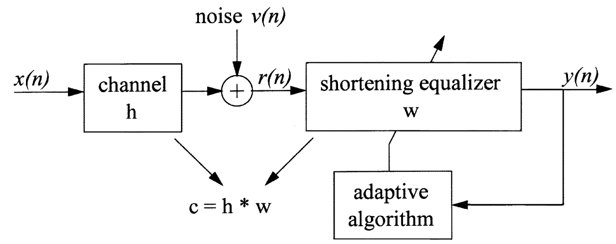


Fig. 1. System model for an adaptive TEQ.

The remainder of this paper is organized as follows. Section II presents the system model and notation. Sections III and IV discuss the SAM cost function and gradient descent algorithm. Section V studies the properties of the cost function. Section VI provides simulations of SAM in an ADSL environment, and Section VII concludes.

II. System Model

The system model is shown in Fig. 1. Let $x(n)$ be the source sequence to be transmitted through a linear finite-impulse-response (FIR) channel $\mathbf{h}$ of length (L_h+1) taps. Let $r(n)$ be the received signal, which will be filtered through an (L_w+1)-tap TEQ with an impulse response vector $\mathbf{w}$ to obtain the output sequence $y(n)$. Let $\mathbf{c}=\mathbf{h}\star\mathbf{w}$ denote the effective channel-equalizer impulse response vector of length (L_c+1) taps, where $L_c=L_h+L_w$. The TEQ will be adapted with the goal of shortening the effective channel $\mathbf{c}$ such that it possesses significant coefficients only within a contiguous window of size $(\nu+1)$ taps. In multicarrier systems, ν is the CP length. That is, we wish to minimize the energy of the coefficients in the effective channel outside the window of interest. The received sequence $r(n)$ is

$$r(n)=\sum_{k=0}^{L_h}h(k)x(n-k)+v(n) \tag{1}$$

and the output of the TEQ is

$$y(n)=\sum_{k=0}^{L_w}w(k)r(n-k)=\mathbf{w}^T\mathbf{r}_n \tag{2}$$

where $\mathbf{r}_n=[r(n)\ r(n-1)\ \cdots\ r(n-L_w)]^T$. Throughout, we make the following assumptions.

1) The source sequence $x(n)$ is white, zero-mean, and wide-sense stationary (W.S.S.).
2) The relation $2L_c<N_{\text{fft}}$ holds for multicarrier (or block-based[1]) systems, i.e., the combined channel has length less than half the FFT (or block) size.
3) The source sequence $x(n)$ is real and has a unit variance.

[1]Vaidyanathan and Vrcelj [18] have proposed the use of a block structure and a cyclic prefix for single-carrier systems, in which case channel shortening may be needed.

4) The noise sequence $v(n)$ is zero-mean, i.i.d., uncorrelated to the source sequence and has a variance σ_v^2.

The first assumption is critical for the proposed channel shortening algorithm. Assumption two is important for analytical reasons, but if it is modestly violated, the performance degradation should be minor. This assumption is irrelevant for the application of SAM to equalization of (non-CP-based) single carrier systems. The last two assumptions are for notational simplicity.

III. Sum-Squared Auto-Correlation Minimization

This section motivates the use of the SAM cost function and shows how to blindly measure it from the data. Consider the auto-correlation sequence of the combined channel-equalizer impulse response, i.e.,

$$R_{cc}(l) = \sum_{k=0}^{L_c} c(k)c(k-l). \tag{3}$$

For the effective response $\mathbf{c}$ to have zero taps outside a window of size $(\nu + 1)$, it is necessary for the auto-correlation values $R_{cc}(l)$ to be zero outside a window of length $(2\nu + 1)$, i.e.,

$$R_{cc}(l) = 0, \quad \forall |l| > \nu. \tag{4}$$

Hence, one possible way of performing channel shortening is by ensuring that (4) is satisfied by the auto-correlation function of the combined response. However, this has a trivial solution when $\mathbf{c} = \mathbf{0}$ or, equivalently, $\mathbf{w} = \mathbf{0}$. This trivial solution can be avoided by imposing a norm constraint on the effective response, for instance $\|\mathbf{c}\|_2^2 = 1$ or, equivalently, $R_{cc}(0) = 1$.

It should be noted that perfect nulling of the auto-correlation values outside the window of interest is not possible since perfect channel shortening is not possible when a finite length baud-spaced TEQ is used. This is because if the channel has L_h zeros, then the effective response will always have $L_h + L_w$ zeros. If we had decreased the length of the channel to, say, $L_s < L_h$ taps, then the combined response would only have L_s zeros, which contradicts our previous statement.

Hence, we define a cost function $J_{\nu+1}$ in an attempt to minimize (instead of nulling) the sum-squared auto-correlation terms, i.e.,

$$J_{\nu+1} = \sum_{l=\nu+1}^{L_c} |R_{cc}(l)|^2. \tag{5}$$

The TEQ optimization problem can then be stated as

$$\mathbf{w}^{\text{opt}} = \arg_{\mathbf{w}} \min_{\|\mathbf{c}\|_2^2=1} J_{\nu+1}. \tag{6}$$

Consider the auto-correlation function of the sequence $y(n)$

$$\begin{aligned} R_{yy}(l) &= E[y(n)y(n-l)] \\ &= E\left[\left(\mathbf{c}^T\mathbf{x}_n + \mathbf{w}^T\mathbf{v}_n\right)\left(\mathbf{x}_{n-l}^T\mathbf{c} + \mathbf{v}_{n-l}^T\mathbf{w}\right)\right] \end{aligned} \tag{7}$$

where $\mathbf{x}_n = [x(n)\ x(n-1)\ \cdots\ x(n-L_h-L_w)]^T$, and $\mathbf{v}_n = [v(n)\ v(n-1)\ \cdots\ v(n-L_w)]^T$. To simplify

$$E\left[\mathbf{v}_n\mathbf{v}_{n-l}^T\right] = \begin{bmatrix} R_{vv}(l) & \cdots & R_{vv}(l+L_w) \\ \vdots & \ddots & \vdots \\ R_{vv}(l-L_w) & \cdots & R_{vv}(l) \end{bmatrix} \tag{8}$$

where $R_{vv}(l) = E[v(n)v(n-l)]$. Since $v(n)$ is i.i.d., this matrix will be Toeplitz, with only one diagonal of nonzero entries. It becomes a shifting matrix, i.e., its affect on a vector is to shift the elements of the vector up or down (depending on l). Since the signal and noise are uncorrelated, $E[\mathbf{x}_n\mathbf{v}_{n-l}^T] = \mathbf{0}$, and $E[\mathbf{v}_n\mathbf{x}_{n-l}^T] = \mathbf{0}$. Finally, $E[\mathbf{x}_n\mathbf{x}_{n-l}^T]$ becomes another shifting matrix, provided that the assumption $2(L_h + L_w) < N_{\text{fft}}$ holds. If this is violated, then the matrix is still Toeplitz, but for some values of l, there will be another diagonal of nonzero entries, corresponding to the correlation between samples in the transmitted symbol end and samples in the transmitted cyclic prefix. Fortunately, assumption 2 is a reasonable one, as can be seen by considering the CSA test loop channels [1] for the case of DSL: $L_h \cong 200, L_w \cong 32$, and $N_{\text{fft}} = 512$, so $2(200 + 32) < 512$.

Now, (7) can be simplified to

$$\begin{aligned} R_{yy}(l) &= \sum_{k=0}^{L_c} c(k)c(k-l) + \sigma_v^2 \sum_{k=0}^{L_w} w(k)w(k-l) \\ &= R_{cc}(l) + \sigma_v^2 R_{ww}(l). \end{aligned} \tag{9}$$

Under the noiseless scenario, $R_{yy}(l) = R_{cc}(l)$, and hence, (5) can be rewritten as

$$J_{\nu+1} = \sum_{l=\nu+1}^{L_c} |R_{cc}(l)|^2 = \sum_{l=\nu+1}^{L_c} |R_{yy}(l)|^2. \tag{10}$$

In the presence of noise, (10) is only approximately true. This suggests approximating the cost function of (5) by

$$\begin{aligned} \hat{J}_{\nu+1} &= \sum_{l=\nu+1}^{L_c} |R_{yy}(l)|^2 \\ &= \sum_{l=\nu+1}^{L_c} |R_{cc}(l)|^2 + 2\sigma_v^2 \sum_{l=\nu+1}^{L_w} R_{cc}(l)R_{ww}(l) \\ &\quad + \sigma_v^4 \sum_{l=\nu+1}^{L_w} |R_{ww}(l)|^2. \end{aligned} \tag{11}$$

In many cases, the equalizer length $L_w + 1$ is comparable to or shorter than the cyclic prefix length ν. (This is true, for example, in [3] and [4].) In such situations, both noise terms in (11) vanish entirely, due to the empty summations. Even if L_w is significantly longer than ν, for typical SNR values, σ_v^4 will be very small (compared with the unit variance source signal); therefore, we can neglect the last term in (11). Furthermore, the summands in the second term will be both positive and negative so they will often add to a small value. Combining this with the fact that the second summation is multiplied by the (small) noise variance, we are justified in ignoring the second term in (11) as well. This leaves us with $\hat{J}_{\nu+1} \cong J_{\nu+1}$ (and $\hat{J}_{\nu+1} = J_{\nu+1}$ exactly if $L_w < \nu + 1$). Accordingly, we will henceforth drop the hat on $J_{\nu+1}$ and ignore the noise terms. The effect of noise on the performance of SAM is investigated in Section VI.

Note that the cost function $J_{\nu+1}$ depends only on the output of the TEQ, namely, $y(n)$, and the choice of ν. Hence, a gradient-descent algorithm over this cost function, with an additional norm constraint on $\mathbf{c}$ or $\mathbf{w}$, requires no knowledge of the source sequence. Such an algorithm will be derived in Section IV. In addition, note that the channel length L_h+1 must be known in order to determine L_c. In ADSL systems, the channel is typically modeled as a length N FIR filter, where $N=512$ is the FFT size. The CSA test loops [1] typically have almost all of their energy in 200 consecutive taps; therefore, the FFT size is a very conservative choice for L_h+1 in this application. For other applications, the user must choose a reasonable estimate (or overestimate) for L_h based on typical delay spread measurements for that application.

IV. Adaptive Algorithm

The steepest gradient-descent algorithm over the cost surface $J_{\nu+1}$ is

$$\mathbf{w}^{\text{new}} = \mathbf{w}^{\text{old}} - \mu \nabla_{\mathbf{w}} \left(\sum_{l=\nu+1}^{L_c} E[y(n)y(n-l)]^2 \right) \tag{12}$$

where μ denotes the step size, and $\nabla_{\mathbf{w}}$ denotes the gradient with respect to $\mathbf{w}$. To implement this algorithm, an instantaneous cost function is defined, where the expectation operation is replaced by a moving average over a user-defined window of length N.

$$J_{\nu+1}^{\text{inst}}(k) = \sum_{l=\nu+1}^{L_c} \left\{ \sum_{n=kN}^{(k+1)N-1} \frac{y(n)y(n-l)}{N} \right\}^2. \tag{13}$$

The value of N is a design parameter. It should be large enough to give a reliable estimate of the expectation, but no larger, as the algorithm complexity is proportional to N. The "stochastic" gradient-descent algorithm is then given by

$$\mathbf{w}^{k+1} = \mathbf{w}^k - 2\mu \sum_{l=\nu+1}^{L_c} \left[\left\{ \sum_{n=kN}^{(k+1)N-1} \frac{y(n)y(n-l)}{N} \right\} \times \left\{ \nabla_{\mathbf{w}} \left(\sum_{n=kN}^{(k+1)N-1} \frac{y(n)y(n-l)}{N} \right) \right\} \right]$$

which simplifies to

$$\mathbf{w}^{k+1} = \mathbf{w}^k - 2\mu \sum_{l=\nu+1}^{L_c} \left[\left\{ \sum_{n=kN}^{(k+1)N-1} \frac{y(n)y(n-l)}{N} \right\} \times \left\{ \sum_{n=kN}^{(k+1)N-1} \left(\frac{y(n)\mathbf{r}_{n-l} + y(n-l)\mathbf{r}_n}{N} \right) \right\} \right]. \tag{14}$$

The TEQ update algorithm described in (14) will be referred to as the SAM algorithm, as it attempts to minimize the cost function described in (5).

An alternate method of implementing the algorithm comes from using auto-regressive (AR) estimates instead of moving average (MA) estimates. Let

$$\begin{aligned}
\mathbf{A}^n &= (1-\alpha)\mathbf{A}^{n-1} + \alpha y(n) \begin{bmatrix} r(n-\nu-1) \\ \vdots \\ r(n-L_c-L_w) \end{bmatrix} \\
\mathbf{B}^n &= \mathbf{W}\mathbf{A}^n \\
\mathbf{C}^n &= (1-\alpha)\mathbf{C}^{n-1} \\
&\quad + \alpha \begin{bmatrix} r(n) \\ \vdots \\ r(n-L_w) \end{bmatrix} \begin{bmatrix} y(n-\nu-1) \\ \vdots \\ y(n-L_c) \end{bmatrix}^T
\end{aligned}$$

where $0 < \alpha < 1$ is a design parameter, and $\mathbf{W}$ is the $(L_c-\nu) \times (L_c+L_w-\nu)$ convolution matrix of the equalizer

$$\mathbf{W} = \begin{bmatrix} w_0 & w_1 & w_2 & \cdots & 0 & 0 \\ 0 & w_0 & w_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & w_{L_w-1} & w_{L_w} \end{bmatrix}. \tag{15}$$

Using these AR estimates, the update rule can be written as

$$\begin{aligned}
\mathbf{w}^{n+1} &= \mathbf{w}^n - 2\mu \sum_{l=\nu+1}^{L_c} \{E[y(n)y(n-l)]\} \\
&\qquad \cdot \{E[y(n)\mathbf{r}_{n-l} + y(n-l)\mathbf{r}_n]\} \\
&\cong \mathbf{w}^n - 2\mu \sum_{l=\nu+1}^{L_c} \{\mathbf{B}_{l-\nu}\} \\
&\qquad \cdot \left\{ \begin{bmatrix} \mathbf{A}_{l-\nu} \\ \vdots \\ \mathbf{A}_{l-\nu+L_w} \end{bmatrix} + \begin{bmatrix} \mathbf{C}_{1,l-\nu} \\ \vdots \\ \mathbf{C}_{L_w+1,l-\nu} \end{bmatrix} \right\}.
\end{aligned} \tag{16}$$

With both implementations, $\mathbf{w}$ must be periodically renormalized (or else the constraint may be implemented in some other fashion, such as by adding a penalty term onto the cost function). The advantage of this implementation is that it allows us to form an update at each time instant, rather than every Nth time instant, where N is the number of samples used in the block averaging of the expectation estimates. The disadvantage is that the estimates now depend more on previous settings of $\mathbf{w}$ rather than the current setting, but if the time variations are reasonably slow, this should not matter. In terms of complexity, the auto-regressive implementation of (16) requires approximately $4L_w(L_c-\nu)$ multiplications and additions (each) per update, plus a division for renormalization, whereas the moving average implementation of (14) requires approximately $3NL_w(L_c-\nu)$ multiplications and additions (each) per update, plus a division for renormalization. Hence, the complexity per unit time is approximately the same for the two if (14) is implemented only once every N samples. However, the moving average implementation is intuitively appealing and is useful for analytic purposes.

The choice of α in the AR implementation is analogous to the choice of N in the MA implementation of (13). Both the MA and the AR estimates are unbiased:

$$\begin{aligned} E\left[\hat{R}_{yy}^{\mathrm{MA}}(l)\right] &= \frac{1}{N}\sum_{n} E[y(n)y(n-l)] \\ &= \frac{1}{N}\cdot N\cdot R_{yy}(l) = R_{yy}(l) \\ E\left[\hat{R}_{yy}^{\mathrm{AR}}(l)\right] &= \sum_{k=0}^{\infty}\alpha(1-\alpha)^k E[y(n-k)y(n-k-l)] \\ &= \alpha\cdot\frac{1}{1-(1-\alpha)}\cdot R_{yy}(l) = R_{yy}(l). \end{aligned} \tag{17}$$

A "fair" comparison of the two approaches should set N and α such that the variances of the two estimates are equal, yet closed-form expressions for the variances of the two estimates are difficult to obtain. An examination of (17) suggests that $\alpha = 1/N$ is a reasonable choice.

As stated earlier, to prevent the algorithm from collapsing the TEQ to an all-zero solution, the equalizer parameters can be normalized after each update to ensure that the norm of the effective response is unity, i.e., $||\mathbf{c}||_2^2 = 1$. As the source sequence is assumed to be white, from (9)

$$E[y^2(n)] = ||\mathbf{c}||_2^2 + \sigma_v^2||\mathbf{w}||_2^2 \approx ||\mathbf{c}||_2^2 \tag{18}$$

and the norm of $\mathbf{c}$ can be approximately determined by monitoring the energy of the output sequence $y(n)$. The approximation does not matter much as it is only used to keep $||\mathbf{c}||_2^2$ nonzero, and the actual value of $||\mathbf{c}||_2^2$ does not matter. Similarly, if the source is nonwhite, (18) does not hold exactly, but maintaining $E[y^2(n)] = 1$ will still keep $||\mathbf{c}||_2^2 \neq 0$. A more easily implementable constraint is the unit norm constraint on $\mathbf{w}$, i.e., $||\mathbf{w}||_2^2 = 1$. This is easier to implement because we have ready access to $\mathbf{w}$ but not to $\mathbf{c}$; therefore, this is the constraint used in the simulations in Section VI.

V. Properties of the Cost Function

As is typical of blind equalization algorithms, for instance the constant modulus algorithm (CMA) [19], SAM's cost surface can be expected to be multimodal. If it has bad local minima, then initialization to ensure convergence to the global minimum becomes important. In general, the SAM cost surface will have local minima. This is a direct result of the following theorem.

Theorem 1: The SAM cost function is invariant to the operation $\mathbf{w} \rightarrow \bar{\mathbf{w}}$, where $\bar{\mathbf{w}}$ denotes $\mathbf{w}$ with the order of its elements reversed.

Proof: Consider the autocorrelation sequences of the combined channels $\mathbf{c}_1 = \mathbf{h} \star \mathbf{w}$ and $\mathbf{c}_2 = \mathbf{h} \star \bar{\mathbf{w}}$.

$$\begin{aligned} R_{c_1c_1} &= \mathbf{c}_1 \star \bar{\mathbf{c}}_1 = (\mathbf{h}\star\mathbf{w}) \star \overline{(\mathbf{h}\star\mathbf{w})} \\ &= \mathbf{h}\star\mathbf{w}\star\bar{\mathbf{h}}\star\bar{\mathbf{w}} \\ &= (\mathbf{h}\star\bar{\mathbf{w}})\star(\bar{\mathbf{h}}\star\mathbf{w}) \\ &= \mathbf{c}_2\star\bar{\mathbf{c}}_2 = R_{c_2c_2}. \end{aligned} \tag{19}$$

Since the autocorrelation sequence is invariant to reversing the order of the elements of $\mathbf{w}$, the SAM cost is also invariant to such a switch. ■

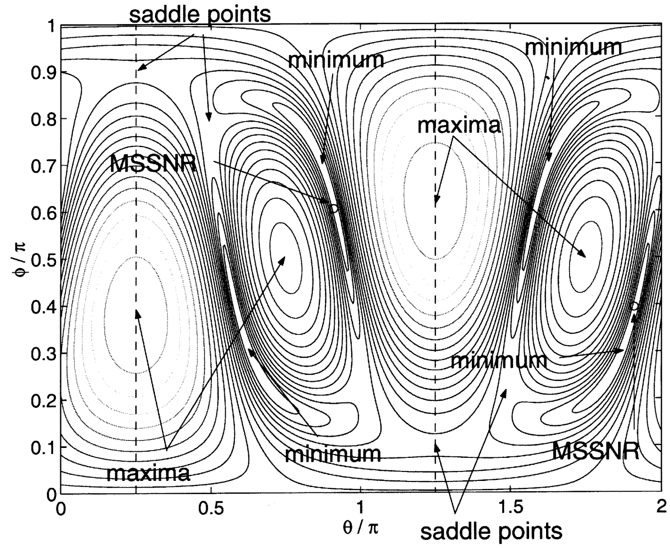


Fig. 2. Contours of the SAM cost function. The two circles are the global minima of the 1/SSNR cost function. The cost function is symmetric about the dashed line.

The upshot of Theorem 1 is that whenever there is a good minimum of the SAM cost surface, say at $\mathbf{w}_o$, there will also be another minimum at $\bar{\mathbf{w}}_o$. There is no reason to expect that the flipped $\mathbf{w}_o$ is as good an equalizer as $\mathbf{w}_o$ (in terms of achievable bit rate, for example); therefore, each good minimum may give rise to a bad minimum. Here, "good" and "bad" mean that even though the SAM cost is the same, the ultimate performance metric (achievable bit rate, for ADSL) will not be the same for the two settings. Another consequence is that the SAM cost surface is symmetric with respect to $\mathbf{w} \Leftrightarrow \bar{\mathbf{w}}$; therefore, there will be minima, maxima, or saddle points along the subspace $\mathbf{w} = \bar{\mathbf{w}}$.

To visualize Theorem 1, consider the following example. The channel is $\mathbf{h} = [1, 0.3, 0.2]$, the cyclic prefix length is 1 (so we want a 2-tap channel), there is no noise, the equalizer $\mathbf{w}$ has three taps, and we use the unit norm constraint $||\mathbf{w}|| = 1$. With this constraint, the equalizer must lie on a unit sphere; therefore, we can represent the equalizer in spherical coordinates: $w_0 \stackrel{\triangle}{=} w_x = \cos(\theta)\sin(\phi)$, $w_1 \stackrel{\triangle}{=} w_z = \cos(\phi)$, $w_2 \stackrel{\triangle}{=} w_y = \sin(\theta)\sin(\phi)$. In this case, $\mathbf{w} \rightarrow \bar{\mathbf{w}}$ is equivalent to switching w_x and w_y (the first and third taps), which is equivalent to reflecting θ over $\pi/4$ or $(5\pi/4)$, and $\mathbf{w} \rightarrow -\mathbf{w}$ is equivalent to the combination of reflecting ϕ over $\pi/2$ and adding π to $\theta (\mathrm{mod}\ 2\pi)$.

A contour plot of the SAM cost function is shown in Fig. 2. The axes represent normalized values of the spherical coordinates θ and ϕ. The contours are logarithmically spaced to show detail in the valleys. There are four minima, but they all have equivalent values of the SAM cost, due to the equivalence relations $\mathbf{w} \Leftrightarrow -\mathbf{w}$ and $\mathbf{w} \Leftrightarrow \bar{\mathbf{w}}$.

We compare the locations of these minima to those of a traditional channel shortening cost function: the shortening SNR (SSNR) [3]. The SSNR is defined as

$$\mathrm{SSNR} = \frac{\mathbf{c}_{\mathrm{win}}^H\mathbf{c}_{\mathrm{win}}}{\mathbf{c}_{\mathrm{wall}}^H\mathbf{c}_{\mathrm{wall}}} \tag{20}$$

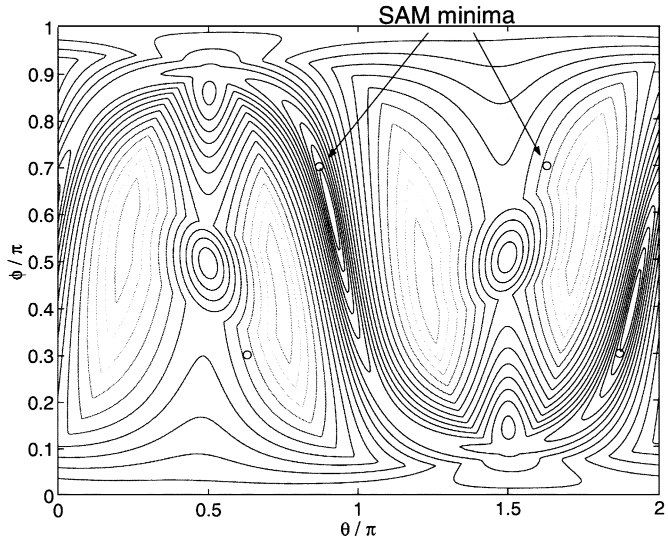


Fig. 3. Contours of the $1/\mathrm{SSNR}$ cost function. The four circles are the global minima of the SAM cost function.

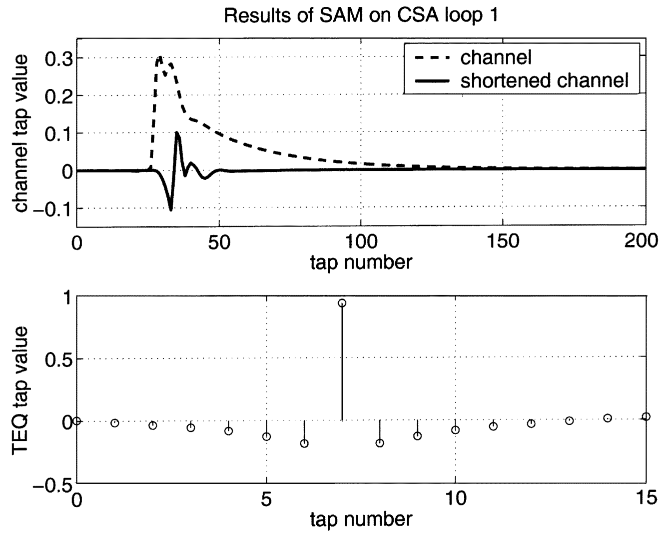


Fig. 4. Channel (dashed) and shortened channel (solid) impulse responses. The shortened channel should have 33 taps.

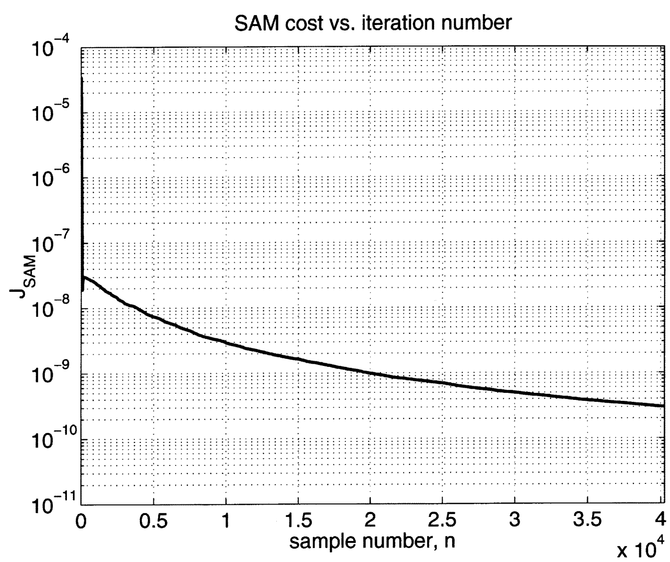


Fig. 5. SAM cost versus iteration (not symbol) number for 40 dB SNR.

where $\mathbf{c}_{\rm win}$ is the effective channel impulse response inside the window of interest (of width $\nu+1$), and $\mathbf{c}_{\rm wall}$ is the effective channel impulse response outside this window. Thus, for our 5-tap effective channel, we pick a 2-tap window and compute the energy of these taps and then divide by the energy of the remaining three taps. For each equalizer setting, we will compute the combined channel, pick the 2-tap window with the highest SSNR, and then plot the inverse of that value (so that we are looking for minima rather than maxima). Contours of this cost function are shown in Fig. 3. Comparison of the two contour plots show that the pair of global minima of $1/\mathrm{SSNR}$ match up nicely with two of the global minima of the SAM cost. Thus, if we find a pair of global minima of the SAM cost and they have a high value of $1/\mathrm{SSNR}$, we can fix this by switching to the other global minima of SAM simply by reversing the order of taps in $\mathbf{w}$.

VI. Simulations

This section provides a numerical performance assessment of SAM in an ADSL environment. All of the Matlab code is available in [20]. Parameters were chosen to match the standard for ADSL downstream transmission: The cyclic prefix ν was 32 samples, the FFT size was 512, the equalizer (TEQ) had 16 taps, and the channel was CSA test loop 1 [1], which is available in [21]. The noise power was set such that $\sigma_x^2\|\mathbf{h}\|^2/\sigma_v^2 = 40$ dB. We used 75 symbols (of 544 samples each), and SAM used the auto-regressive implementation of (16) with $\alpha = 1/100$ and with the unit norm equalizer constraint. The initialization was a single spike, and the step size was 5 (such a large step size can be used because the SAM cost is very small, so the update size is still small). SAM is compared with the maximum shortening SNR solution [3], which was obtained using the code at [21], and the matched filter bound (MFB) on capacity, which assumes no ICI.

Two types of noise are considered: white Gaussian noise and near-end cross-talk (NEXT) [13], which is highly colored. The NEXT was generated by exciting a coupling filter with spectrum $|H_{\rm next}(f)|^2 = H_o H_{\rm mask}(f) f^{(3/2)}$ with white noise [14]. The constant H_o was chosen so that the variance of the NEXT was σ_v^2, with σ_v^2 chosen to achieve the desired SNR. The filter $H_{\rm mask}$ is an ADSL upstream spectral mask that passes frequencies between 28 and 138 kHz since the upstream signal is the source of the NEXT for the downstream signal. The code to generate the NEXT was obtained from [21].

Fig. 4 shows the channel and the combined channel-equalizer after running SAM. Figs. 5 and 6 show the SAM cost and achievable bit rate versus the iteration number. The fact that the SAM cost is not monotonically decreasing in the first few hundred samples is because of the renormalization. After each iteration, the equalizer is divided by its norm, and this projection causes the algorithm to no longer be a gradient descent algorithm (though it is approximately so). The bit rate is not monotonically increasing because the SAM cost bears no direct relation to the bit rate. At 340 iterations, SAM achieves 96% of the MFB but then drops and eventually rises again to 74% of the

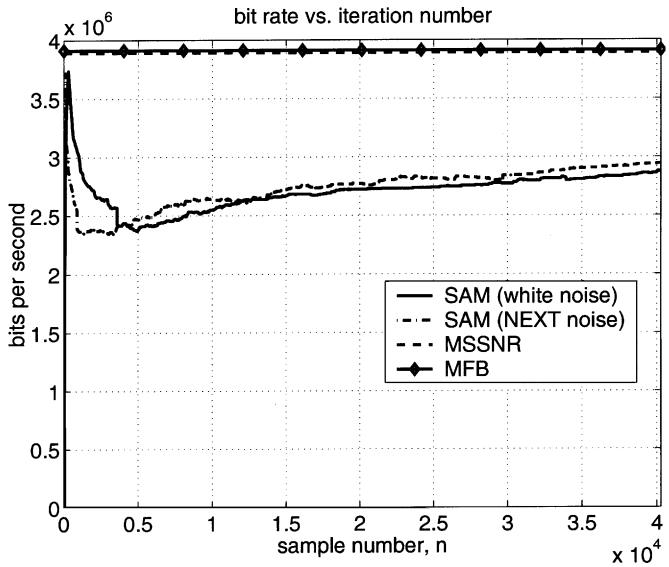


Fig. 6. Achievable bit rate versus iteration number (not symbol number) for 40 dB SNR. The dashed line and the diamonds correspond to the maximum SSNR solution and the matched filter bound.

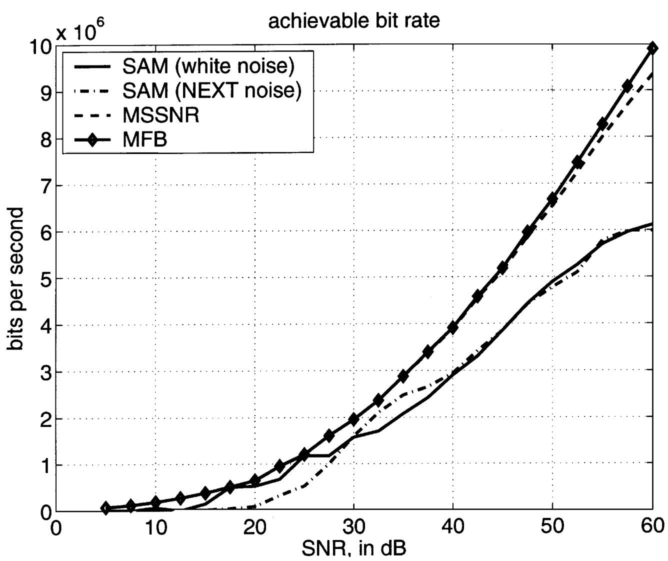


Fig. 7. Achievable bit rate versus SNR for SAM and the maximum SSNR algorithm for white noise and for colored noise (NEXT).

bound. The fact that the SAM cost is steadily decreasing when the bit rate decreases and then increases again is very important It indicates that the SAM minima and the bit rate maxima are not in exactly the same location. Note that SAM performs similarly for white and for colored noise.

Fig. 7 shows the achievable bit rate versus SNR for SAM and for the maximum shortening SNR algorithm of [3] for white noise and for NEXT. The bit rate is determined based on

$$R = \sum_{i=1}^{N_{\text{fft}}} \log_2\left(1 + \frac{\text{SNR}_i}{\Gamma}\right). \tag{21}$$

The bit rate was computed using a 6-dB margin and a 4.2-dB coding gain. For more details, see [4] or [21]. The bit rate was determined for the settings SAM arrived at after 75 DMT symbols. Observe that for low SNR, the performance of SAM and the MSSNR method are comparable, and the performance of SAM degrades (relatively) for high SNR. This is because when the noise is high, SAM only needs to reduce the ICI below the noise floor, but when the SNR is 60 dB, the excess ICI becomes more noticeable. For very low SNR (less than 15 dB for white noise, less than 25 dB for NEXT), the performance of SAM degrades, presumably due to the noiseless assumption in the derivations. However, typical SNR values for ADSL are 40 to 60 dB, and an SNR less than 25 dB is very unusual. BER curves are not included because for ADSL, the bit allocation on each tone is increased until the BER becomes 10^{-7}; therefore, a BER curve would be flat as a function of SNR.

VII. Conclusions and Future Work

A new blind, adaptive channel shortening algorithm based on a windowed sum-squared auto-correlation minimization has been proposed. The effectiveness of the algorithm to blindly shorten the channel has been demonstrated numerically. However, SAM may perform poorly in situations such as when the source sequence is not white or when there is extremely strong cross-talk (or other forms of colored noise).

Proper initialization of the TEQ is necessary to ensure the convergence of the SAM algorithm to a good minima. Further studies are needed to characterize the cost function and formulate suitable design rules to ensure good performance. Robustness of the algorithm to receiver noise and violation of the assumption of source whiteness need to be investigated further as well.

References

[1] K. Sistanizadeh, "Loss characteristics of the proposed canonical ADSL loops with 100-ohm termination at 70, 90, and 120 F," Amer. Nat. Stand. Inst., New York, Washington, DC, ANSI T1E1.4 Committee Contribution, no. 161, Nov. 1991.

[2] D. D. Falconer and F. R. Magee, "Adaptive channel memory truncation for maximum likelihood sequence estimation," *Bell Syst. Tech. J.*, pp. 1541–1562, Nov. 1973.

[3] P. J. W. Melsa, R. C. Younce, and C. E. Rohrs, "Impulse response shortening for discrete multitone transceivers," *IEEE Trans. Commun.*, vol. 44, pp. 1662–1672, Dec. 1996.

[4] G. Arslan, B. L. Evans, and S. Kiaei, "Equalization for discrete multitone receivers to maximize bit rate," *IEEE Trans. Signal Processing*, vol. 49, pp. 3123–3135, Dec. 2001.

[5] W. Lesch, "Impulse response shortening for OFDM in a single frequency network," M.S. Thesis, Royal Inst. Technol., Stockholm, Sweden, 1998.

[6] N. Al-Dhahir and J. M. Cioffi, "Optimum finite-length equalization for multicarrier transceivers," *IEEE Trans. Commun.*, vol. 44, pp. 56–64, Jan. 1996.

[7] B. Farhang-Boroujeny and M. Ding, "Design methods for time-domain equalizers in DMT transceivers," *IEEE Trans. Commun.*, vol. 49, pp. 554–562, Mar. 2001.

[8] K. Van Acker, G. Leus, M. Moonen, O. van de Wiel, and T. Pollet, "Per tone equalization for DMT-based systems," *IEEE Trans. Commun.*, vol. 49, pp. 109–119, Jan. 2001.

[9] T. Pollet, M. Peeters, M. Moonen, and L. Vandendorpe, "Equalization for DMT-based broadband modems," *IEEE Commun. Mag.*, vol. 38, pp. 106–113, May 2000.

[10] L. Vandendorpe, J. Louveaux, B. Maison, and A. Chevreuil, "About the asymptotic perofrmance of MMSE MIMO DFE for filter-bank based multicarrier transmission," *IEEE Trans. Commun.*, vol. 47, pp. 1472–1475, Oct. 1999.

[11] S. Trautmann and N. J. Fliege, "Perfect equalization for DMT systems without guard interval," *IEEE J. Select. Areas Commun.*, vol. 20, pp. 987–996, June 2002.

[12] J. S. Chow, J. M. Cioffi, and J. A. C. Bingham, "Equalizer training algorithms for multicarrier modulation systems," *Proc. IEEE Int. Conf. Commun.*, pp. 761–765, May 1993.

[13] T. Starr, J. M. Cioffi, and P. T. Silverman, *Understanding Digital Subscriber Line Technology*. Upper Saddle River, NJ: Prentice-Hall, 1999.

[14] N. Lashkarian and S. Kiaei, "Optimum equalization of multicarrier systems: A unified geometric approach," *IEEE Trans. Commun.*, vol. 49, pp. 1762–1769, Oct. 2001.

[15] W. Henkel and T. Kessler, "Maximizing the channel capacity of multicarrier transmission by suitable adaptation of the time-domain equalizer," *IEEE Trans. Commun.*, vol. 48, pp. 2000–2004, Dec. 2000.

[16] M. de Courville, P. Duhamel, P. Madec, and J. Palicot, "Blind equalization of OFDM systems based on the minimization of a quadratic criterion," in *Proc. Int. Conf. Commun.*, Dallas, TX, June 1996, pp. 1318–1321.

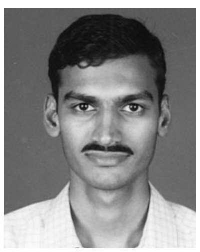
[17] R. K. Martin, J. Balakrishnan, W. A. Sethares, and C. R. Johnson, Jr., "A blind, adaptive TEQ for multicarrier systems," *IEEE Signal Processing Lett.*, vol. 9, pp. 341–343, Nov. 2002.

[18] P. P. Vaidyanathan and B. Vrcelj, "Fast and robust blind-equalization based on cyclic prefix," *Proc. IEEE Int. Conf. Commun.*, vol. 1, pp. 1–5, Apr.–May 2002.

[19] C. R. Johnson, Jr., P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: A review," *Proc. IEEE*, vol. 86, pp. 1927–1950, Oct. 1998.

[20] R. K. Martin. Matlab Code for Papers by R. K. Martin. [Online]. Available: http://bard.ece.cornell.edu/matlab/martin/index.html

[21] G. Arslan, M. Ding, B. Lu, Z. Shen, and B. L. Evans. TEQ Design Toolbox. Univ. Texas, Austin, TX. [Online]. Available: http://www.ece.utexas.edu/~bevans/projects/adsl/dmtteq/dmtteq.html

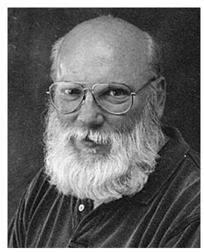


Jaiganesh Balakrishnan (M'03) was born in Madras, India, in 1976. He received the B.Tech. degree from the Department of Electrical Engineering, Indian Institute of Technology, Madras, in 1997 and the M.S. and Ph.D. degrees from the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, in 1999 and 2002, respectively.

He is presently with Texas Instruments, Dallas. His research interests include equalization, detection and estimation, adaptive signal processing, and wireless communications.

Richard K. Martin (S'01) received the B.S. degrees (*Summa Cum Laude*) in physics and electrical engineering from the University of Maryland, College Park, in 1999 and the M.S. degree in electrical engineering from Cornell University, Ithaca, NY, in 2001. He is currently pursuing the Ph.D. degree from Cornell University.

His research focus is adaptive equalization for multicarrier systems.



C. Richard Johnson, Jr. (F'89) was born in Macon, GA, in 1950. He received the B.E.E. degree with high honors from the Georgia Institute of Technology, Atlanta, in 1973 and the M.S.E.E. and Ph.D. degrees in electrical engineering with minors in engineering-economic systems and art history from Stanford University, Stanford, CA, in 1975 and 1977.

He is currently a Professor with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY. His current research interests are in adaptive parameter estimation theory that is useful in applications of digital signal processing to communications systems.

Dr. Johnson was selected by Eta Kappa Nu as the Outstanding Young Electrical Engineer in 1982 and as the C. Holmes MacDonald Outstanding Teacher of 1983. He was elected Fellow of the IEEE for "contributions to adaptive parameter estimation theory with applications in digital control and signal processing." In 1991, he was selected a Distinguished Lecturer of the Signal Processing Society of the IEEE.