Low Complexity MIMO Blind, Adaptive Channel Shortening

Richard K. Martin, John M. Walsh, and C. Richard Johnson, Jr.

Richard K. Martin and C. Richard Johnson, Jr.
Cornell University
School of Electrical & Computer Engineering
Ithaca, NY 14853-3801, USA
{frodo,johnson}@ece.cornell.edu, jmw56@cornell.edu

Low Complexity MIMO Blind, Adaptive Channel Shortening

R. K. Martin, *Member, IEEE,* J. M. Walsh, *Student Member, IEEE,* and C. R. Johnson, Jr., *Fellow, IEEE*

Abstract— **Channel shortening is often employed as a means of mitigating inter-symbol and inter-carrier interference in systems using multicarrier modulation. The MERRY algorithm has previously been shown to blindly and adaptively shorten a channel to the length of the guard interval in a multicarrier system. This paper addresses synchronization and complexity reduction issues which were not dealt with in previous work, and provides extensions to and generalizations of the MERRY algorithm. A modification is presented which removes the square root and division needed at each iteration without introducing additional complexity, with the added benefit of allowing the use of constraints other than a unit norm equalizer; an extension is proposed which allows for the use of more data in the MERRY update; the algorithm is generalized to the MIMO and fractionally-spaced cases; a low-complexity, blind symbol synchronization technique is proposed; and a method is proposed for blind initialization of the algorithm to avoid slow modes of convergence. Each extension to the basic MERRY algorithm is accompanied by an illustrative simulation example.**

Index Terms— **Adaptive, Blind, Channel Shortening, Cyclic Prefix, Equalization, MERRY, Multicarrier.**

I. INTRODUCTION

CHANNEL shortening is a generalization of equalization, since equalization amounts to shortening the channel to length one. Channel shortening was first applied to maximum likelihood sequence estimation (MLSE) [1]. Since the complexity of MLSE grows exponentially with the channel length, a prefilter can be employed to shorten the channel and reduce the complexity [2], [3]. Channel shortening can also be applied to systems employing multicarrier modulation (MCM) [4], single carrier modulation with a cyclic prefix, and even CDMA systems [5].

MCM techniques like orthogonal frequency division multiplexing (OFDM) and discrete multi-tone (DMT) have been deployed in applications such as IEEE 802.11a and HIPER-LAN/2 wireless LANs, Digital Audio and Video Broadcast, digital subscriber loops (DSL), power line communications, and satellite radio. MCM can easily combat channel dispersion when the channel delay spread is not greater than the length

of the cyclic prefix (CP), and when the Doppler spread is negligible. However, when the CP is not long enough, the orthogonality of the sub-carriers is lost, causing inter-carrier interference (ICI) and inter-symbol interference (ISI).

The most common technique for mitigating the ICI/ISI caused by the inadequate CP length is the use of a time-domain equalizer (TEQ) at the receiver front-end [3], [6] – [20]. The TEQ is a filter that shortens the channel so that the delay spread of the effective channel-equalizer impulse response is no larger than the length of the CP. Most TEQs in the literature have been designed in the context of DSL, which runs over twisted pair telephone lines [8], [9], [11], [12]. As a consequence, most of the TEQ designs in the literature are trained and non-adaptive.

Recently, blind and/or adaptive TEQ design has received increasing attention. The MERRY (Multicarrier Equalization by Restoration of RedundancY) algorithm [16], [21] induces channel shortening by restoring the redundancy in the received data that is due to the CP. The algorithm is low-complexity and converges to the minimum MSE solution of [2] (for a white input). The SAM (Sum-squared Auto-correlation Minimization) algorithm [22] attempts to shorten the auto-correlation of the TEQ output sequence. Although SAM converges quickly, it is multimodal and computationally intensive. To avoid multi-modality, a blind, non-adaptive fractionally-spaced TEQ that is also based on second-order output statistics was proposed in [23], [24]. A "carrier nulling algorithm" (CNA) was proposed in [25], [26], which exploits the fact that many MCM systems transmit zeros on some input tones at the band edges. The TEQ can be adapted blindly to force the corresponding output tones to zero. In [25], it was shown that CNA equalizes the channel to a single spike (i.e. an impulse), rather than shortening it to a window, hence CNA is primarily suited to MCM systems that do not use a CP [25].

In many cases, the receiver may need to jointly shorten multiple channels using a single TEQ. In a multicarrier code division multiple access (MC-CDMA) system, multiple users each spread their signals using a spreading code before multi-carrier modulation takes place [27]. To enhance performance, the receiver can jointly shorten all of the users' channels to mitigate ISI before de-spreading takes place. In DSL, each modem receives the desired signal as well as crosstalk from other signals in the same cable bundle. In this case, joint channel shortening can be combined with multiuser detection to improve the receiver's performance. If a DSL system is operating in echo cancelling mode [6], then the channel and the echo path must be jointly shortened [8], [28]. As another example, multiple receive antennas or oversampling of the received data may be employed, leading to multiple outputs

$r_1(k)$
$r_P(k)$
$y_1(k)$
$y_P(k)$
$y(k)$

Fig. 1. SISO multicarrier system model. (I)FFT: (inverse) fast Fourier transform, P/S: parallel to serial, S/P: serial to parallel, CP: add cyclic prefix, and xCP: remove cyclic prefix.

for each input. This motivates a multiple input, multiple output (MIMO) system model, in which multiple channels need to be shortened simultaneously. Joint channel shortening has been studied in [8], [28] – [33]. However, these works involved extending the training-based, non-adaptive single-input, single output (SISO) TEQs proposed in [2], [3], and [8] to the MIMO case; whereas we consider the blind, adaptive MIMO case.

The contributions of this paper are as follows:

- a generalization of MERRY to the MIMO and fractionally-spaced cases,
- a modification to the MERRY algorithm which removes the square root and division at each iteration without introducing additional complexity, and which allows for constraints that may be more appropriate than the unit norm equalizer constraint used in [16],
- a method to increase the amount of usable data (the basic MERRY algorithm only uses one sample per block),
- a low-complexity symbol synchronization method (an issue not addressed in [16]),
- and a method for blind initialization of MERRY to avoid slow modes of convergence.

This paper is organized as follows. Section II reviews the system model and notation. Section III proposes the modifications and extensions of the MERRY algorithm. Section IV provides a method for choosing the symbol synchronization parameter. Section V provides illustrative simulations, and Section VI concludes the paper.

II. SYSTEM MODEL AND NOTATION

The SISO multicarrier system model is shown in Fig. 1. The input stream is divided into blocks of N bins, and each bin is viewed as a quadrature amplitude modulated (QAM) signal that will be modulated by a different carrier. The modulation can be efficiently implemented in discrete time via an inverse fast Fourier transform (IFFT), which converts the frequency-domain data into a time-domain signal. After transmission through a dispersive channel $\mathbf{h}$, the receiver can use an FFT to recover the data.

If the received data is a circular convolution of the channel and transmitted data, then the received frequency-domain output is a *pointwise* multiplication of the transmitted frequency-domain data with the discrete Fourier transform (DFT) of the channel. Since the convolution is actually linear rather than circular, it is made to appear circular by adding a cyclic prefix (CP) to the start of each data block. The CP is obtained by repeating the last ν samples of each block at the beginning of

TABLE I
MIMO CHANNEL SHORTENING NOTATION

Notation	Definition
$x_l(k)$	transmitted signal for user l
$n_p(k)$	additive noise on p^{th} received sequence
$r_p(k)$	p^{th} sequence of received data
$y_p(k)$	output of TEQ p
$y(k)$	recombined output $= \sum_p y_p(k)$
N, ν, M	sizes of FFT, CP, and symbol
Δ	transmission delay
$\mathbf{h}_{p,l}$	channel from user l to receiver p
$\mathbf{w}_p$	p^{th} TEQ impulse response
$\mathbf{c}_{p,l}$	effective channel $= \mathbf{h}_{p,l} \star \mathbf{w}_p$
L_h, L_w, L_c	order of $\mathbf{h}$, $\mathbf{w}$, or $\mathbf{c}$
$\tilde{L}_h = L_h + 1$	length of $\mathbf{h}$, e.g.
$\mathbf{R}_{n,p}$	autocorrelation matrix of $n_p(k)$
$\mathbf{A}^*, \mathbf{A}^T, \mathbf{A}^H$	conjugate, transpose, and Hermitian

the block. If the CP length is shorter than the channel, then the convolution appears to be circular and the signals in the bins can be equalized by a bank of complex gains, referred to as a frequency domain equalizer (FEQ) [6].

Since transmitting the CP wastes time that could be used to transmit data, the CP is usually set to a reasonably small value, and a TEQ $\mathbf{w}$ is employed to shorten the channel to this length. Trained, non-adaptive TEQ designs have been well explored [2], [3], [6] – [19].

Table I gives the MIMO notation, and Fig. 2 shows the time-domain portion of the MIMO system. We assume the use of L transmit antennas (with up to L users) and P receive antennas (or oversampling by P). The received signal $r_p(k)$ from antenna p (or the pth sub-sampling sequence), $p \in \{1, \cdots, P\}$, is obtained by passing each signal from user $l \in \{1, \cdots, L\}$ through channel $h_{p,l}$, adding the L channel outputs, and adding noise sequence $n_p(k)$. In an MC-CDMA system, each user's signal $x_l(k)$ is obtained by spreading one or more symbols in the frequency domain, taking an IFFT, and adding a CP; see [27], e.g. If the multiple users arise due to cross-talk in a wireline system, then each signal $x_i(k)$ is generated in the manner of Fig. 1.

After the CP is added, the first and last ν samples are identical in each transmitted symbol,

$$x_l(Mk + i) = x_l(Mk + i + N),$$
$$i \in \{1, \ldots, \nu\}, \quad l \in \{1, \ldots, L\}, \tag{1}$$
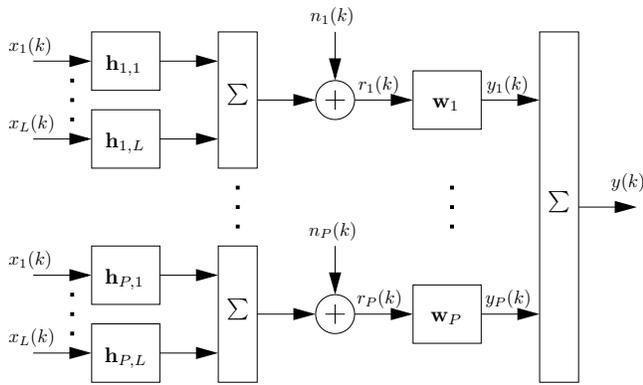
Fig. 2.   MIMO TEQ model, for L transmitters and P receive antennas (or oversampling by a factor of P).

where $M = N + \nu$ is the total symbol duration and k is the symbol index. The received data r_p is obtained from $\{x_l : l = 1, \cdots, L\}$ by

$$r_p(Mk+i) = \sum_{l=1}^{L} \sum_{j=0}^{L_h} h_{p,l}(j)\, x_l(Mk+i-j) + n_p(Mk+i),$$
$$(2)$$

and y_p, the output of TEQ p, is obtained from r_p by

$$y_p(Mk+i) = \sum_{j=0}^{L_w} w_p(j)\, r_p(Mk+i-j) \qquad (3)$$

$$= \sum_{l=1}^{L} \sum_{j=0}^{L_c} c_{p,l}(j)\, x_l(Mk+i-j)$$

$$+ \sum_{j=0}^{L_w} w_p(j)\, n_p(Mk+i-j). \qquad (4)$$

Then the final, recombined output is obtained by

$$y(Mk+i) = \sum_{p=1}^{P} y_p(Mk+i) \qquad (5)$$

$$= \sum_{l=1}^{L} \sum_{j=0}^{L_c} c_l(j) x_l(kM+i-j)$$

$$+ \sum_{p=1}^{P} \sum_{j=0}^{L_w} w_p(j)\, n_p(Mk+i-j), \qquad (6)$$

where

$$c_l(j) = \sum_{p=1}^{P} c_{p,l}(j), \quad j \in \{0, \ldots, L_c\}, \quad l \in \{1, \ldots, L\}. \quad (7)$$

One could either work with the collection of P sequences $\{y_p(k)\}$ or the single output sequence $y(k)$. The weights for the linear combination in (5) have implicitly been absorbed into the P TEQs. Each of the $P \cdot L$ channels is modeled as an FIR filter of length L_h+1, each of the P TEQs is an FIR filter of length L_w+1, and each effective channel $\mathbf{c}_{p,l} = \mathbf{h}_{p,l} \star \mathbf{w}_p$ has length L_c+1, where $L_c = L_h+L_w$. The symbol $\star$ denotes linear convolution.

III. EXTENSIONS TO MERRY

This section discusses several extensions to the MERRY algorithm. Section III-A proposes and analyzes the FRODO cost function as a generalization of the MERRY cost function. The generalization allows for a MIMO design, the use of more than one sample in the update rule, and channel shortening to variable window lengths. Section III-B re-casts the optimal solution in several formulations that are equivalent but quite different in appearance. Section III-C derives the algorithm, with modifications (relative to the method used in [16]) to remove the square root and division in the update rule and to allow the use of alternate constraints.

A. FRODO: generalizing the MERRY cost function

Throughout this section we make the following assumptions:

1. The L input sequences of the IFFTs are each zero-mean, white, and wide sense stationary (implying that the output bins of each IFFT are uncorrelated), with transmit power $\sigma_{x,l}^2$.
2. $L_c + 1 \leq N$ (the length of the effective channel is no larger than the FFT size).
3. For each p, the noise autocorrelation function $\mathbf{R}_{n,p}(\delta) = 0$ for $\delta \geq N - L_w$.
4. All data sequences x_l and noise sequences n_p are pairwise uncorrelated.

An uncorrelated IFFT input produces an uncorrelated IFFT output because the DFT matrix is unitary. We may alter assumption 4 in some cases to set $x_1 = x_2 = \cdots = x_L$, i.e. for the case of oversampling. If there is one user with L antennas, assumption 4 may still hold if space-time coding is applied.

The idea behind MERRY is that if the channel length $L_h + 1 \leq \nu$, then the last sample in the CP should match the last sample in the symbol. The MERRY cost function reflects this goal:

$$J_{merry} = \mathrm{E}\left[\left| y(Mk + \nu + \Delta) - y(Mk + \nu + N + \Delta) \right|^2 \right],$$
$$\Delta \in \{0, \ldots, M-1\}, \qquad (8)$$

where the symbol synchronization parameter Δ represents the desired delay. Minimizing (8) minimizes the energy of the channel outside of a length-ν window [21]. Since there are ν samples in the cyclic prefix, a natural generalization is to compare more than one of these samples to their counterparts at the end of the symbol. Thus, a more general cost function is

$$J_{frodo} = \sum_{i \in S_f} \mathrm{E}\left[\left| y(Mk+i+\Delta) - y(Mk+i+N+\Delta) \right|^2 \right],$$
$$\Delta \in \{0, \ldots, M-1\}, \qquad (9)$$

where $S_f \subset \{1, \cdots, \nu\}$ is an index set. For MERRY, $S_f = \{\nu\}$. Different sets allow for the use of more or less data, as well as for shortening to different channel lengths, which will be shown momentarily. Since the modified cost function allows the option of using all of the data in the CP ($S_f = \{1, \cdots, \nu\}$), or a single sample ($S_f = \{\nu\}$), or anything in

between, we use the name Forced Redundancy with Optional Data Omission (FRODO) to refer to the algorithm using this cost function. An equalization (not channel shortening) algorithm equivalent to using FRODO with the set $S_f = \{1, \cdots, \nu\}$ was proposed in [34]. The general FRODO cost function includes the cost functions of [34] and MERRY [16] as special cases. We now analyze the general FRODO cost function as a means of demonstrating its utility.

Theorem 1: The FRODO cost function (9) simplifies to

$$J_{frodo} = 2 \sum_{i \in S_f} \sum_{l=1}^{L} \sigma_{x,l}^2 \|\mathbf{c}_{l,wall}^{i+\Delta}\|^2 + 2 |S_f| \sum_{p=1}^{P} \mathbf{w}_p^H \mathbf{R}_{n,p} \mathbf{w}_p, \tag{10}$$

where $|S_f|$ is the cardinality of the set S_f, and

$$\|\mathbf{c}_{l,wall}^{i+\Delta}\|^2 = \sum_{j=0}^{\Delta+i-\nu-1} |c_l(j)|^2 + \sum_{j=\Delta+i}^{L_c} |c_l(j)|^2, \quad l \in \{1, \ldots, L\}. \tag{11}$$

with $c_l(j)$ as in (7).

Proof: For simplicity, we consider the case $\sigma_{x,l}^2 = \sigma_x^2 \ \forall l$, though the general case is a simple extension. Consider eq. (6) in [16],

$$J_\delta = 2 \sigma_x^2 \left(\sum_{j=0}^{\delta-1} |c_j|^2 + \sum_{j=\nu+\delta}^{T+L-2} |c_j|^2 \right) + 2 \ \mathbf{w}^H \mathbf{R}_n \mathbf{w}, \tag{12}$$

and make the following substitutions: $\delta \to \Delta + i - \nu$, $\mathbf{c} \to \mathbf{c}_l$, $T \to L_w + 1$, $L \to L_h + 1$, and

$$\mathbf{w} = \left[\mathbf{w}_1^T, \mathbf{w}_2^T, \cdots, \mathbf{w}_P^T \right]^T, \tag{13}$$

$$\mathbf{n}(j) = [n_1(j), \cdots, n_1(j - L_w), \cdots,$$
$$n_P(j), \cdots, n_P(j - L_w)]^T, \tag{14}$$

$$\mathbf{R}_n = \mathrm{E} \left[\mathbf{n}^*(Mk + i) \ \mathbf{n}^T(Mk + i) \right]. \tag{15}$$

Since the noise sequences are uncorrelated, the result follows by summing over $i \in S_f$. ∎

Remarks on Theorem 1: For the case $L = P = 1$, and for $S_f = \{\nu\}$, we have the term $i = \nu$ only. Thus, the cost function is the power of the "wall" of the channel (as opposed to a "window"), and minimizing this cost function leads to shortening to a ν-length window. For the case $L = 1$, $P > 1$, and $S_f = \{\nu\}$, the cost function suppresses the tails of the averaged channel, $\mathbf{c} = \sum_p \mathbf{c}_p$, allowing for diversity gain. For the case $L > 1$, $P = 1$, and $S_f = \{\nu\}$, the cost function suppresses the average of the tail energies of the L channels (rather than the tail energy of the average channel in the previous case), thus shortening all L channels at once. In an MC-CDMA scenario, the L demodulated signals can now be separated using the spreading codes, which would not have been possible if the channels were not shortened first. The case $L = P = 1$ with $S_f \supsetneq \{\nu\}$ will be discussed after Theorem 2.

Theorem 2: If we relax assumption 4 so that $x_l(k) = x(k) \ \forall l \in \{1, \cdots, L\}$, i.e. multiple transmit antennas for a single user, then the FRODO cost function (9) simplifies to

$$J_{frodo} = 2\sigma_x^2 \sum_{i \in S_f} \|\mathbf{c}_{wall}^{i+\Delta}\|^2 + 2 |S_f| \sum_{p=1}^{P} \mathbf{w}_p^H \mathbf{R}_{n,p} \mathbf{w}_p, \tag{16}$$

where

$$\|\mathbf{c}_{wall}^{i+\Delta}\|^2 = \sum_{j=0}^{\Delta+i-\nu-1} |c(j)|^2 + \sum_{j=\Delta+i}^{L_c} |c(j)|^2, \tag{17}$$

and where

$$c(j) = \sum_{p=1}^{P} \sum_{l=1}^{L} c_{p,l}(j), \quad j \in \{0, \ldots, L_c\}. \tag{18}$$

Remarks on Theorem 2: The proof of Theorem 2 parallels the proof of Theorem 1, except with the definition of (18) instead of (7), hence the details of the proof are omitted. The cost function is the tail energy of the average channel (averaged over p and l).

The effect of using more than one comparison [more than one value of i in (9)] can be illustrated as follows. For simplicity, let $L = P = 1$ and drop the subscripts l and p. Consider using the "full" index set, $i \in \{1, \cdots, \nu\}$. From Theorem 1, the cost function is

$$J_{frodo} = 2\sigma_x^2 \sum_{i=1}^{\nu} \|\mathbf{c}_{wall}^{i+\Delta}\|^2 + 2 \sum_{i=1}^{\nu} \mathbf{w}^H \mathbf{R}_n \mathbf{w} \tag{19}$$

$$= 2 \sigma_x^2 \left[\nu \ c^2(0) + \nu \ c^2(1) + \cdots + \nu \ c^2(\Delta - \nu) \right.$$
$$+ (\nu - 1) \ c^2(\Delta - \nu + 1) + \cdots + 2 \ c^2(\Delta - 2) + c^2(\Delta - 1)$$
$$+ c^2(\Delta + 1) + 2 \ c^2(\Delta + 2) + \cdots + (\nu - 1) \ c^2(\Delta + \nu - 1)$$
$$\left. + \nu \ c^2(\Delta + \nu) + \nu \ c^2(\Delta + \nu + 1) + \cdots + \nu \ c^2(L_c) \right]$$
$$+ 2 \ \nu \ \mathbf{w}^H \mathbf{R}_n \mathbf{w}. \tag{20}$$

Fig. 3 shows a pictorial example for $\nu = 4$. Each time i is incremented, the window location shifts over by one sample. Thus, the "full" FRODO cost function tries to suppress all of the taps of the effective channel except for the Δth tap, and the noise gain is limited as well. Taps farther from the center are more heavily weighted, and hence should be smaller. This makes the cost function very similar to the minimum delay spread (MDS) algorithm [14] which minimizes

$$J_{MDS} = \sum_{j=0}^{L_c} |j - \Delta|^2 \ c^2(j), \tag{21}$$

force to zero	don't care	force to zero
force to zero	don't care	force to zero
force to zero	don't care	force to zero
force to zero	don't care	force to zero
don't care		
force to zero	↓	force to zero

weighting	4	4	4	3	2	1	0	1	2	3	4	4	4

$i = 1$, $i = 2$, $i = 3$, $i = 4 = \nu$, summed

$\mathbf{c} = \mathbf{h} \star \mathbf{w}$

Fig. 3. The relation of the "don't care" windows in the different terms of the FRODO cost function, for $\nu = 4$. The line "summed" indicates the effect of considering all four terms at once, and the line "weighting" indicates how much emphasis the total cost function places on forcing each tap to zero.

subject to a unit norm constraint on the effective channel, $\|\mathbf{c}\| = 1$. A variant of the MDS algorithm proposed in [15] uses linear weights, rather than quadratic:

$$\hat{J}_{MDS} = \sum_{j=0}^{L_c} |j - \Delta| \, c^2(j). \tag{22}$$

This alternate MDS penalty increases linearly with the distance from tap Δ, and the FRODO penalty [in this example which uses all possible values of i in the summation in (9)] increases linearly for a distance of ν on each side of tap Δ, and then remains fixed at that penalty level for larger distances. Consequently, the "full" FRODO algorithm will attempt to suppress all taps save the Δth tap, but with a tendency to minimize delay spread rather than to equalize. Since the ICI and ISI caused by channel taps increase with their distance from the Δth tap [14], minimizing the delay spread is more advantageous than equalizing to an impulse, yet less advantageous than an algorithm explicitly designed for channel shortening (such as the MSSNR design [8]). The linearly increasing penalty function of (22) may amplify the effects of channel estimation errors [19]. The leveling off of the FRODO penalty function mitigates this.

If a window size between 1 and ν is desired, then the index set for FRODO can be changed accordingly. Moreover, if two comparisons are made in (9) rather than one as in (8), then the algorithm has access to more data per update. Hence, convergence should be faster and smoother. The penalty is that the window size will be smaller by one sample. However, if ν is large (as in DVB, where ν can be as large as 2048; or even in ADSL, in which $\nu = 32$), then the difference between a ν-length window and a $(\nu - 1)$-length window is minor.

B. Equivalent problem statements

This section formulates the FRODO design problem in a variety of mathematically equivalent ways. This equivalence will be used in Section III-C to pick a problem statement which allows for an update rule that has no divisions or square roots. Defining the "stacked" vectors

$$\mathbf{r}_p(j) = [r_p(j), r_p(j-1), \cdots, r_p(j-L_w)]^T,$$
$$p \in \{1, \cdots, P\}, \tag{23}$$

$$\mathbf{r}(j) = \left[\mathbf{r}_1^T(j), \mathbf{r}_2^T(j), \cdots, \mathbf{r}_P^T(j)\right]^T, \tag{24}$$

$$\tilde{\mathbf{r}}_i(k) = \mathbf{r}(Mk + i + \Delta) - \mathbf{r}(Mk + i + N + \Delta),$$
$$\forall i \in S_f \tag{25}$$

with $\mathbf{w}$ as in (13), the FRODO cost function (9) can be rewritten as

$$J_{frodo} = \sum_{i \in S_f} \mathrm{E}\left[\left|\tilde{\mathbf{r}}_i^T(k)\mathbf{w}\right|^2\right] \tag{26}$$

$$= \mathbf{w}^H \underbrace{\sum_{i \in S_f} \underbrace{\mathrm{E}\left[\tilde{\mathbf{r}}_i^*(k)\tilde{\mathbf{r}}_i^T(k)\right]}_{\mathbf{A}_i}}_{\mathbf{A}} \mathbf{w}. \tag{27}$$

Under our four assumptions, it can be shown that

$$\mathbf{A}_i = 2\sum_{l=1}^{L} \sigma_{x,l}^2 \begin{bmatrix} \mathbf{H}_{1,l,wall}^H \mathbf{H}_{1,l,wall}, & \cdots, & \mathbf{H}_{1,l,wall}^H \mathbf{H}_{P,l,wall} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{P,l,wall}^H \mathbf{H}_{1,l,wall}, & \cdots, & \mathbf{H}_{P,l,wall}^H \mathbf{H}_{P,l,wall} \end{bmatrix}$$
$$+ 2 \begin{bmatrix} \mathbf{R}_{n,1}, & \cdots, & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0}, & \cdots, & \mathbf{R}_{n,P} \end{bmatrix} \tag{28}$$

where $\mathbf{H}_{p,l,wall}$ is obtained from the convolution matrix $\mathbf{H}_{p,l}$ for channel p, l by removing rows $\Delta + i - \nu$ through $\Delta + i - 1$, similar to $\mathbf{H}_{wall}$ in [8]. (The proof is quite similar to the proof of Theorem 1, and hence is omitted.) We wish to minimize (27), with some constraint to avoid the trivial solution $\mathbf{w} = \mathbf{0}$.

The FRODO cost function is a measure of the energy in the "wall" portion of the effective channel. Also of interest are the energy in the "window" portion of the effective channel, and the total energy of the effective channel. To this end, we define

$$J_{win} = 2 \sum_{i \in S_f} \mathrm{E}\left[y^*(Mk + i + \Delta) \, y(Mk + i + N + \Delta)\right]$$

$$= \mathbf{w}^H \underbrace{\left(\sum_{i \in S_f} \underbrace{2\,\mathrm{E}\left[\mathbf{r}^*(Mk + i + \Delta)\,\mathbf{r}^T(Mk + i + N + \Delta)\right]}_{\mathbf{B}_i}\right)}_{\mathbf{B}} \mathbf{w},$$
$$\tag{29}$$

and

$$J_{total} = 2\,|S_f|\,\mathrm{E}\left[|y(Mk + i_o + \Delta)|^2\right], \quad i_o \in \{0, \cdots, M-1\}$$
$$= \mathbf{w}^H \underbrace{\left(2\,|S_f|\,\mathrm{E}\left[\mathbf{r}^*(Mk + i_o + \Delta)\,\mathbf{r}^T(Mk + i_o + \Delta)\right]\right)}_{\mathbf{C}} \mathbf{w}. \tag{30}$$

Here, i_o is simply a specific (but arbitrary) value of i representing the time at which the power is calculated in (30). It is arbitrary since (30) does not depend on i_o. It can be shown that the $\mathbf{B}_i$ matrices have the same form as the $\mathbf{A}_i$ matrices in (28), except with $\mathbf{H}_{p,l,wall}$ replaced by $\mathbf{H}_{p,l,win}$ (which equals rows $\Delta + i - \nu$ through $\Delta + i - 1$ of the channel convolution matrix $\mathbf{H}_{p,l}$), and the $\mathbf{C}$ matrix has the same form as the $\mathbf{A}_i$ matrices, except with $\mathbf{H}_{p,l,wall}$ replaced by $\mathbf{H}_{p,l}$.

Theorem 3: Under the assumptions in Section III-A, the following optimization problems all produce the same solution $\mathbf{w}_{opt}$, up to a scale factor:

$$\mathbf{w}_{opt}^1 = \arg\min_{\mathbf{w}} J_{frodo} \text{ such that } J_{win} = 1 \tag{31}$$

$$\mathbf{w}_{opt}^2 = \arg\max_{\mathbf{w}} J_{win} \text{ such that } J_{frodo} = 1 \tag{32}$$

$$\mathbf{w}_{opt}^3 = \arg\min_{\mathbf{w}} J_{frodo} \text{ such that } J_{total} = 1 \tag{33}$$

$$\mathbf{w}_{opt}^4 = \arg\max_{\mathbf{w}} J_{total} \text{ such that } J_{frodo} = 1 \tag{34}$$

$$\mathbf{w}_{opt}^5 = \arg\min_{\mathbf{w}} J_{total} \text{ such that } J_{win} = 1 \tag{35}$$

$$\mathbf{w}_{opt}^6 = \arg\max_{\mathbf{w}} J_{win} \text{ such that } J_{total} = 1. \tag{36}$$

Proof: For simplicity of notation, we consider the case $S_f = \{\nu\}$. The general case is straightforward but more tedious. Let $\mathbf{u}_1 = \mathbf{r}(Mk + \nu + \Delta)$, $\mathbf{u}_2 = \mathbf{r}(Mk + \nu + N + \Delta)$, and $\mathbf{u}_3 = \mathbf{u}_1 - \mathbf{u}_2$. Note that $\mathrm{E}\left[\mathbf{u}_1^*\mathbf{u}_1^T\right] = \mathrm{E}\left[\mathbf{u}_2^*\mathbf{u}_2^T\right]$ and

$E\left[\mathbf{u}_1^*\mathbf{u}_2^T\right] = E\left[\mathbf{u}_2^*\mathbf{u}_1^T\right]$ for uncorrelated source sequences, under our assumption that $L_c + 1 \leq N$. Then

$$\underbrace{E\left[\mathbf{u}_3^*\mathbf{u}_3^T\right]}_{\mathbf{A}} = E\left[\mathbf{u}_1^*\mathbf{u}_1^T + \mathbf{u}_2^*\mathbf{u}_2^T - \mathbf{u}_1^*\mathbf{u}_2^T - \mathbf{u}_2^*\mathbf{u}_1^T\right]$$

$$= \underbrace{E\left[2\mathbf{u}_1^*\mathbf{u}_1^T\right]}_{\mathbf{C}} - \underbrace{E\left[2\mathbf{u}_1^*\mathbf{u}_2^T\right]}_{\mathbf{B}}, \qquad (37)$$

i.e. $\mathbf{A} + \mathbf{B} = \mathbf{C}$. Thus,

$$\mathbf{w}_{opt}^1 = \arg\min_{\mathbf{w}} \frac{\mathbf{w}^H\mathbf{A}\mathbf{w}}{\mathbf{w}^H\mathbf{B}\mathbf{w}} \qquad (38)$$

$$= \arg\max_{\mathbf{w}} \frac{\mathbf{w}^H\mathbf{B}\mathbf{w}}{\mathbf{w}^H\mathbf{A}\mathbf{w}} = \mathbf{w}_{opt}^2 \qquad (39)$$

$$= \arg\max_{\mathbf{w}} \left(\frac{\mathbf{w}^H\mathbf{B}\mathbf{w}}{\mathbf{w}^H\mathbf{A}\mathbf{w}} + \underbrace{\frac{\mathbf{w}^H\mathbf{A}\mathbf{w}}{\mathbf{w}^H\mathbf{A}\mathbf{w}}}_{1} \right) \qquad (40)$$

$$= \arg\max_{\mathbf{w}} \frac{\mathbf{w}^H\mathbf{C}\mathbf{w}}{\mathbf{w}^H\mathbf{A}\mathbf{w}} = \mathbf{w}_{opt}^4. \qquad (41)$$

The remaining equivalence relations are proven in a similar fashion. ∎

Thus, we can transform our original constrained minimization problem into various constrained maximization problems. Chatterjee, *et al.* [35], have proposed an iterative algorithm which can solve maximization problems of the form of (32), (34), (36), rather than the form of (31), (33), (35). We will combine Chatterjee's algorithm with the results of Theorem 3 in the next section.

C. Division-free update rule

This section derives the FRODO algorithm that is free of the periodic square root and division of other adaptive and iterative TEQ designs [2], [7], [16], [22], [25]. An iterative generalized eigen-decomposition algorithm was proposed in [35] for neural networks, and it was applied to trained, iterative (not adaptive) TEQ design in [19]. The update algorithm is a gradient ascent of a cost function $(\mathbf{w}^H\mathbf{C}\mathbf{w})$ with a Lagrangian constraint $(\mathbf{w}^H\mathbf{A}\mathbf{w} = 1)$, and it is given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\left(\mathbf{C}\mathbf{w} - \mathbf{A}\mathbf{w}\left(\mathbf{w}^H\mathbf{C}\mathbf{w}\right)\right). \qquad (42)$$

This algorithm is globally convergent to the maximum generalized eigenvalue and eigenvector of the matrix pencil [36] $(\mathbf{C}, \mathbf{A})$ for real parameters [35]. In our case, we have blind, stochastic approximations of $\mathbf{A}$ and $\mathbf{C}$ available at the receiver, which are obtained by removing the expectations in (27) and (30). Combining these estimates with (42), the FRODO update rule is

---

Given Δ and i_o, for symbol $k = 0, 1, 2, \ldots$,

$$\tilde{\mathbf{r}}_i(k) = \mathbf{r}(Mk + i + \Delta) - \mathbf{r}(Mk + i + N + \Delta),$$
$$\forall i \in S_f$$
$$e_i(k) = \mathbf{w}^T(k)\,\tilde{\mathbf{r}}_i(k), \quad \forall i \in S_f$$
$$y_{i_o}(k) = y(Mk + i_o + \Delta) = \mathbf{w}^T\mathbf{r}(Mk + i_o + \Delta)$$
$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu\,y_{i_o}(k)\,[\mathbf{r}^*(Mk + i_o + \Delta)$$
$$- \sum_{i \in S_f} \left(y_{i_o}^*(k)\,e_i(k)\right)\tilde{\mathbf{r}}_i^*(k)]$$

$$\qquad (43)$$

When $S_f = \{\nu\}$ and $P = L = 1$, we obtain an algorithm that is similar to the MERRY algorithm without the renormalization. For comparison, the MERRY algorithm of [16] is

Given Δ, for symbol $k = 0, 1, 2, \ldots$,
$$\tilde{\mathbf{r}}(k) = \mathbf{r}(Mk + \nu + \Delta) - \mathbf{r}(Mk + \nu + N + \Delta),$$
$$e(k) = \mathbf{w}^T(k)\,\tilde{\mathbf{r}}(k),$$
$$\hat{\mathbf{w}}(k+1) = \mathbf{w}(k) + \mu\,e(k)\,\tilde{\mathbf{r}}^*(k), \qquad (44)$$
$$\mathbf{w}(k+1) = \frac{\hat{\mathbf{w}}(k+1)}{\sqrt{\hat{\mathbf{w}}^T(k+1)\,\hat{\mathbf{w}}(k+1)}}$$

The algorithm of (43) requires approximately $(2\,|S_f|+2)P\tilde{L}_w$ multiplications, where $|S_f|$, the size of the set S_f, is usually 1. The algorithm of (44), which uses $|S_f| = 1$, requires $4P\tilde{L}_w$ multiplications, a square root, and a division. (The $P\tilde{L}_w$ divisions are implemented as one division and $P\tilde{L}_w$ multiplications.)

The FRODO update rule of (43) uses rank 1 approximations of $\mathbf{A}$ and $\mathbf{C}$. One blind method of initializing the FRODO TEQ is to accumulate better estimates of $\mathbf{A}$ and $\mathbf{C}$ from the data by replacing the expectations in (27) and (30) by time averages, and then find the generalized eigenvector corresponding to the maximum generalized eigenvalue of $(\hat{\mathbf{C}}, \hat{\mathbf{A}})$.

IV. SYMBOL SYNCHRONIZATION

The FRODO algorithm requires a choice of the delay Δ. This section shows how the core idea in MERRY/FRODO can be used to obtain a reasonable heuristic delay choice. The performance metric used in this section is the shortening SNR [8], which is the ratio of energy in the desired window of the channel impulse response to the energy outside of this window.

The heuristic is: given the delay Δ_{peak} which maximizes the energy of the average (unshortened) channel in a window of taps Δ_{peak} through $\Delta_{peak} + \nu - 1$, a near-optimum delay is

$$\Delta = \Delta_{peak} + \left\lfloor \frac{L_w}{2} \right\rfloor. \qquad (45)$$

There are two issues to be addressed: (1) the means of obtaining Δ_{peak}, and (2) the validity of this heuristic. These will be addressed in order.

In the absence of a TEQ, $\mathbf{c}_{p,l} = \mathbf{h}_{p,l}$. From Theorem 1, if we only make one comparison using $i = \nu$,

$$J_{frodo}(\Delta) = 2\sum_{l=1}^{L}\sigma_{x,l}^2\|\mathbf{h}_{l,wall}^{\nu+\Delta}\|^2 + 2\sum_{p=1}^{P}\sigma_{n,p}^2, \qquad (46)$$

with an analogous definition for $\mathbf{h}_{l,wall}^{\nu+\Delta}$ as for $\mathbf{c}_{l,wall}^{\nu+\Delta}$. Since

$$\|\mathbf{h}_l\|^2 = \|\mathbf{h}_{l,win}^{\nu+\Delta}\|^2 + \|\mathbf{h}_{l,wall}^{\nu+\Delta}\|^2, \qquad (47)$$

the index Δ_{peak} in which the average windowed channel energy is highest is the index in which the average walled channel energy is smallest. Thus, Δ_{peak} can be estimated by minimizing an estimate of $J_{frodo}(\Delta)$ over Δ, as measured on the received data rather than the TEQ output:

$$\hat{\Delta}_{peak} = \arg\min_{0 \leq \Delta \leq M-1}\sum_{k=1}^{K}|r(Mk + \nu + \Delta)$$
$$-r(Mk + \nu + N + \Delta)|^2 \qquad (48)$$
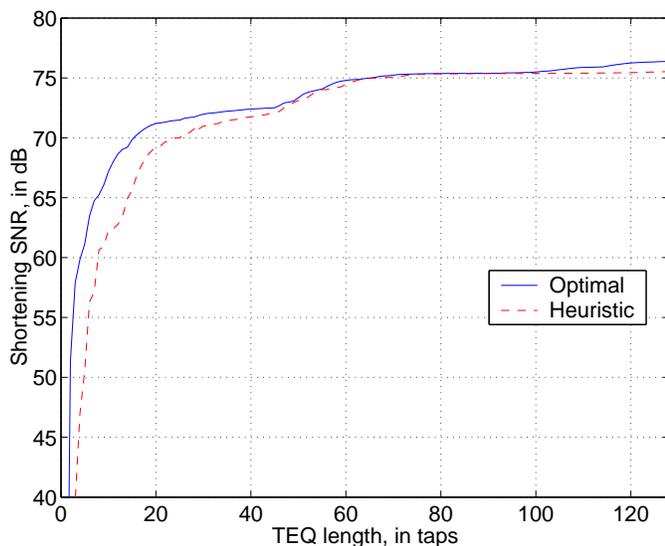
Fig. 4. Shortening SNR versus TEQ length for FRODO using the optimal and a heuristic delay.

Fig. 5. Example 1: the plots show the shortened channel impulse response magnitudes using (a) no TEQ, (b) an MMSE TEQ, (c) an MSSNR TEQ, and (d) a (zero-forcing) MSSNR TEQ with a window of size 1. Here, $\nu = 16$.

Fig. 6. Example 1: the plots show the shortened channel impulse response magnitudes using (a) an MDS TEQ with linear weights, (b) a MERRY TEQ, (c) a FRODO TEQ with two windows, and (d) a "full" FRODO TEQ with $\nu = 16$ windows.

for some number of symbols K. This approach only requires $M \cdot K$ multiplications, $M \cdot (2K - 1)$ additions, and $M - 1$ comparisons. Thus, a large value of K can be used for an accurate estimate with low computational complexity. Moreover, this heuristic can be applied to other design methods (besides FRODO) to avoid a global delay search.

The second question is whether or not this heuristic is valid. Fig. 4 shows a plot of the shortening SNR achieved by the delay-optimized FRODO design and by the FRODO design using the heuristic delay of (45). For simplicity, $P = L$, $|S_f| = 1$. The performance was averaged over carrier serving area (CSA) loops 1 through 8 [11] (standard synthetic ADSL test channels), and the window size was 32 taps. The heuristic delay provides reasonable performance relative to the optimal delay for TEQs with at least 8 taps, and very nearly optimal performance for TEQs with at least 32 taps. For TEQs shorter than 8 taps, the range of "good" delay choices will be small, so a heuristic approach may not be adequate. For ADSL, typical TEQ lengths are 16 or 32 taps. Other heuristics may be used; the proposed approach is merely one method which generally works and is blind.

V. SIMULATIONS

This section presents simulation results. The first example compares the exact (non-adaptive) FRODO design to existing TEQs, with the goal of characterizing FRODO's asymptotic performance. The second example explores the convergence behavior of FRODO. The third example demonstrates that FRODO can be used to jointly shorten multiple channels at once. The fourth example provides SISO and SIMO BER curves.

In all four examples, the FFT size is $N = 64$, the CP length is $\nu = 16$, and the channel model consists of three parts [37]: $\mathbf{h}_{local,1}$, scatterers near the transmitter; $\mathbf{h}_{mid}$, remote scatterers; and $\mathbf{h}_{local,2}$, scatterers near the receiver. The channel is then

$$\mathbf{h} = \mathbf{h}_{local,1} \star \mathbf{h}_{mid} \star \mathbf{h}_{local,2}, \qquad (49)$$
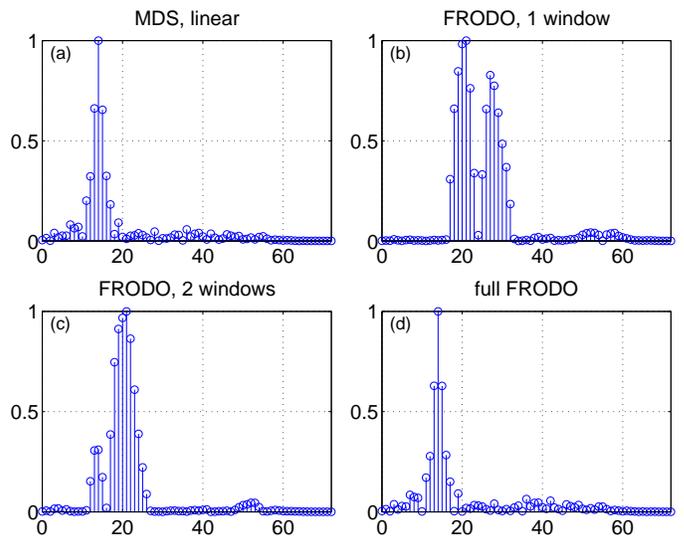
where $\star$ denotes convolution. $\mathbf{h}_{mid}$ consists of 32 uncorrelated Rayleigh fading taps with an exponential delay profile, and $\mathbf{h}_{local,1}$ and $\mathbf{h}_{local,2}$ each consist of 6 uncorrelated Rayleigh fading taps with a uniform delay profile.

Example 1: In this example, the TEQ has 32 taps and the SNR is 20 dB (AWGN). Figs. 5 and 6 show the shortened channel impulse responses magnitudes using various designs, for the channel realization shown in Fig. 5(a). The "full" FRODO impulse response is quite similar to the linear MDS impulse response (rather than the zero-forcing impulse response), FRODO with one window has characteristics like the MSSNR design, and FRODO with two windows has similar characteristics but is slightly narrower. Fig. 7 shows the shortening SNR [8], the inverse of the MSE [2], and the inverse of
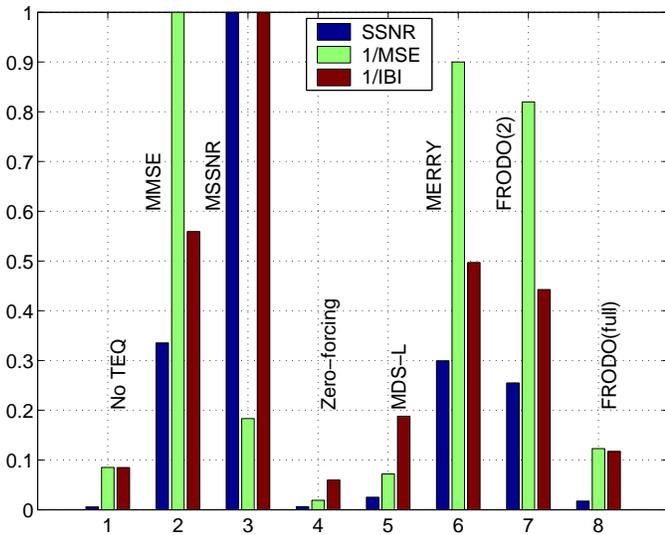
Fig. 7. Performance of various shortened channels for example 1. The shortening SNR [8], the inverse of the MSE [2], and the inverse of the inter-block interference [38] were averaged over 10000 channel realizations and normalized relative to the largest (i.e. best) value obtained from the 8 TEQ designs.

Fig. 8. Performance metrics vs. time: SSNR (top left), MSE (top right), IBI (bottom left), and MERRY cost (bottom right).

the inter-block interference (IBI) [38], averaged over 10000 channel realizations and normalized relative to the largest (i.e. best) value obtained from the 8 designs. The FRODO cost function with 1 window (i.e. MERRY) performs much like the MMSE design. The use of two windows for FRODO only slightly degrades the performance. Hence, we may use two windows in the adaptive version of FRODO without fear of significantly affecting the asymptotic performance. However, the use of all 16 windows causes the FRODO TEQ to achieve performance that is slightly worse than the (linear) MDS TEQ. Thus, the number of windows for FRODO should be relatively small, since we want to improve the convergence speed without adversely affecting the asymptotic solution.

Another insight gained from this example is that if the blind, non-adaptive FRODO initialization is used, then a performance comparable to the MMSE and MSSNR designs can be achieved without the need for training.

Example 2: We now examine the convergence rate of FRODO using various numbers of windows. Here, the TEQ has 16 taps and the SNR is 25 dB (AWGN). For a fair comparison, all algorithms used the same step size, normalized by the number of windows $|S_f|$. The synchronization was performed blindly using the method of Section IV.

The performance of FRODO versus time is shown in Fig. 8, in terms of the shortening SNR [8], the MSE [2], the IBI [19], and the MERRY cost [i.e. FRODO with only one term in (9)]. For this example, FRODO takes about 5000 iterations to converge. This corresponds to 300 iterations per tap, which is roughly consistent with the folklore that adaptive algorithms often take 100 iterations per tap to converge. By adding additional comparisons, the algorithm converges faster, but the quality of the final solution is not as good. Ideally, one would choose the parameters such that the final performance of the two algorithms were equal and then compare convergence rates, but that is not possible here since the use of more
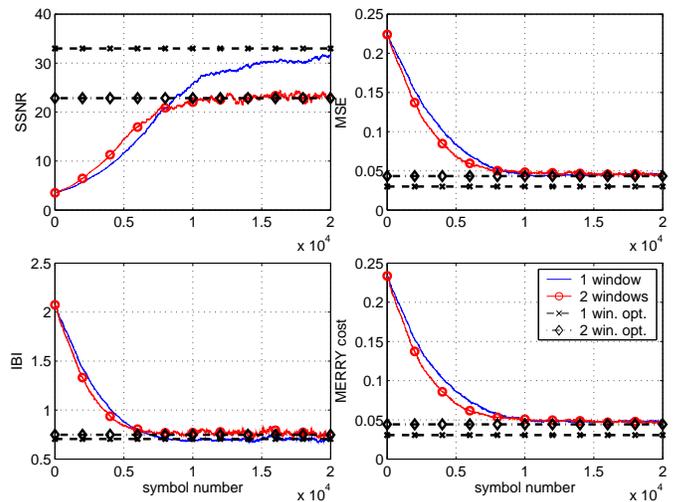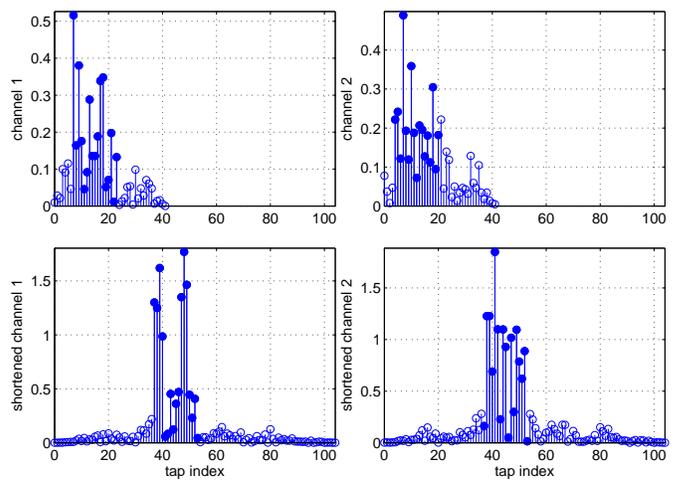
Fig. 9. Joint shortening of two Rayleigh fading channels. Top: channel impulse responses magnitudes; bottom: impulse responses magnitudes of the shortened channels. The "filled" stems in the channels indicate the window of $\nu + 1$ taps with largest energy (for the channel) or the window of $\nu + 1$ taps starting with the desired delay (for the shortened channel).

comparisons changes the asymptotic performance. The moral is that multiple comparisons should only be used to speed convergence or tracking, but near convergence, the algorithm should drop down to only one comparison.

Example 3: This MISO example demonstrates MERRY's ability to jointly shorten multiple channels, blindly and adaptively. There are $L = 2$ transmitters and $P = 1$ receivers. The TEQ has 64 taps, and the SNR is 20 dB (AWGN). The two input sequences and the noise sequence were independent. We assume that the transmitted sequences are coarsely synchronized, i.e. that the two cyclic prefixes arrive very roughly at the same time as each other, otherwise no joint channel shortening algorithm will succeed.

The implementation of MERRY in this case is no different than for a single channel. Fig. 9 shows the two channel impulse responses and the two effective channels shortened by MERRY after convergence. Fig. 10 shows the joint SSNR [8],
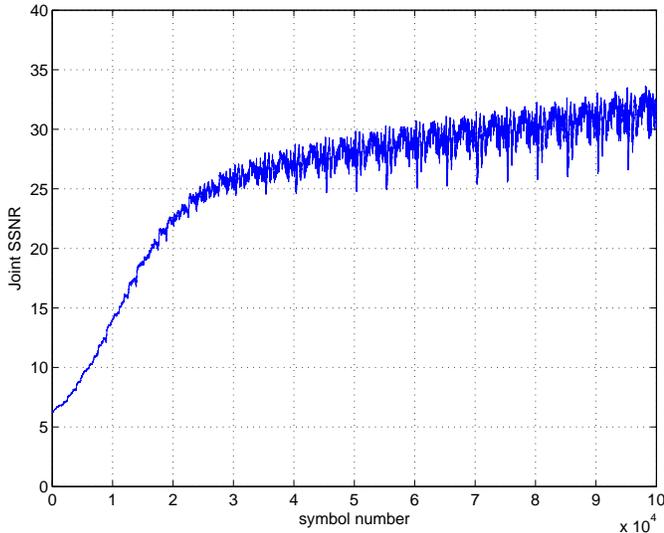
Fig. 10. The joint shortening SNR versus time as FRODO adapts to jointly shorten two Rayleigh fading channels.
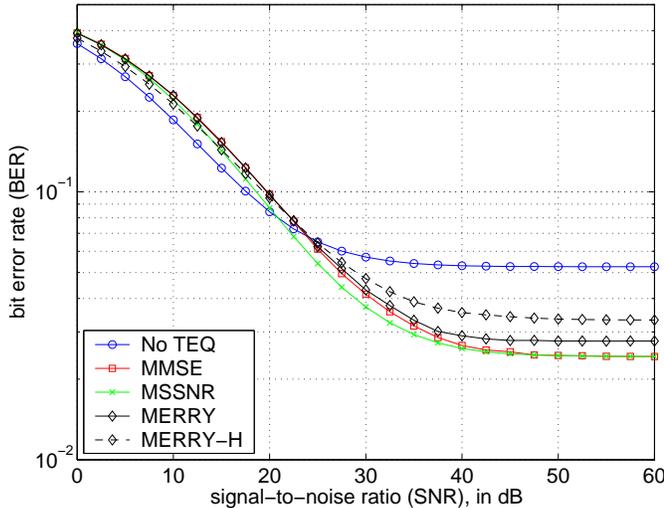
Fig. 11. BER vs. SNR for the SISO case.

Fig. 12. BER vs. SNR for the SIMO case.

similarly, and MERRY with a heuristic delay performs almost as well. The SISO curves level off for high SNR because the channel cannot be perfectly shortened. The SIMO BER values are much lower because the effective channel can be almost perfectly shortened.

VI. CONCLUSIONS

The MERRY algorithm has previously been shown to blindly, adaptively shorten a transmission channel in order to perform equalization in a multicarrier receiver. This paper has proposed extensions to the MERRY algorithm that remove the square root and division in the update, allow for the use of alternate (and possibly more appropriate) constraints, and allow fractionally-spaced and MIMO adaptive equalization. A low-complexity method was proposed for choosing the symbol synchronization, and a method was proposed for blind initialization of the algorithm to avoid slow modes of convergence. Each of the proposed methods was illustrated via simulation. The Matlab code to reproduce the figures in this paper is available at [39].

[28] versus time. The synchronization was performed blindly as in Section IV. The fact that the joint SSNR increased from 6 to 32 is evidence that MERRY can jointly shorten multiple channels, blindly and adaptively.

Example 4: Figs. 11 and 12 show BER curves for the SISO and SIMO cases, respectively, using the channel model of (49), with $L = 1$ and $P = 2$. The BER values were averaged over 200 channels (for SNRs of 30–60 dB, Fig. 12 used 2000 channels), with 100 data blocks per channel. The frequency domain signal was differentially encoded BPSK, so that no FEQ was needed. The blind, non-adaptive MERRY TEQ was compared to the non-adaptive MMSE [3], [29] and MSSNR [8] designs, all using delay optimization. The MIMO MSSNR design is simply the MMSE design with assumptions of white input and no noise. The performance of MERRY with the heuristic delay choice of (45) is denoted "MERRY-H." For low SNRs, all TEQs have very little effect on the BER. For larger SNRs, the three delay-optimized methods perform

REFERENCES

[1] G. D. Forney, "Maximum-likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," *IEEE Trans. on Info. Theory*, vol. 18, pp. 363–378, May 1972.
[2] D. D. Falconer and F. R. Magee, "Adaptive Channel Memory Truncation for Maximum Likelihood Sequence Estimation," *Bell Sys. Tech. Journal*, pp. 1541–1562, Nov. 1973.
[3] N. Al-Dhahir and J. M. Cioffi, "Efficiently Computed Reduced-Parameter Input-Aided MMSE Equalizers for ML Detection: A Unified Approach," *IEEE Trans. on Info. Theory*, vol. 42, no. 3, pp. 903–915, May 1996.
[4] A. N. Akansu, P. Duhamel, X. Lin, and M. de Courville, "Orthogonal Transmultiplexers in Communication: A Review," *IEEE Transactions on Signal processing*, vol. 46, no. 4, pp. 979–995, Apr. 1998.

[5] S. Barbarossa, G. Scutari, and A. Swami, "MUI-free CDMA Systems Incorporating Space-Time Coding and Channel Shortening," in *International Conference on Acoustics, Speech, and Signal Processing*, May 2002, vol. 3, pp. 2213–2216.

[6] T. Starr, J.M. Cioffi, and P.J. Silvermann, *Understanding Digital Subscriber Line Technology*, Prentice Hall PTR, Upper Saddle River, NJ, 1999.

[7] J. S. Chow, J. M. Cioffi, and J. A. C. Bingham, "Equalizer Training Algorithms for Multicarrier Modulation Systems," in *Proc. IEEE Int. Conf. on Comm.*, Geneva, Switzerland, May 1993, pp. 761–765.

[8] P. J. W. Melsa, R. C. Younce, and C. E. Rohrs, "Impulse Response Shortening for Discrete Multitone Transceivers," *IEEE Trans. on Comm.*, vol. 44, pp. 1662–1672, Dec. 1996.

[9] N. Al-Dhahir and J. M. Cioffi, "Optimum Finite-Length Equalization for Multicarrier Transceivers," *IEEE Trans. on Comm.*, vol. 44, no. 1, pp. 56–64, Jan. 1996.

[10] M. Nafie and A. Gatherer, "Time-Domain Equalizer Training for ADSL," in *Proc. IEEE Int. Conf. on Comm.*, Montreal, Canada, June 1997, vol. 2, pp. 1085–1089.

[11] G. Arslan, B. L. Evans, and S. Kiaei, "Equalization for Discrete Multitone Receivers To Maximize Bit Rate," *IEEE Trans. on Signal Processing*, vol. 49, no. 12, pp. 3123–3135, Dec. 2001.

[12] B. Farhang-Boroujeny and M. Ding, "Design Methods for Time-Domain Equalizers in DMT Transceivers," *IEEE Trans. on Comm.*, vol. 49, no. 3, pp. 554–562, Mar. 2001.

[13] N. Lashkarian and S. Kiaei, "Optimum Equalization of Multicarrier Systems: A Unified Geometric Approach," *IEEE Trans. on Comm.*, vol. 49, pp. 1762–1769, Oct. 2001.

[14] R. Schur and J. Speidel, "An Efficient Equalization Method to Minimize Delay Spread in OFDM/DMT Systems," in *Proc. IEEE Int. Conf. on Comm.*, Helsinki, Finland, June 2001, vol. 5, pp. 1481–1485.

[15] A. Tkacenko and P. P. Vaidyanathan, "Noise Optimized Eigenfilter Design of Time-domain Equalizers for DMT Systems," in *Proc. IEEE Int. Conf. on Comm.*, New York, NY, Apr.–May 2002, vol. 1.

[16] R. K. Martin, J. Balakrishnan, W. A. Sethares, and C. R. Johnson, Jr., "A Blind, Adaptive TEQ for Multicarrier Systems," *IEEE Signal Processing Letters*, vol. 9, no. 11, pp. 341–343, Nov. 2002.

[17] M. Milosevic, L. F. C. Pessoa, B. L. Evans, and R. Baldick, "DMT Bit Rate Maximization with Optimal Time Domain Equalizer Filter Bank Architecture," in *Proc. IEEE Asilomar Conf.*, Pacific Grove, CA, Nov. 2002, vol. 1, pp. 377–382.

[18] K. Vanbleu, G. Ysebaert, G. Cuypers, M. Moonen, and K. Van Acker, "Bitrate Maximizing Time-Domain Equalizer Design for DMT-based Systems," to appear in *IEEE Trans. on Comm.*, 2004.

[19] S. Celebi, "Interblock Interference (IBI) Minimizing Time-Domain Equalizer (TEQ) for OFDM," *IEEE Signal Processing Letters*, vol. 10, no. 8, pp. 232–234, Aug. 2003.

[20] D. Daly, C. Heneghan, and A. D. Fagan, "Minimum Mean-Squared Error Impulse Response Shortening for Discrete Multitone Transceivers," *IEEE Trans. Signal Processing*, vol. 52, no. 1, pp. 301–306, 2004.

[21] R. K. Martin, J. Balakrishnan, W. A. Sethares, and C. R. Johnson, Jr., "Blind, Adaptive Channel Shortening for Multicarrier Systems," in *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2002.

[22] J. Balakrishnan, R. K. Martin, and C. R. Johnson, Jr., "Blind, Adaptive Channel Shortening by Sum-squared Auto-correlation Minimization (SAM)," *IEEE Trans. on Signal Processing*, vol. 51, no. 12, pp. 3086–3093, Dec. 2003.

[23] T. Miyajima and Z. Ding, "Multicarrier Channel Shortening Based on Second-Order Output Statistics," in *Proc. IEEE Workshop on Signal Proc. Advances in Wireless Comm.*, Rome, Italy, June 2003.

[24] T. Miyajima and Z. Ding, "Second-Order Statistical Approaches to Channel Shortening in Multicarrier Systems," to appear in *IEEE Trans. on Signal Processing*, 2004.

[25] 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. IEEE Int. Conf. on Comm.*, Dallas, TX, June 1996, pp. 1318–1321.

[26] F. Romano and S. Barbarossa, "Non-Data Aided Adaptive Channel Shortening for Efficient Multi-Carrier Systems," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc.*, Hong Kong SAR, China, Apr. 2003, vol. 4, pp. 233–236.

[27] D. N. Kalofonos, M. Stojanovic, and J. G. Proakis, "Performance of Adaptive MC-CDMA Detectors in Rapidly Fading Rayleigh Channels," *IEEE Trans. on Wireless Comm.*, vol. 2, no. 2, pp. 229–239, Mar. 2003.

[28] M. Milosevic, L. F. C. Pessoa, and B. L. Evans, "Simultaneous Multichannel Time Domain Equalizer Design Based On The Maximum Composite Shortening SNR," in *Proc. IEEE Asilomar Conf.*, Pacific Grove, CA, Nov. 2002, vol. 2, pp. 1895–1899.

[29] N. Al-Dhahir, "FIR Channel-Shortening Equalizers for MIMO ISI Channels," *IEEE Trans. on Comm.*, vol. 49, no. 2, pp. 213–218, Feb. 2001.

[30] W. Younis and N. Al-Dhahir, "Joint Prefiltering and MLSE Equalization of Space-Time Coded Transmissions Over Frequency-Selective Channels," *IEEE Trans. on Vehicular Tech.*, vol. 51, pp. 144–154, Jan. 2002.

[31] A. Tkacenko and P. P. Vaidyanathan, "Eigenfilter Design of MIMO Equalizers for Channel Shortening," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Orlando, FL, May 2002, vol. 3, pp. 2361–2364.

[32] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir, "Intercarrier Interference in MIMO OFDM," *IEEE Trans. on Signal Processing*, vol. 50, no. 10, pp. 2451–2464, Oct. 2002.

[33] G. Leus and M. Moonen, "Per-Tone Equalization for MIMO OFDM Systems," *IEEE Trans. on Signal Processing, Special Issue on Signal Processing for MIMO Wireless Comm. Systems*, vol. 51, no. 11, pp. 2965–2975, Nov. 2003.

[34] D. L. Jones, "Property-Restoral Algorithms for Blind Equalization of OFDM," in *Proc. IEEE Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2003, pp. 619–622.

[35] C. Chatterjee, V. P. Roychowdhury, J. Ramos, and M. D. Zoltowski, "Self-Organizing Algorithms for Generalized Eigen-Decomposition," *IEEE Trans. on Neural Networks*, vol. 8, no. 6, pp. 1518–1530, Nov. 1997.

[36] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1996.

[37] A. J. Paulraj and B. C. Ng, "Space-time Modems for Wireless Personal Communications," *IEEE Personal Communications*, vol. 5, pp. 36–48, Feb. 1998.

[38] S. Celebi, "Interblock Interference (IBI) and Time of Reference (TOR) Computation in OFDM Systems," *IEEE Trans. on Comm.*, vol. 49, no. 11, pp. 1895–1900, Nov. 2001.

[39] R. K. Martin, "Matlab code for papers by R. K. Martin," [Online]. Available: http://bard.ece.cornell.edu/matlab/martin/index.html.

Richard K. Martin obtained dual B.S. degrees in physics and electrical engineering from the University of Maryland, College Park in 1999 (Summa Cum Laude), and an M.S. and a Ph.D. in electrical engineering from Cornell University in 2001 and 2004, respectively. In the fall of 2004, he plans to join the faculty of the Air Force Institute of Technology in Dayton, Ohio, as an Assistant Professor. His research interests include equalization for multicarrier systems; blind, adaptive algorithms; reduced complexity equalizer design; and exploiting sparsity for performance improvement of adaptive filters. He has published six journal papers, fifteen conference papers, and the book *Theory and Design of Adaptive Filters Answer Book*; and he has three patents pending.

John M. Walsh was born in Carbondale, IL on September 23, 1981, and spent his youth in Illinois and Virginia. He received the B.S. degree magna cum laude and the M.S. degree from Cornell University in 2002 and 2004, respectively. Since then he has continued his studies at Cornell University in pursuit of a Ph.D. He is a member of Tau Beta Pi and Eta Kappa Nu. His research interests include developing a behavior theory for connected adaptive elements, turbo joint synchronization/equalization and decoding, and multi-input multi-output channel shortening; and he loves to play the guitar.

$\mathbf{h}_{P,L}$
$\mathbf{w}_1$
$\mathbf{w}_P$
$n_1(k)$
$n_P(k)$ 12
$r_1(k)$
$r_P(k)$
$y_1(k)$
$y_P(k)$
$y(k)$
$y(k)$
$x(k)$
$n(k)$
$r(k)$
$\mathbf{h}$
$\mathbf{w}$
$\mathbf{c} = \mathbf{h} \star \mathbf{w}$

C. Richard Johnson, Jr. was born in Macon, GA in 1950. He received a B.E.E. with high honors from the Georgia Institute of Technology (1973); and an M.S.E.E. and a Ph.D. in E.E. with minors in Engineering-Economic Systems and Art History from Stanford University (1975, 1977). He is currently a Professor in the School of Electrical and Computer Engineering at Cornell University.

Dr. Johnson has received school, college, and national teaching awards, including selection by Eta Kappa Nu as the Outstanding Young Electrical Engineer in 1982 and as the C. Holmes MacDonald Outstanding Teacher among young professors of electrical engineering in the USA in 1983. Dr. Johnson was elected a Fellow of the IEEE in 1989 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. In 1999, Dr. Johnson became a member of the Acme Oyster House Raw Oyster Eating (15 Dozen) Club, New Orleans, LS. Dr. Johnson's current research interests are in adaptive parameter estimation theory that is useful in applications of digital signal processing to communications systems.

The current broadband adaptive receiver design project of his research group at Cornell is described at http://bard.ece.cornell.edu/ . The group's activity is currently supported by the National Science Foundation, Applied Signal Technology, and Texas Instruments.