Low Complexity MIMO Blind Adaptive Channel Shortening

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

Cornell University
School of Elec. & Comp. Engineering
Ithaca, NY 14853
{frodo,johnson}@ece.cornell.edu, jmw56@cornell.edu

ABSTRACT

Channel shortening is often necessary for mitigation of intersymbol and inter-carrier interference in multicarrier transceivers. 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. In this paper, MERRY is modified to remove the square root and division needed at each iteration, 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 case; and blind symbol synchronization and initialization techniques are proposed. Simulations demonstrate the success of the algorithm and the synchronization technique in a MIMO setting.

1. INTRODUCTION

Channel shortening is a generalization of equalization, since equalization amounts to shortening the channel to length one. Channel shortening to a length greater than one is frequenctly used to facilitate equalization in systems employing multicarrier modulation (MCM). MCM techniques like discrete multitone (DMT) and orthogonal frequency division multiplexing (OFDM) have been deployed in applications such as IEEE 802.11a and HIPERLAN/2 wireless LANs, digital audio/video broadcast, digital subscriber loops (DSL), power line communications, and satellite radio.

In a system employing MCM, the linear convolution between the channel and data is made to appear circular by adding a cyclic prefix (CP) to each data block. If the channel is shorter than the CP, then the convolution appears circular and the effect of the channel in the frequency domain is a pointwise multiplication, which is easily equalized. However, the channel is often longer than the CP, leading to inter-carrier and intersymbol interference (ICI, ISI). In order to mitigate this ICI/ISI, a time-domain equalizer (TEQ), i.e. a channel shortener, can be employed at the receiver to shorten the effective channel to the desired length [1] – [8]. Most TEQs in the literature have been designed in the context of DSL, which runs over twisted pair telephone lines. Consequently, most TEQ designs are trained and non-adaptive.

Recently, blind, adaptive TEQ design has received some attention. The MERRY (Multicarrier Equalization by Restoration of RedundancY) algorithm [7] induces channel shortening by restoring the redundancy in the received data that is due to the addition of the CP. The algorithm is low-complexity and converges to the minimum mean squared error (MMSE) solution [1] for a white input. The SAM (Sum-squared Autocorrelation Minimization) algorithm [9] attempts to shorten the auto-correlation of the TEQ output sequence, since a short channel leads to a short auto-correlation. Although SAM converges quickly, it is multimodal and computationally intensive. The “Carrier Nulling Algorithm” (CNA) [10] exploits the fact that many MCM systems transmit zeros on some tones (usually at the band edges). The TEQ can be adapted blindly to force the corresponding output tones to zero. 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 [10].

Often, a multiple input, multiple output (MIMO) system model is of interest, in which multiple channels need to be shortened simultaneously. In a multicarrier code division multiple access (MC-CDMA) system, each user spreads its signal before multicarrier modulation takes place. To enhance performance, the receiver can jointly shorten all of the user’s channels to mitigate ISI before de-spreading. In DSL, joint channel shortening can be combined with multiuser detection to mitigate crosstalk. If a DSL system is operating in echo canceling mode, then the channel and the echo path must be jointly shortened [3], [11]. As another example, multiple receive antennas or oversampling of the received data may be employed. Joint channel shortening has been studied in, e.g., [3], [6], [11], and [12]. However, these works involved extending the training-based, non-adaptive designs of [1] – [3] to the MIMO case; whereas this paper considers the blind, adaptive case.

The contributions of this paper are a modification to the MERRY algorithm which removes the square root and division at each iteration and allows for constraints that may be more appropriate than the unit norm equalizer constraint of [7]; a method to increase the amount of usable data (MERRY only uses one sample per block); a generalization to the MIMO case; and methods for blind initialization and symbol synchronization (issues not addressed in [7]).

2. SYSTEM MODEL AND NOTATION

A single-input, single output (SISO) MCM system model is shown in Fig. 1. Each block of N samples is passed through an inverse fast Fourier transform (IFFT) and a dispersive chan-

The authors are supported in part by NSF Grant CCR-0310023, Applied Signal Technology, Texas Instruments, and the Olin Fellowship from Cornell University.

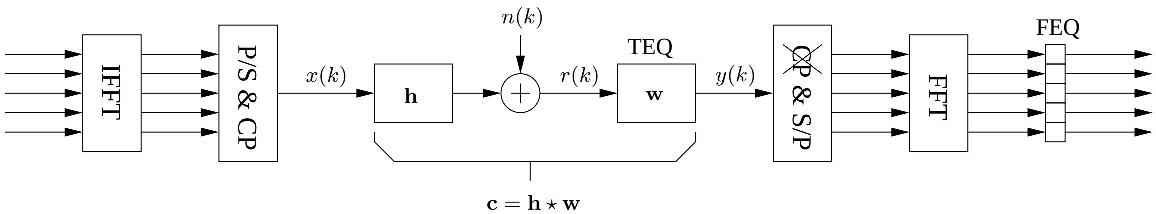

Fig. 1. SISO multicarrier system model. (I)FFT: (inverse) fast Fourier transform, P/S: parallel to serial, S/P: serial to parallel, CP/xCP: add/remove cyclic prefix.

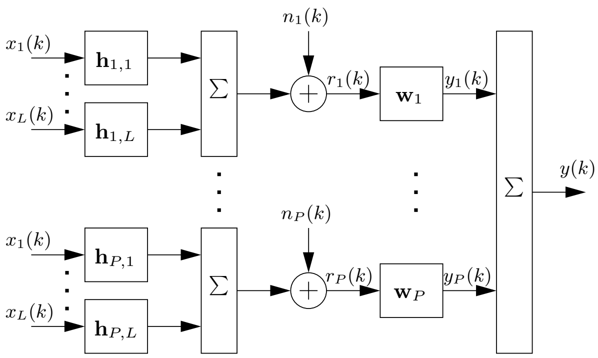

Fig. 2. MIMO channel shortening model.

nel $\mathbf{h}$ with additive noise. The channel convolution is made to appear circular by adding a CP, i.e. the last ν samples of each block are copied and inserted before the beginning of the block. If the effective channel length is less than $\nu + 1$, then the data can be demodulated by an FFT and a bank of complex gains, called a frequency domain equalizer (FEQ). The goal of the TEQ is to ensure this channel length condition.

Fig. 2 shows the model for MIMO channel shortening. The received signal $r_p(k)$ from antenna $p \in \{1, \cdots, P\}$, is obtained by passing each signal from user $l \in \{1, \cdots, L\}$ through channel $h_{p,l}$, adding the L outputs, and adding noise sequence $n_p(k)$.

After the CP is added, the last ν samples are identical to the first ν samples in each transmitted symbol, i.e.

$$x_l\left(Mk+i\right) = x_l\left(Mk+i+N\right), \quad i \in \{1, \ldots, \nu\}, \quad l \in \{1, \ldots, L\} \tag{1}$$

where $M = N + \nu$ is the total symbol duration and k is the symbol index. The received data r_p is given by

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

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

$$y_p(Mk+i) = \sum_{j=0}^{L_w} w_p(j)\, r_p(Mk+i-j). \tag{2}$$

Then the final, recombined output is obtained by

$$y(Mk+i) = \sum_{p=1}^{P} y_p(Mk+i). \tag{3}$$

The weights for the linear combination in (3) have implicitly been absorbed into the P TEQs. Each channel is modeled as a length $L_h + 1$ filter, each TEQ is a length $L_w + 1$ filter, and each effective channel $\mathbf{c}_{p,l} = \mathbf{h}_{p,l} \star \mathbf{w}_p$ has length $L_c + 1$, where $L_c = L_h + L_w$, and $\star$ denotes linear convolution.

3. EXTENSIONS TO MERRY

This section generalizes the MERRY cost function, and proposes a division-free update rule, an initialization method, and a symbol synchronization technique, all of which are blind.

3.1. Cost function and algorithm

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. Since there are ν samples in the CP, a natural generalization is to compare more than one of these samples to their counterparts at the end of the symbol:

$$J = \sum_{i \in S_f} \mathrm{E}\left[\left|y(Mk+i+\Delta) - y(Mk+i+N+\Delta)\right|^2\right], \quad \Delta \in \{0, \ldots, M-1\}, \tag{4}$$

where Δ is the delay and $S_f \subseteq \{1, \cdots, \nu\}$ is an index set. For MERRY, $S_f = \{\nu\}$. Different sets allow for the use of more or less data. Since this cost function allows the option of using all of the data in the CP, a single sample, or anything in between, we use the name Forced Redundancy with Optional Data Omission (FRODO) to refer to a stochastic gradient descent of this cost function. An equalization (not channel shortening) algorithm similar to using FRODO with the set $S_f = \{1, \cdots, \nu\}$ was proposed in [13]. The FRODO cost function includes [7] and [13] as special cases.

Theorem 1 *The FRODO cost function (4) simplifies to*

$$J = 2 \sum_{i \in S_f} \sum_{l=1}^{L} \sigma_{x,l}^2 \|\mathbf{c}_{l,wall}^{i+\Delta}\|^2 + 2\,|S_f| \sum_{p=1}^{P} \mathbf{w}_p^H \mathbf{R}_{n,p} \mathbf{w}_p, \tag{5}$$

where

$$\|\mathbf{c}_{l,wall}^{i+\Delta}\|^2 = \sum_{j=0}^{\Delta+i-\nu-1} |c_l(j)|^2 + \sum_{j=\Delta+i}^{L_c} |c_l(j)|^2, \quad \forall l, \tag{6}$$

and where

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

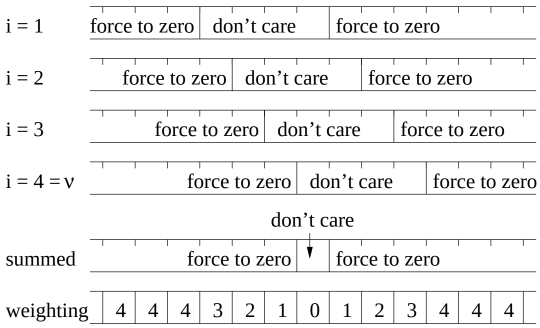

Fig. 3. The relation of the windows in the different terms of the FRODO cost function.

Remarks: The proof is straightforward but consists of tedious algebraic manipulation, and is omitted. For $L = P = 1$ and $S_f = \{\nu\}$, we have the term $i = \nu$ only. The cost function is the energy of the channel outside of a ν-length window. For $|S_f| = L = 1$ and $P > 1$, we suppress the tails of the averaged channel, $\mathbf{c} = \sum_p \mathbf{c}_p$. For $L > 1$ and $|S_f| = P = 1$, we suppress the average of the tail energies of the L channels, effectively shortening all L channels at once. The demodulated signal will be the sum of the L transmitted signals, but will be free of ISI and ICI, if this simultaneous channel shortening is successful. Thus, in an MC-CDMA scenario, the L signals can now be separated using the spreading codes.

The effect of using $|S_f| > 1$ can be illustrated as follows. Consider using the "full" index set, $S_f = \{1, \cdots, \nu\}$, as shown in Fig. 3. Each index term contributes to the cost function, and the windows for adjacent indices overlap by all but one sample. Thus, taps farther from the center are more heavily suppressed. This makes the cost function similar to the minimum delay spread algorithm [5]. The larger the index set, the more data gets used in the update (implying faster convergence), but the smaller the final window size.

Defining "stacked" amalgamations of various vectors as

$$\begin{aligned} \mathbf{r}_p(j) &= \left[r_p(j), r_p(j-1), \cdots, r_p(j-L_w)\right]^T, \\ \mathbf{r}(j) &= \left[\mathbf{r}_1^T(j), \mathbf{r}_2^T(j), \cdots, \mathbf{r}_P^T(j)\right]^T, \\ \tilde{\mathbf{r}}_i(k) &= \mathbf{r}(Mk+i+\Delta) - \mathbf{r}(Mk+i+N+\Delta), \\ \mathbf{w} &= \left[\mathbf{w}_1^T, \mathbf{w}_2^T, \cdots, \mathbf{w}_P^T\right]^T, \end{aligned}$$

the FRODO cost function (4) can be rewritten as

$$J = \sum_{i \in S_f} \mathrm{E}\left[\left|\tilde{\mathbf{r}}_i^T(k)\mathbf{w}\right|^2\right] = \mathbf{w}^H \mathbf{A} \mathbf{w}. \qquad (8)$$

To avoid $\mathbf{w} = \mathbf{0}$, we minimize J relative to the output power,

$$J_2 = \mathrm{E}\left[\left|y(Mk+i_o+\Delta)\right|^2\right] = \mathbf{w}^H \mathbf{C} \mathbf{w}, \qquad (9)$$

for some i_o. Equivalently, we may maximize J_2/J.

A generalized eigendecomposition algorithm was proposed in [14]. The algorithm is a gradient ascent of $\mathbf{w}^H\mathbf{C}\mathbf{w}$ with the constraint $\mathbf{w}^H\mathbf{A}\mathbf{w} = 1$. In the case of FRODO, we have blind, stochastic estimates of $\mathbf{A}$ and $\mathbf{C}$ available at the receiver. Combining these estimates with the method of [14], the FRODO algorithm is:

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

$$\begin{aligned} e_i(k) &= \mathbf{w}^T(k)\, \tilde{\mathbf{r}}_i(k), \quad \forall i \in S_f \\ y_{i_o}(k) &= y(Mk+i_o+\Delta) = \mathbf{w}^T \mathbf{r}(Mk+i_o+\Delta) \\ \mathbf{w}(k+1) &= \mathbf{w}(k) + \mu\, y_{i_o}(k) \\ &\times \left(\mathbf{r}^*(Mk+i_o+\Delta) - y_{i_o}^*(k) \sum_{i \in S_f} e_i(k) \tilde{\mathbf{r}}_i^*(k)\right) \end{aligned} \qquad (10)$$

When $|S_f| = P = L = 1$, we obtain an algorithm that is similar to MERRY, but without the renormalization.

3.2. Initialization and synchronization

If $\mathbf{A}$ and $\mathbf{C}$ are ill-conditioned, then FRODO will have slow modes of convergence. One way to avoid this is to accumulate estimates of $\mathbf{A}$ and $\mathbf{C}$ from the data,

$$\hat{\mathbf{A}} = \frac{1}{K}\sum_{k=1}^{K}\sum_{i \in S_f} \tilde{\mathbf{r}}_i^*(k)\, \tilde{\mathbf{r}}_i^T(k),$$

$$\hat{\mathbf{C}} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{r}^*(Mk+i_o+\Delta)\, \mathbf{r}^T(Mk+i_o+\Delta),$$

and then find the generalized eigenvector corresponding to the maximum generalized eigenvalue of $(\hat{\mathbf{C}}, \hat{\mathbf{A}})$.

MERRY (and hence FRODO) requires a choice of the delay Δ. This issue was not addressed in [7]. We propose the following heuristic: given the delay Δ_{peak} which maximizes the energy of the *channel* in a window of taps Δ_{peak} through $\Delta_{peak} + \nu - 1$, a near-optimum delay is

$$\Delta = \Delta_{peak} + \left\lfloor \frac{L_w}{2} \right\rfloor. \qquad (11)$$

Δ_{peak} can be obtained as follows. In the absence of a TEQ (i.e. $\mathbf{w} = 1$), $\mathbf{c}_{p,l} = \mathbf{h}_{p,l}$. From Theorem 1, if $S_f = \{\nu\}$,

$$J = 2\sum_{l=1}^{L} \sigma_{x,l}^2 \|\mathbf{h}_{l,wall}^{\nu+\Delta}\|^2 + 2\sum_{p=1}^{P} \sigma_{n,p}^2, \qquad (12)$$

Since the delay for which the average windowed channel energy is highest is the delay for which the average walled channel energy is smallest, Δ_{peak} can be estimated by minimizing an estimate of J over Δ, as measured on $r(k)$ rather than $y(k)$:

$$\hat{\Delta}_{peak} = \arg\min_{0 \le \Delta \le M-1} \sum_{k=1}^{K} |r(Mk+\nu+\Delta) - r(Mk+\nu+N+\Delta)|^2$$

This only requires MK multiplications and $M(2K-1)$ additions. This heuristic can also be applied to other design methods to avoid a global search over the delay parameter.

Fig. 4 shows a plot of the shortening SNR [3] achieved by FRODO using the optimal and heuristic delays. The performance was averaged over ADSL carrier serving area loops 1 through 8 [4]. The heuristic delay provides reasonable performance for TEQs with at least 8 taps, and nearly optimal performance for TEQs with at least 32 taps. For ADSL, typical TEQ lengths are 16 or 32 taps. Other heuristics may be used; the proposed approach is merely one blind method which works.

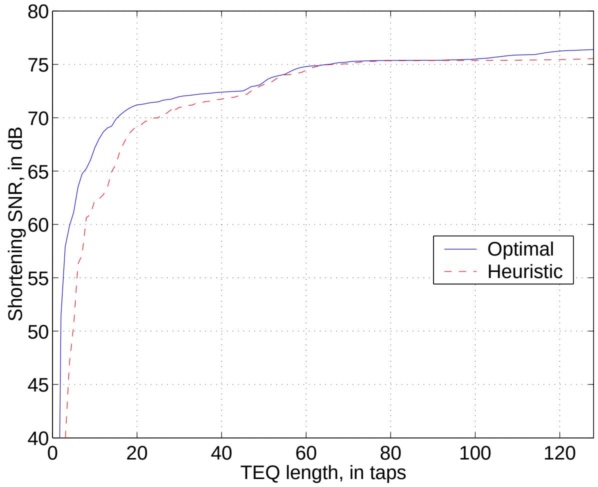

Fig. 4. Shortening SNR for FRODO using the optimal and heuristic delays.

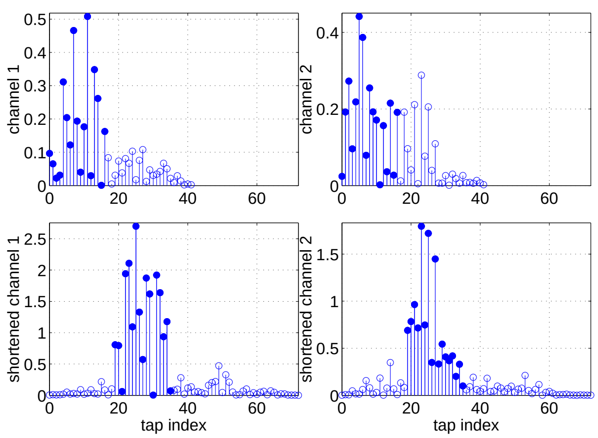

Fig. 5. Joint shortening simulation. The "filled" stems indicate the window of $\nu + 1$ taps with largest energy.

4. SIMULATIONS

Fig. 5 shows two channel impulse response magnitudes and the two effective channels as jointly shortened by FRODO after 20000 symbols. The parameters were $L = 2$, $P = 1$, $N = 64$, $\nu = 16$, 30 dB SNR, and the TEQ had 32 taps. We assume that the transmitted sequences are coarsely synchronized, i.e. that the two cyclic prefixes arrive very roughly at the same time, otherwise no joint channel shortening algorithm will succeed. The delay was chosen blindly using the method of Section 3.2. The tails of the impulse responses of both channels have been jointly suppressed by a single TEQ, without suppressing the desired windows of the channels.

ACKNOWLEDGEMENTS

The authors thank Dr. Jaiganesh Balakrishnan of Texas Instruments, Dallas for his comments, and Dr. Vince Poor of Princeton University for discussions of MIMO channel shortening.

5. REFERENCES

[1] D. D. Falconer and F. R. Magee, "Adaptive Channel Memory Truncation for Maximum Likelihood Sequence Estimation," *Bell Sys. Tech. J.*, pp. 1541–1562, Nov. 1973.

[2] 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.

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

[13] D. L. Jones, "Property-Restoral Algorithms for Blind Equalization of OFDM," *Proc. IEEE Asilomar Conf.*, Pacific Grove, CA, Nov. 2003.

[14] 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.