A Blind Adaptive TEQ for Multicarrier Systems

Richard K. Martin, Jaiganesh Balakrishnan, William A. Sethares, *Member, IEEE*, and C. Richard Johnson, Jr., *Fellow, IEEE*

***Abstract*—This letter exploits the cyclic prefix to create a blind adaptive globally convergent channel-shortening algorithm, with a complexity like least mean squares. The cost function is related to that of the shortening signal-to-noise solution of Melsa *et al.*, and simulations are provided to demonstrate the performance of the algorithm.**

***Index Terms*—Adaptive, blind, cyclic prefix, equalization, multicarrier, orthogonal frequency division multiplexing (OFDM), TEQ.**

I. Introduction

MULTICARRIER systems, such as orthogonal frequency division multiplexing or discrete multitone, have less stringent equalization requirements than single-carrier systems. If the channel is shorter than the cyclic prefix (CP), then the effect of the channel is flat fading on each carrier. If the channel exceeds this length, then interchannel interference (ICI) and intersymbol interference (ISI) will be present. The standard solution is to use a channel-shortening (time-domain) equalizer (TEQ).

There are currently many methods which, when given a channel, can compute the optimal equalizer (for some metric) [1]–[3]. There are also many suboptimal and/or adaptive approaches, such as [4]–[7]. Most approaches to TEQ design are nonadaptive, have high complexity, and require training or a channel estimate. While there are methods for blind channel identification for multicarrier systems, there is only one "blind" adaptive method that directly equalizes the channel [7]. However, [7] performs complete equalization rather than channel shortening. This not the desired criterion, so the overall performance is expected to be worse. Furthermore, [7] requires two matrix–vector multiplies per update, which is more computationally intensive than the proposed method. We propose a blind adaptive channel-shortening algorithm that has significantly lower complexity than [7].

Manuscript received February 27, 2002; revised July 25, 2002. This work was supported in part by the National Science Foundation under Grant ECS-9811297 and by NxtWave Communications. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ananthram Swami.

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

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.

W. A. Sethares is with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706 USA (e-mail: sethares@ece.wisc.edu).

Digital Object Identifier 10.1109/LSP.2002.804423

II. System Model

The (baseband) system model is as follows. Each of the N frequency bins is modulated with a quadrature amplitude-modulated signal, although often some bins are left as null carriers [7]. Modulation is performed via an inverse fast Fourier transform (FFT), and demodulation is accomplished via an FFT. Channel shortening is performed by a TEQ, and the resulting shortened combined channel is equalized by a bank of one-tap frequency-domain equalizers (FEQs). After the CP is added, the last ν samples are identical to the first ν samples in the symbol, i.e.,

$$x(M\cdot k+i)=x(M\cdot k+i+N), \qquad i\in\{1,2,\ldots,\nu\} \tag{1}$$

where $M=N+\nu$ is the total symbol duration, and k is the symbol index. To simplify the notation, henceforth we assume $k=0$ (without loss of generality). The received data $\mathbf{r}$ is obtained from $\mathbf{x}$ by

$$r(i)=\sum_{l=0}^{L-1} h(l)\cdot x(i-l)+n(i) \tag{2}$$

and the equalized data $\mathbf{y}$ is similarly obtained from $\mathbf{r}$ by

$$y(i)=\sum_{j=0}^{T-1} w(j)\cdot r(i-j) \tag{3}$$

where T is the length of the equalizer $\mathbf{w}$. The combined channel is denoted by $\mathbf{c}=\mathbf{h}\star\mathbf{w}$.

III. Merry Algorithm

The channel destroys the relationship in (1), because the ICI and ISI that affect the CP are different from the ICI and ISI that affect the last ν samples in the symbol. Consider a system with $N=8$, $\nu=2$, and $\mathbf{h}=[h(0),\ldots,h(4)]$. The CP contains $x(1), x(2)$, and the symbol contains $x(3),\ldots,x(10)$. Note that $x(2)=x(10)$, but at the receiver, the interfering samples before sample 2 are not all equal to their counterparts before sample 10. If $h(2)$, $h(3)$, and $h(4)$ were zero, then $r(2)=r(10)$.

If the channel order $L\leq\nu$, then the last sample in the CP should match the last sample in the symbol. One cost function that reflects this is

$$J_\delta=\mathrm{E}\left[|y(\nu+\delta)-y(\nu+N+\delta)|^2\right], \qquad \delta\in\{0,\ldots,M-1\} \tag{4}$$

where δ is the symbol synchronization parameter, which is included because this approach requires knowledge of where the symbol begins. The choice of δ will change the cost function.

The proposed algorithm "Multicarrier Equalization by Restoration of RedundancY" (MERRY) performs a stochastic gradient descent of (4), with a constraint to avoid the trivial solution $\mathbf{w} = \mathbf{0}$. This algorithm is as follows:

For symbol $k = 1, 2, \ldots$ and for tap $j = 0, 1, \ldots, T-1$

$$\begin{aligned} w_j(k+1) = & w_j(k) - \mu[y(M \cdot k + \nu + \delta) \\ & - y(M \cdot k + \nu + N + \delta)] \cdot [r^*(\nu + \delta - j) \\ & - r^*(\nu + N + \delta - j)] \end{aligned} \quad (5)$$

where $*$ denotes complex conjugation.

Possible implementations of the constraint include renormalizing $\mathbf{w}$ after each iteration or fixing one tap to unity. Note that the algorithm only updates once per symbol.

IV. Properties of the Solution

We now analyze the cost function and relate it to the shortening signal-to-noise (SSNR) solution of [3]. We assume the following.

1) Inverse FFT (IFFT) input is zero-mean, white, and wide-sense stationary (thus, the IFFT output bins are uncorrelated).
2) $N \geq T + L - 1$ (the FFT size is at least as large as the combined channel-equalizer length).
3) Noise autocorrelation function $R_n(\Delta) = 0$ for $\Delta > (N - T)$.
4) Noise is uncorrelated with the data.

A. Cost Function

The following theorem relates our work to that in [3]. It says that MERRY attempts to produce a "don't care" region with a width of ν taps.

Theorem 1: The cost function (4) simplifies to

$$J_\delta = 2\sigma_x^2 \left(\sum_{j=0}^{\delta-1} |c_j|^2 + \sum_{j=\nu+\delta}^{T+L-2} |c_j|^2 \right) + 2\mathbf{w}^t R_n \mathbf{w}^* \quad (6)$$

where $R_n = \mathrm{E}[\mathbf{n}_j \mathbf{n}_j^h]$, and $\mathbf{c} = \mathbf{h} \star \mathbf{w}$.

Proof: Consider the following definitions:

$$\begin{aligned} \mathbf{x}_j &= [x(j), x(j-1), \ldots, x(j - (T+L-1) + 1)]^t \\ \mathbf{x}'_j &= \mathbf{x}_j - \mathbf{x}_{j+N} \\ \mathbf{n}_j &= [n(j), n(j-1), \ldots, n(j-T+1)]^t \\ \mathbf{n}'_j &= \mathbf{n}_j - \mathbf{n}_{j+N}. \end{aligned} \quad (7)$$

Then (4) simplifies to

$$\begin{aligned} J_\delta &= \mathrm{E}\left[|\mathbf{c}^t \mathbf{x}_{\nu+\delta} - \mathbf{c}^t \mathbf{x}_{\nu+N+\delta} + \mathbf{w}^t \mathbf{n}_{\nu+\delta} - \mathbf{w}^t \mathbf{n}_{\nu+N+\delta}|^2 \right] \\ &= \mathrm{E}[|\mathbf{c}^t \mathbf{x}'_{\nu+\delta}|^2] + \mathrm{E}[|\mathbf{w}^t \mathbf{n}'_{\nu+\delta}|^2] \\ &= \mathbf{c}^t \underbrace{\mathrm{E}[\mathbf{x}'_{\nu+\delta}(\mathbf{x}'_{\nu+\delta})^h]}_{A_x^{\nu+\delta}} \mathbf{c}^* + \mathbf{w}^t \underbrace{\mathrm{E}[\mathbf{n}'_{\nu+\delta}(\mathbf{n}'_{\nu+\delta})^h]}_{A_n^{\nu+\delta}} \mathbf{w}^*. \end{aligned} \quad (8)$$

In going to the second line, we have assumed that the noise and the data are uncorrelated. Observe that

$$\begin{aligned} \mathbf{x}'_{\nu+\delta} = & [x(\nu+\delta) - x(\nu+\delta+N) \\ & x(\nu+\delta-1) - x(\nu+\delta+N-1), \ldots \\ & x(\nu+\delta-T-L+2) \\ & - x(\nu+\delta-T-L+2+N)]^t. \end{aligned} \quad (9)$$

The values of x that enter additively have a highest index of $(\nu+\delta)$, whereas the values of x that enter with a minus sign have a smallest index of $(\nu+\delta-T-L+2+N)$. If $N \geq T+L-1$, then the highest index in the first group will be lower than the lowest index in the second group. If this requirement and assumption 1) above hold, then $A_x^{\nu+\delta}$ will be diagonal. Furthermore, the middle ν terms in (9) are all zero, due to (1). Thus

$$A_x^{\nu+\delta} = 2\sigma_x^2[\mathrm{diag}(\mathbf{1}_{\delta\times 1}, \mathbf{0}_{\nu\times 1}, \mathbf{1}_{(T+L-1-\nu-\delta)\times 1})]. \quad (10)$$

The matrix $A_n^{\nu+\delta}$ simplifies as well:

$$\begin{aligned} A_n^{\nu+\delta} = & \mathrm{E}[\mathbf{n}_{\nu+\delta}\mathbf{n}_{\nu+\delta}^h] - \mathrm{E}[\mathbf{n}_{\nu+\delta}\mathbf{n}_{\nu+N+\delta}^h] \\ & - \mathrm{E}[\mathbf{n}_{\nu+N+\delta}\mathbf{n}_{\nu+\delta}^h] + \mathrm{E}[\mathbf{n}_{\nu+N+\delta}\mathbf{n}_{\nu+N+\delta}^h]. \end{aligned} \quad (11)$$

If assumption 3) holds, then the middle two terms are zero. The remaining two terms each equal the noise autocorrelation matrix, so $A_n^{\nu+\delta} = 2R_n$. Substituting into (8), we have

$$J_\delta = 2\sigma_x^2 \left(\sum_{j=0}^{\delta-1} |c_j|^2 + \sum_{j=\nu+\delta}^{T+L-2} |c_j|^2 \right) + 2\mathbf{w}^t R_n \mathbf{w}^*. \quad (12)$$

Thus, J_δ is proportional to the energy of the combined impulse response outside of a length ν window, plus a noise gain term. This completes the proof. ■

Theorem 1 suggests that MERRY finds a solution similar to the one found in [3]. MERRY minimizes the sum of the energy outside of a length ν window plus the energy of the filtered noise, subject to a constraint (e.g., $||\mathbf{w}|| = 1$). In contrast, [3] minimizes the energy of the combined impulse response outside of a window of length $\nu + 1$ (rather than ν), subject to the constraint that the energy inside the window is unity. Note that [3] does not limit the noise gain, whereas MERRY does.

B. Minima

This section outlines a proof of global convergence of a gradient descent of (4). (MERRY is a *stochastic* gradient descent of (4), so it should follow the average system and converge as well for small μ. See [8, Sect. 2.5] for a discussion of how a stochastic gradient algorithm performs the averaging.) Define

$$\begin{aligned} \mathbf{r}_j &= [r(j), r(j-1), \ldots, r(j-T+1)]^t \\ \mathbf{r}'_j &= \mathbf{r}_j - \mathbf{r}_{j+N}. \end{aligned} \quad (13)$$

Adding a Lagrangian constraint to the cost function, we have

$$\begin{aligned} J_\delta &= \mathrm{E}\left[|y(\nu+\delta) - y(\nu+\delta+N)|^2 \right] + \lambda(1 - \mathbf{w}^h\mathbf{w}) \\ &= \mathrm{E}\left[|\mathbf{w}^t \mathbf{r}_{\nu+\delta} - \mathbf{w}^t \mathbf{r}_{\nu+\delta+N}|^2 \right] + \lambda(1 - \mathbf{w}^h\mathbf{w}) \\ &= \mathbf{w}^t \underbrace{\mathrm{E}[\mathbf{r}'_{\nu+\delta}(\mathbf{r}'_{\nu+\delta})^h]}_{A_r^{\nu+\delta}} \mathbf{w}^* + \lambda(1 - \mathbf{w}^h\mathbf{w}). \end{aligned}$$

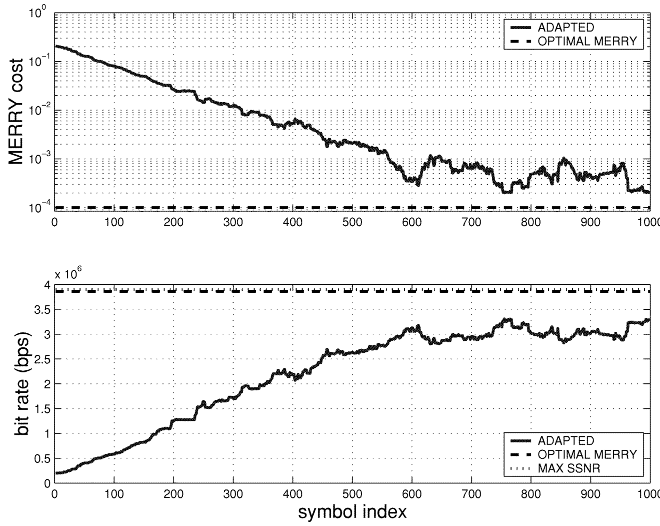

Fig. 1. Achievable bit rate versus time.

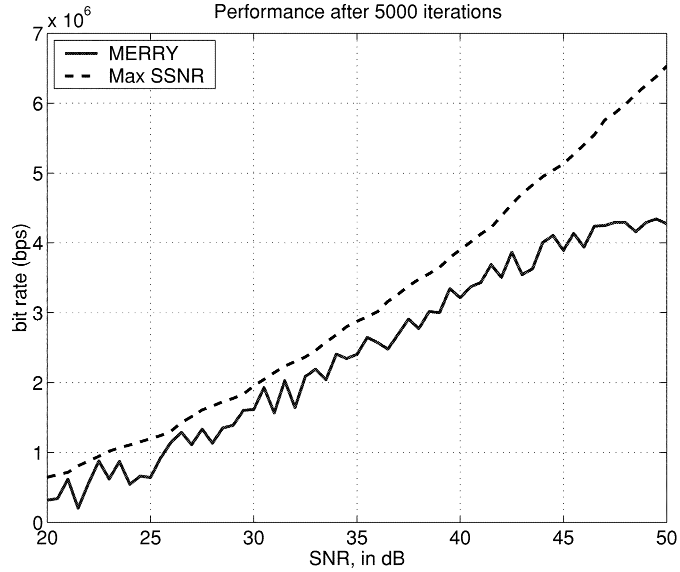

Fig. 2. Achievable bit rate versus SNR at input to receiver.

The gradient $\bigtriangledown_{\mathbf{w}} J_\delta = 0$ if and only if $(\lambda, \mathbf{w})$ are an eigenpair of $A_r^{\nu+\delta}$. Then, the Hessian $A_r^{\nu+\delta} - \lambda I$ is positive definite only if we choose λ to be the smallest eigenvalue. If the smallest eigenvalue is repeated, then there will be multiple minima, but all will have the same cost (equal to the eigenvalue). This proves global convergence of the gradient descent algorithm. Furthermore, in the noiseless case, $A_r^{\nu+\delta} = 2\sigma_x^2 A$, where $A = H_{\text{wall}}^h H_{\text{wall}}$, and H_{wall} is the channel convolution matrix with the middle ν rows removed, as in [3]. Using this, it can be shown that modifying the constraint from $||\mathbf{w}|| = 1$ to $||\mathbf{c}|| = 1$ makes the minima of (6) and [3] identical.

V. Simulations

Fig. 1 shows a digital subscriber loop (DSL) simulation using carrier serving area (CSA) loop 1, a standard DSL test channel [5]. The Matlab code is available at [9]. The CP length is 32; the TEQ had 16 taps; and $\sigma_x^2 ||\mathbf{h}||^2/\sigma_n^2 = 40$ dB. (Robustness to crosstalk will be considered in future work.) Initialization was a single spike. The DSL performance metric is the achievable bit rate for a fixed probability of error

$$B = \sum_i \ln\left(1 + \frac{\text{SNR}_i}{\Gamma}\right)$$

where SNR_i is the signal to interference and noise ratio in frequency bin i. (We assume a 6-dB margin and 4.2-dB coding gain; for more details, refer to [1].) Fig. 1 shows that MERRY can rapidly provide a solution approaching the maximum SSNR solution and the optimal MERRY solution. For DSL, MERRY should converge within 16 000 symbols in order to perform bit allocation at the end of the initialization period, whereas in a broadcast environment, tracking speed is more of an issue than converging within a set time. Fig. 2 shows the bit rate versus SNR. Here, the bit rate was computed by running for 5000 symbols and gradually decreasing the step size over time. For all these SNR values, MERRY approaches the max SSNR solution. The jaggedness is due to the random input.

VI. Conclusion

The MERRY algorithm performs blind adaptive channel shortening, which has not hitherto been attempted (although de Courville *et al.* [7] attempt full equalization for a multicarrier system without a cyclic prefix). The MERRY algorithm is low complexity and globally convergent. Future work will involve improving the convergence speed of the algorithm and expanding upon the analysis of the cost function.

Acknowledgment

The authors thank the Embedded Signal Processing Laboratory for providing the TEQ design toolbox [10].

References

[1] 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.

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

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

[5] 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.

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

[7] 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.

[8] O. Macchi, *Adaptive Processing: The Least Mean Squares Approach With Applications in Transmission*. Chichester, U.K.: Wiley, 1995.

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

[10] Embedded Signal Processing Laboratory, University of Texas at Austin. TEQ design toolbox. [Online]. Available: http://www.ece.utexas.edu/~bevans/projects/adsl/dmtteq/dmtteq.html.