Adaptive Minimum Symbol-Error Rate Equalization for Quadrature-Amplitude Modulation

Chen-Chu Yeh, *Member, IEEE,* and John R. Barry, *Member, IEEE*

***Abstract*—We propose the *adaptive minimum symbol-error rate* algorithm, which is a low-complexity technique for adapting the coefficients of a linear equalizer in systems using pulse-amplitude or quadrature-amplitude modulation. The proposed algorithm very nearly minimizes error probability in white Gaussian noise and can significantly outperform the minimum-mean-squared error equalizer (by as much as 16 dB) when the number of equalizer coefficients is small relative to the severity of the intersymbol interference.**

***Index Terms*—Adaptive equalization, MMSE criterion.**

I. Introduction

THE minimum-mean-squared error (MMSE) criterion is by far the most popular strategy for linear equalizer design, in part because of its amenability to adaptive implementation, and in part because it is widely believed to offer good performance with respect to more relevant performance measures such as error probability. Indeed, when the number of equalizer coefficients is sufficiently large, the MMSE equalizer closely approximates the linear equalizer whose coefficients are chosen to minimize symbol-error rate (SER). In contrast, however, when the number of equalizer coefficients is small relative to the severity of the intersymbol interference (ISI), the MMSE equalizer is far from optimal.

While the superiority of minimum-SER equalization has been known for decades [1]–[3], there were no known adaptive techniques for its implementation until the work of Chen *et al.* [4]. Since then, others have proposed techniques for realizing a minimum-SER equalizer [5]–[8], but all were restricted to binary modulation schemes. In this paper, we propose an adaptive algorithm for realizing the minimum-SER equalizer that is applicable to nonbinary pulse-amplitude modulation (PAM) and higher order complex quadrature amplitude modulation (QAM). First reported in [9], the proposed algorithm is a generalization of the binary scheme of [8]. Related work recently appeared in [10]. Simulation results show that the proposed algorithm very nearly minimizes SER, outperforming the MMSE equalizer by 16 dB in one example. Nevertheless, the complexity of the proposed algorithm is no greater than that of the least-mean square (LMS) algorithm.

Manuscript received January 29, 2002; revised April 10, 2003. This work was supported in part by the National Science Foundation under Grant CCR-0082329. This paper was presented in part at the IEEE International Conference on Communications, Atlanta, GA, June 7–11, 1998. The associate editor coordinating the review of this paper and approving it for publication was Prof. Zhi Ding.

C.-C. Yeh is with Intersil Corporation, Palm Bay, FL 32905 USA (e-mail: cyeh@intersil.com).

J. R. Barry is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA (e-mail: barry@ece.gatech.edu).

Digital Object Identifier 10.1109/TSP.2003.819081

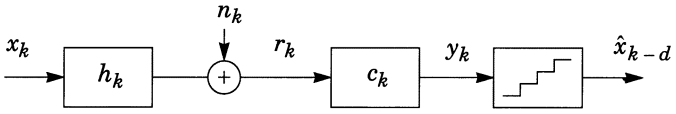


Fig. 1. System block diagram.

II. Problem Statement and Symbol-Error Probability

We first consider a real-valued linear discrete-time channel whose output at time k is given by

$$r_k = \sum_{i=0}^{M} h_i x_{k-i} + n_k \tag{1}$$

as depicted in Fig. 1, where the input symbols $\{x_k\}$ are drawn independently and uniformly from the alphabet $\mathcal{A} = \{\pm 1, \pm 3, \ldots \pm (L-1)\}$, where h_k is the causal channel impulse response with memory M and satisfying $h_k = 0$ for $k \notin \{0, \ldots M\}$, and where the noise samples $\{n_k\}$ are independent zero-mean Gaussian random variables with variance σ^2. The equalizer output at time k can be expressed as an inner product $y_k = \mathbf{c}^T \mathbf{r}_k$, where $\mathbf{c} = [c_0, \ldots, c_{N-1}]^T$ is a vector of N equalizer coefficients, and $\mathbf{r}_k = [r_k, \ldots, r_{k-N+1}]^T$ is a vector of channel outputs. This latter vector can be expressed as

$$\mathbf{r}_k = \mathbf{H}\mathbf{x}_k + \mathbf{n}_k \tag{2}$$

where $\mathbf{H}$ is an $N \times (M+N)$ Toeplitz matrix satisfying $H_{i,j} = h_{j-i}$, where $\mathbf{x}_k = [x_k, \ldots x_{k-M-N+1}]^T$ is a vector of transmitted symbols, and where $\mathbf{n}_k = [n_k, \ldots n_{k-N+1}]^T$ is a vector of noise samples.

We constrain the decision device to be memoryless so that the decision $\hat{x}_{k-d} \in \mathcal{A}$ regarding symbol x_{k-d} depends only on the kth equalizer output y_k. The delay parameter $d \in \{0, 1, \ldots M+N-1\}$ accounts for the delay of the channel and the equalizer, and it must be optimized for best performance.

When higher order (nonbinary) alphabets are used, the thresholds of the decision device must also be optimized. Let $f_k = \sum_{i=0}^{M} h_i c_{k-i}$ denote the *overall* impulse response of the cascade of the channel and equalizer. In vector form, we have

$$\mathbf{f} = [f_0, \ldots f_{M+N-1}] = \mathbf{c}^T \mathbf{H}.$$

The equalizer output at time k can then be expressed as

$$y_k = f_d x_{k-d} + \sum_{i \neq d} f_i x_{k-i} + v_k. \tag{3}$$

The first term $f_d x_{k-d}$ represents the desired signal, the second term represents residual ISI, and the third term represents the filtered noise, as defined by $v_k = \mathbf{c}^T \mathbf{n}_k$. Because the residual ISI plus noise has a symmetric probability distribution, the minimum-error probability decision thresholds are $\{0, \pm 2f_d, \ldots \pm (L-2)f_d\}$. Equivalently, we can apply the integer thresholds $\{0, \pm 2, \ldots \pm (L-2)\}$ to the *normalized* equalizer output $z_k = y_k/f_d$. We remark that although the cursor f_d is often close to unity and often ignored, it actually introduces a bias [11], and normalizing the equalizer output removes the bias.

Let $\mathcal{A}^{M+N}$ denote the set of all L^{M+N} possible $\mathbf{x}_k$ vectors, and let $\tilde{\mathbf{x}}$ denote a random vector uniformly distributed over the subset of L^{M+N-1} such vectors for which the *desired* symbol is 1 (i.e., $x_{k-d} = 1$). The problem addressed in this paper is that of finding a method for adapting $\mathbf{c}$ to minimize SER. The following result provides a concise characterization of SER in terms of $\tilde{\mathbf{x}}$.

Lemma 1: With optimal decision thresholds, the symbol-error probability $P_e = \Pr[\hat{x}_k \neq x_k]$ after any equalizer $\mathbf{c}$ may be expressed in an exact and compact form as

$$P_e = \frac{2L-2}{L} E\left[Q\left(\frac{\mathbf{c}^T \mathbf{H}\tilde{\mathbf{x}}}{\|\mathbf{c}\|\sigma}\right)\right] \tag{4}$$

where Q is the Gaussian error function [12], and where the expectation is with respect to $\tilde{\mathbf{x}}$.

Proof: Let $\epsilon_k = y_k - f_d x_{k-d} = \sum_{i \neq d} f_i x_{k-i} + v_k$ denote the equalizer error, containing both residual ISI and noise. Because v_k and $\{x_i\}$ are independent and symmetric random variables, the probability distribution of the equalizer error is also symmetric, satisfying $p(-\epsilon_k) = p(\epsilon_k)$. This symmetry implies that

$$\Pr[\epsilon_k < -f_d] = \Pr[\epsilon_k > f_d] = \frac{1}{2}\Pr[|\epsilon_k| > f_d]. \tag{5}$$

The key to the proof is the following derivation of a concise expression for the above probabilities:

$$\Pr[\epsilon_k < -f_d] = \Pr\left[\sum_{i \neq d} f_i x_{k-i} + v_k < -f_d\right] \tag{6}$$

$$= E\left[\Pr\left[-v_k > f_d + \sum_{i \neq d} f_i x_{k-i} \mid \{x_{k-i} : i \neq d\}\right]\right] \tag{7}$$

$$= E\left[Q\left(\frac{f_d + \sum_{i \neq d} f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right)\right] \tag{8}$$

$$= E\left[Q\left(\frac{\mathbf{c}^T \mathbf{H}\tilde{\mathbf{x}}}{\|\mathbf{c}\|\sigma}\right)\right]. \tag{9}$$

The expression in (7) results from conditioning (6) on a particular outcome of the set of interfering symbols $\{x_{k-i} : i \neq d\}$ and then averaging over all L^{M+N-1} equally likely outcomes. Now, the SER can be decomposed into the sum of two terms, according to whether one of the endpoints or one of the inner points of the L-ary alphabet $\{\pm 1, \pm 3, \ldots, \pm(L-1)\}$ is transmitted:

$$P_e = \frac{2}{L}\Pr[\hat{x}_{k-d} \neq x_{k-d} \mid x_{k-d} = \pm(L-1)] + \frac{L-2}{L}\Pr[\hat{x}_{k-d} \neq x_{k-d} \mid x_{k-d} \neq \pm(L-1)]. \tag{10}$$

With optimal decision thresholds $\{0, \pm 2f_d, \cdots \pm (L-2)f_d\}$, the two conditional probabilities reduce to

$$\Pr[\hat{x}_{k-d} \neq x_{k-d} | x_{k-d} = \pm(L-1)] = \Pr[\epsilon_k > f_d]$$
$$\Pr[\hat{x}_{k-d} \neq x_{k-d} | x_{k-d} \neq \pm(L-1)] = \Pr[|\epsilon_k| > f_d]. \tag{11}$$

Substituting (9) and (11) into (10) yields (4), which is the desired result.

III. Adaptive Algorithm for Minimizing SER

By setting to zero the gradient of (4), we find that the equalizer $\mathbf{c}$ that minimizes error probability must satisfy the following fixed-point relationship:

$$\mathbf{c} = aF(\mathbf{c}) \text{ for any } a > 0 \tag{12}$$

where the vector function $F: \mathcal{R}^N \to \mathcal{R}^N$ is defined by

$$F(\mathbf{c}) = E\left[\exp\left(\frac{-(\mathbf{c}^T \mathbf{H}\tilde{\mathbf{x}})^2}{2\|\mathbf{c}\|^2\sigma^2}\right)\mathbf{H}\tilde{\mathbf{x}}\right]. \tag{13}$$

As in the binary case [8], the fixed-point condition of (12) is necessary but not sufficient to minimize error probability. This is because the error probability surface can have local minima that are not global minima. Although (12) cannot be solved in closed form, the following recursion can be used to find a fixed-point solution:

$$\mathbf{c}_{k+1} = \mathbf{c}_k + \mu F(\mathbf{c}_k) \tag{14}$$

where μ is a positive step size. For known channels, (14) provides a convenient method for numerically computing the coefficients of the minimum-SER equalizer. As in the binary case, careful initialization is needed to avoid undesirable local minima [8].

We now derive an adaptive algorithm that, unlike (14), is suitable for unknown and time-varying channels. Let us first introduce an *error indicator function* I_k that is unity if a decision error occurs at time k and zero otherwise. In terms of the normalized equalizer output $z_k = y_k/f_d$, we may express the indicator function mathematically as $I_k = \psi(z_k \mid x_{k-d})$, where

$$\psi(z \mid x) = \begin{cases} 1, & \text{if } z < x-1 \text{ and } x \neq -L+1 \\ 1, & \text{if } z > x+1 \text{ and } x \neq L-1 \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

In Appendix A, we prove that the indicator function satisfies the following key relationship:

$$E[I_k \text{sign}(-e_k)\mathbf{H}\mathbf{x}_k] = \frac{2L-2}{L} E\left[Q\left(\frac{\mathbf{c}^T \mathbf{H}\tilde{\mathbf{x}}}{\|\mathbf{c}\|\sigma}\right)\mathbf{H}\tilde{\mathbf{x}}\right] \tag{16}$$

where $e_k = z_k - x_{k-d}$ is the error at the output of the normalized equalizer. This relationship is important because the right-hand side of (16) is very nearly the function $F(\mathbf{c})$ of (13). Indeed,

replacing $Q(x)$ in (16) by the approximation $Q(x) \approx e^{-x^2/2}$ [12] leads to the following approximation for (13):

$$F(\mathbf{c}) \approx \frac{L}{2L-2} E[I_k \text{sign}(-e_k)\mathbf{H}\mathbf{x}_k]. \tag{17}$$

Thus, we can use the indicator function to approximate and simplify the deterministic recursion of (14):

$$\begin{aligned}\mathbf{c}_{k+1} &= \mathbf{c}_k + \mu' F(\mathbf{c}_k) \\ &\approx \mathbf{c}_k + \mu E[I_k \text{sign}(-e_k)\mathbf{H}\mathbf{x}_k] \\ &= \mathbf{c}_k - \mu E[I_k \text{sign}(e_k)(\mathbf{r}_k - \mathbf{n}_k)] \\ &\approx \mathbf{c}_k - \mu E[I_k \text{sign}(e_k)\mathbf{r}_k]. \end{aligned} \tag{18}$$

The first approximation follows from (17), with $\mu' = 2\mu(1 - 1/L)$, and the second approximation neglects noise. Although the accuracy of the two approximations in (18) can be quantified analytically [13], they are best justified by the good performance of the resulting algorithm, as demonstrated in Section V.

Removing the expectation in (18) leads to the following asymptotically unbiased stochastic update:

$$\mathbf{c}_{k+1} = \mathbf{c}_k - \mu I_k \text{sign}(e_k)\mathbf{r}_k. \tag{19}$$

This defines the adaptive minimum-SER (AMSER) algorithm. When $L = 2$, (19) reverts back to the binary algorithm proposed in [8].

The AMSER algorithm has the same form as two other well-known adaptive equalization techniques, namely, the LMS algorithm and the *sign-error* LMS algorithm [14]:

$$\begin{aligned}\text{LMS: } \mathbf{c}_{k+1} &= \mathbf{c}_k - \mu(y_k - x_{k-d})\mathbf{r}_k \\ \text{sign-LMS: } \mathbf{c}_{k+1} &= \mathbf{c}_k - \mu \text{sign}(y_k - x_{k-d})\mathbf{r}_k \end{aligned} \tag{20}$$

but with the important distinction that AMSER updates the coefficients *only* when an error occurs (i.e., when $I_k \neq 0$). In retrospect, one could partially justify the presence of the error indicator I_k in the AMSER update by arguing that an equalizer that does not produce an error should not be changed. In contrast, the LMS and sign-LMS algorithms will update the coefficients at every iteration, regardless of whether an error occurs. Upon closer inspection, we see that the AMSER algorithm differs from the sign-LMS algorithm in only two ways: First, the AMSER equalizer only updates when a decision error occurs, and second, the AMSER error signal $e_k = y_k/f_d - x_{k-d}$ is based on the normalized equalizer output, whereas the sign-LMS error signal $y_k - x_{k-d}$ is not. Because it can avoid a floating-point multiplication, AMSER is slightly less complex than the LMS algorithm, with complexity comparable with the sign-LMS algorithm. Despite the striking similarities between AMSER and the LMS-based algorithms, Section V will show that their performance is vastly different.

Evaluating the indicator function I_k in (19) requires knowledge of f_d, which changes with time as $\mathbf{c}$ adapts. Let $\hat{f}_d(k)$ denote the receiver's estimate of f_d at time k. From (3), we see that the ratio y_k/x_{k-d} has mean f_d. A receiver with training can thus track f_d using the simple moving average

$$\hat{f}_d(k+1) = (1-\lambda)\hat{f}_d(k) + \lambda y_k/x_{k-d}$$

where λ is a positive step size.

Because the AMSER algorithm (19) updates only when an error occurs, the convergence rate will be slow when the error rate is low. To speed convergence, the AMSER algorithm can be modified so that it updates not only when an error is made but also when an error is *almost* made as well, i.e., when the distance between the equalizer output and the nearest decision threshold is less than some small positive constant τ. Specifically, the indicator function of (15) may be modified to

$$\psi_\tau(z \mid x) = \begin{cases} 1, & \text{if } z < x - 1 + \tau \text{ and } x \neq -L+1 \\ 1, & \text{if } z > x + 1 - \tau \text{ and } x \neq L-1 \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

When $\tau = 0$, (21) reverts to (15). A nonzero τ can adversely affect the steady-state SER, however, so the choice of τ is a tradeoff between convergence speed and steady-state performance. The design and analysis of algorithm modifications for speeding convergence is an important area for further research but is beyond the scope of this paper. We only note that the similarities of the proposed algorithm with the binary algorithm of [8] will allow the direct application of binary speed-up techniques such as the multistep and infinite-step algorithms of [13], as well as the related technique of [15].

IV. Extension to Complex QAM Systems

In this section, we derive an extension of the AMSER algorithm that is applicable to a complex L^2-QAM system. We still use the model of (1)–(3), as illustrated in Fig. 1, but we now assume that the symbols, filters, and noise are all complex valued. We assume that the real and imaginary parts of the symbols are chosen independently and uniformly from $\{\pm 1, \pm 3, \cdots \pm(L-1)\}$ and that the noise has independent real and imaginary parts, where each is white and Gaussian with zero mean and variance σ^2. In the following, we will use subscripts of R and I to denote real and imaginary parts, respectively, so that we may decompose $\mathbf{x}_k = \mathbf{x}_R + j\mathbf{x}_I$, $\mathbf{r}_k = \mathbf{r}_R + j\mathbf{r}_I$, $\mathbf{c} = \mathbf{c}_R + j\mathbf{c}_I$, $\mathbf{H} = \mathbf{H}_R + j\mathbf{H}_I$, and $\mathbf{n}_k = \mathbf{n}_R + j\mathbf{n}_I$. When convenient, the time index k will be dropped.

By symmetry, the real and imaginary parts of the symbol decisions are equally likely to be incorrect, so that

$$\Pr[\Re\{\hat{x}_k\} \neq \Re\{x_k\}] = \Pr[\Im\{\hat{x}_k\} \neq \Im\{x_k\}]$$

for any complex equalizer $\mathbf{c}$. The derivation that follows will use this probability, call it SER_1, as the optimization criterion. This criterion is essentially equivalent to the error probability for the complex symbols since application of the union bound yields $\Pr[\hat{x}_k \neq x_k] < 2\text{SER}_1$, which is tight for even moderate values of SNR.

The key to our derivation is to express the complex system with N equalizer coefficients as a pair of real systems with $2N$ coefficients. Specifically, the real part of the output of the complex equalizer at time k can be expressed as

$$\begin{aligned} y_R &= \Re\{\mathbf{c}^T\mathbf{r}_k\} \\ &= \mathbf{c}_R^T\mathbf{r}_R - \mathbf{c}_I^T\mathbf{r}_I \\ &= \bar{\mathbf{c}}^T\bar{\mathbf{r}}_k \end{aligned} \tag{22}$$

where we have introduced the real vectors $\bar{\mathbf{c}} = [\mathbf{c}_R^T, -\mathbf{c}_I^T]^T$ and $\bar{\mathbf{r}}_k = [\mathbf{r}_R^T, \mathbf{r}_I^T]^T$, each having $2N$ components. Because

$\mathbf{r}_R = \mathbf{H}_R\mathbf{x}_R - \mathbf{H}_I\mathbf{x}_I + \mathbf{n}_R$ and $\mathbf{r}_I = \mathbf{H}_I\mathbf{x}_R + \mathbf{H}_R\mathbf{x}_I + \mathbf{n}_I$, we may express $\bar{\mathbf{r}}_k$ as

$$\bar{\mathbf{r}}_k = \begin{bmatrix} \mathbf{H}_R & -\mathbf{H}_I \\ \mathbf{H}_I & \mathbf{H}_R \end{bmatrix} \begin{bmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{bmatrix} + \begin{bmatrix} \mathbf{n}_R \\ \mathbf{n}_I \end{bmatrix} \tag{23}$$

$$= \bar{\mathbf{H}}\bar{\mathbf{x}} + \bar{\mathbf{n}} \tag{24}$$

with obvious definitions for $\bar{\mathbf{H}}$, $\bar{\mathbf{x}}$, and $\bar{\mathbf{n}}$. We see that the vector of channel outputs $\bar{\mathbf{r}}_k$ in (24) differs from the vector of channel outputs $\mathbf{r}_k$ in (2) for a *real*-valued PAM system in only two inconsequential ways: First, there are $2N$ outputs instead of N; second, the channel matrix $\bar{\mathbf{H}}$ is not Toeplitz. The latter implies that the components of $\bar{\mathbf{H}}\bar{\mathbf{x}}$ cannot be expressed as a convolution of a real impulse response with the components of $\bar{\mathbf{x}}$, or, in other words, we cannot convert from (24) back to an equation of the form (1). However, this has no impact whatsoever on the analysis and derivations presented in this paper because it was never required that $\mathbf{H}$ be Toeplitz. (Indeed, the AMSER algorithm applies to *arbitrary* multiple-input multiple-output channels [16].)

If we were to adapt the $2N$ real components of $\bar{\mathbf{c}}$ to minimize $\Pr[\Re\{\hat{x}_{k-d}\} \neq \Re\{x_{k-d}\}]$, without regard to the imaginary part of the decision, then a direct application of (19) to the real system of (24) yields

$$\bar{\mathbf{c}}_{k+1} = \bar{\mathbf{c}}_k - \mu I_R \text{sign}(e_R)\bar{\mathbf{r}}_k \tag{25}$$

where $z_R = y_R/\bar{f}_d$ is the normalized equalizer output, and $e_R = z_R - \Re\{x_{k-d}\}$ is the resulting error, where $I_R = \psi(z_R \mid \Re\{x_{k-d}\})$, and where $\bar{f}_d$ is defined by

$$\bar{\mathbf{f}} = [\bar{f}_0, \ \ldots \ \bar{f}_{2M+2N-1}] = \bar{\mathbf{c}}^T\bar{\mathbf{H}}.$$

In terms of the complex filter $\mathbf{f} = \mathbf{c}^T\mathbf{H}$, we may write $\bar{f}_d = \Re\{f_d\}$. Equivalently, in terms of the complex equalizer $\mathbf{c} = \mathbf{c}_R + j\mathbf{c}_I$, (25) can be rewritten as

$$\mathbf{c}_{k+1} = \mathbf{c}_k - \mu I_R \text{sign}(e_R)\mathbf{r}_k^* \tag{26}$$

where $\mathbf{r}^*$ denotes the conjugate of $\mathbf{r}$.

Alternatively, we could instead adapt $\mathbf{c}$ to minimize the probability that the *imaginary* part of the decision is incorrect, ignoring the real part. Specifically, consider the imaginary part of the equalizer output

$$\begin{aligned} y_I &= \Im\{\mathbf{c}^T\mathbf{r}_k\} \\ &= \mathbf{c}_I^T\mathbf{r}_R + \mathbf{c}_R^T\mathbf{r}_I \\ &= \bar{\mathbf{c}}^T\mathbf{P}\bar{\mathbf{r}}_k \end{aligned} \tag{27}$$

where again $\bar{\mathbf{r}}_k$ is given by (24), and where we have introduced the following permutation matrix:

$$\mathbf{P} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{bmatrix}.$$

Like $\bar{\mathbf{r}}_k$, the permuted observation vector $\mathbf{P}\bar{\mathbf{r}}_k$ may also be modeled using (24) but with $\mathbf{P}\bar{\mathbf{H}}$ replacing $\bar{\mathbf{H}}$. If we were to adapt $\mathbf{c}$ in an attempt to minimize $\Pr[\Im\{\hat{x}_{k-d}\} \neq \Im\{x_{k-d}\}]$, ignoring the real part of the decision, then application of (19) yields

$$\bar{\mathbf{c}}_{k+1} = \bar{\mathbf{c}}_k - \mu I_I \text{sign}(e_I)\mathbf{P}\bar{\mathbf{r}}_k \tag{28}$$

where $z_I = y_I/g_{d+N}$ is the normalized equalizer output, and $e_I = z_I - \Im\{x_{k-d}\}$ is the resulting error, where $I_I = \psi(z_I \mid \Im\{x_{k-d}\})$, and where g_{d+N} is defined by

$$\mathbf{g} = [g_0, \ \ldots \ g_{2M+2N-1}] = \bar{\mathbf{c}}^T\mathbf{P}\bar{\mathbf{H}}.$$

In terms of the complex filter $\mathbf{f} = \mathbf{c}^T\mathbf{H}$, we may write $g_{d+N} = \Re\{f_d\}$. In terms of the complex vector $\mathbf{c} = \mathbf{c}_R + j\mathbf{c}_I$, (28) becomes

$$\mathbf{c}_{k+1} = \mathbf{c}_k - \mu j I_I \text{sign}(e_I)\mathbf{r}_k^*. \tag{29}$$

A symmetry argument demonstrates that (26) and (29) converge in the mean to the same steady-state solution so that they are, in that sense, redundant. However, each by itself converges slowly, because each would update the equalizer only when an error occurs in its dimension. The speed of convergence can be roughly doubled by adding (26) to (29), in effect running both simultaneously. This leads to the main result of this section, namely, the following complex version of the AMSER algorithm:

$$\mathbf{c}_{k+1} = \mathbf{c}_k - \mu Q_k \mathbf{r}_k^* \tag{30}$$

where

$$Q_k = \Re\{I_k\}\text{sign}(\Re\{e_k\}) + j\Im\{I_k\}\text{sign}(\Im\{e_k\}) \tag{31}$$

where

$$I_k = \psi(\Re\{z_k\} \mid \Re\{x_{k-d}\}) + j\psi(\Im\{z_k\} \mid \Im\{x_{k-d}\})$$

where $\psi(\cdot \mid \cdot)$ is defined by (21), where $z_k = \mathbf{c}^T\mathbf{r}_k/\Re\{f_d\}$ is the normalized equalizer output and $e_k = z_k - x_{k-d}$ is the resulting error, and where $f_d = \sum_{i=0}^{M} h_i c_{d-i}$. Although the notation in (31) is somewhat cumbersome, it has a straightforward interpretation. Specifically, we may interpret the factor $Q_k \in \{0, \pm 1, \pm j, \pm 1 \pm j\}$ in (30) as a *complex error indicator*; its real part indicates the presence and sign of an error for the real part of the decision, and its imaginary part independently indicates the presence and sign of an error for the imaginary part of the decision.

V. Numerical Results

The performance difference between an *exact minimum-SER (EMSER)* equalizer and an MMSE equalizer is most pronounced when the number of equalizer coefficients is small relative to the severity of the ISI. For example, consider the real channel $H(z) = 0.5 + z^{-1}$ with 4-PAM and only $N = 2$ equalizer coefficients. In Fig. 2, we compare the learning curves for AMSER and LMS by plotting SER versus time, averaged over 1000 independent trials. The AMSER equalizer used the LMS update for the first 200 symbol periods as initialization and then switched to the AMSER update of (19) with $\tau = 0.1$ thereafter. Both equalizers used a delay of $d = 2$, which minimizes SER for the MMSE equalizer, and a step size $\mu = 10^{-3}$. The SNR was 27 dB, where $\text{SNR} = E_a \sum_k |h_k|^2/(2\sigma^2)$, and $E_a = E[|a_k|^2]$ is the alphabet energy. We observe from the figure that the AMSER algorithm significantly outperforms the LMS algorithm, achieving a steady-state SER that is 15 times smaller. Furthermore, just as the LMS algorithm closely matches the performance of the exact MMSE equalizer, the AMSER algorithm comes close to achieving the optimal performance of the EMSER equalizer.

We now test the performance of the proposed algorithm as a function of SNR by considering another 4-PAM example, this time against the channel $H(z) = 0.66 + z^{-1} - 0.66z^{-2}$, with

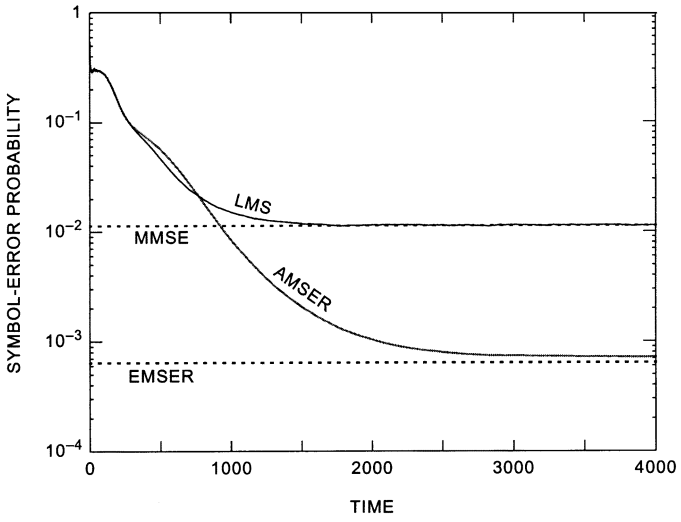


Fig. 2. Learning curves for the LMS and AMSER adaptive equalizers.

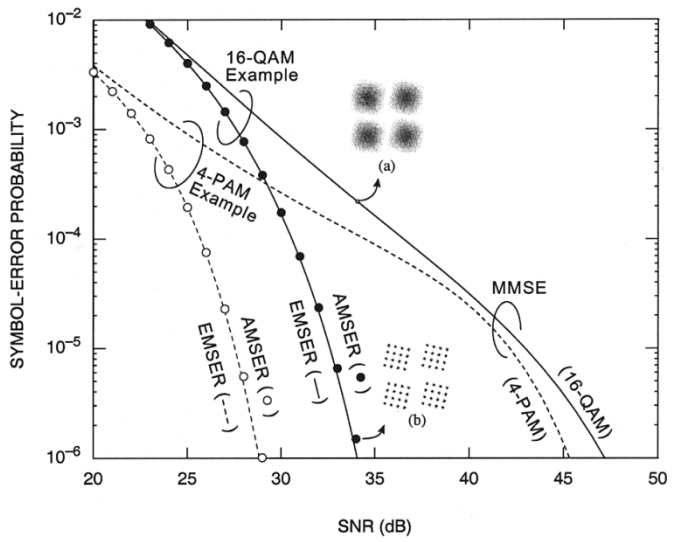


Fig. 3. Performance comparison for the five-tap 4-PAM example (dashed lines and open circles) and the 16-QAM example (solid lines and filled circles). The points labeled (a) and (b) correspond to the constellations of Fig. 4(a) and (b), respectively.

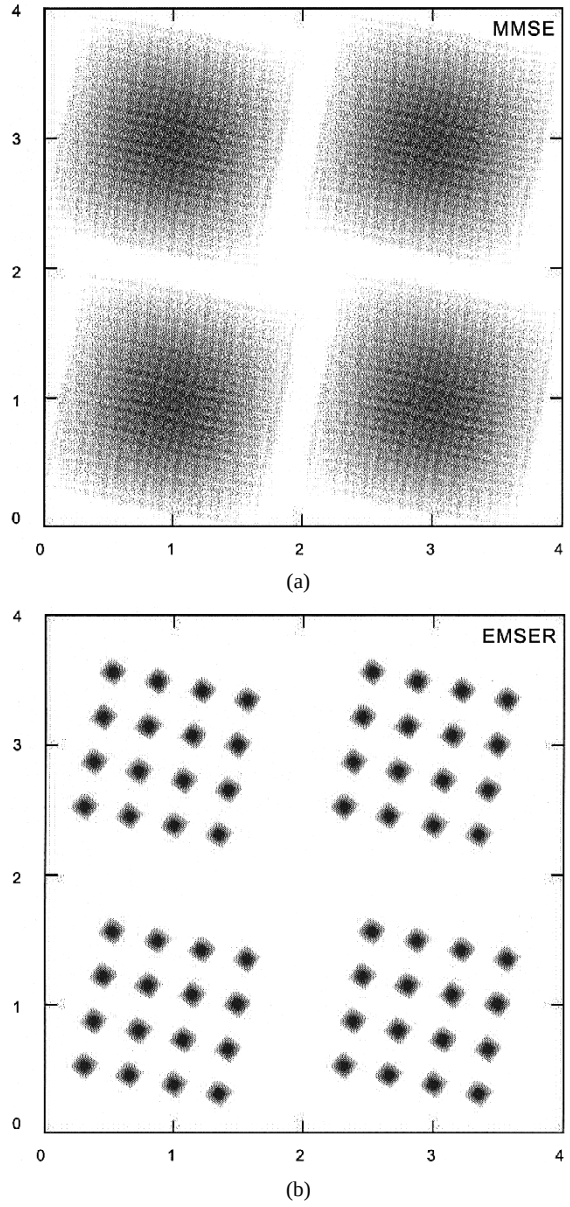


Fig. 4. First quadrants of the noiseless constellations for the 16-QAM example with SNR = 34 dB. (a) After the MMSE equalizer (with MSE $= -6.3$ dB, and $\text{SER}_1 = 2.2 \times 10^{-4}$). (b) After the EMSER equalizer (with MSE $= -4.4$ dB and $\text{SER}_1 = 1.1 \times 10^{-6}$).

$N = 5$ equalizer coefficients, and with a delay of $d = 3$ (which minimizes SER for the MMSE equalizer). In Fig. 3, we plot SER versus SNR $= E_a \sum_k |h_k|^2/(2\sigma^2)$ for the five-tap EMSER, AMSER, and MMSE equalizers. The coefficients of the MMSE and EMSER equalizers were calculated exactly, whereas the AMSER coefficients were obtained via the stochastic update (19), with $\mu = 0.0002$, $\tau = 0.05$, and after 10^6 iterations with training. The small step size and long training period were chosen to diminish the impact of transient effects and allow us to test the steady-state performance of the algorithm. The SER for all three equalizers was then evaluated using (4). Observe from Fig. 3 that the performance of AMSER is almost indistinguishable from that of the EMSER equalizer and that the AMSER equalizer outperforms the MMSE equalizer by over 16 dB at high SNR.

To demonstrate the effectiveness of AMSER for large complex alphabets, consider a 16-QAM system with $H(z) = (1.2 + j) + (1.6 - 1.7j)z^{-1}$ and $N = 4$ equalizer coefficients. A plot of SER_1 versus SNR for the four-tap EMSER, AMSER, and MMSE equalizers is shown in Fig. 3. In all cases, the delay is $d = 4$, which minimizes SER_1 for the MMSE equalizer. The MMSE and EMSER coefficients were calculated exactly, and the AMSER coefficients were obtained adaptively via (30) with $\mu = 10^{-5}$ and $\tau = 0.05$ after 10^6 training symbols. The SER_1 for all three equalizers was evaluated using (4). AMSER outperforms MMSE by more than 13 dB.

It is instructive to compare the noiseless constellation diagrams after the MMSE and EMSER equalizers for the previous 16-QAM example, as illustrated in Fig. 4 for SNR =3.4 dB. (Only the first quadrant is shown; the other three quadrants are translates of the first.) The constellations from Fig. 4(a) and (b) correspond to the MMSE point labeled (a) and the EMSER point in Fig. 3, respectively. The EMSER equalizer is scaled to have

the same norm—and therefore the same noise enhancement—as the MMSE equalizer. The MMSE equalizer penalizes all points that stray from the desired, even those that are unlikely to cause an error. The clusters that result after the MMSE equalizer are roughly Gaussian in shape with significant tails, causing the distance between neighboring clusters to be small. In contrast, the EMSER equalizer exploits the non-Gaussianity of the ISI. The clusters after the EMSER equalizer are markedly non-Gaussian, and the distance between clusters after EMSER is significantly larger than for the MMSE case. Thus, although the MSE of the EMSER equalizer is 1.9 dB greater than that of the MMSE equalizer, its SER is smaller by a factor of 200.

A rigorous convergence analysis is beyond the scope of this paper. We only mention two facts. First, empirically, the AMSER algorithm has exhibited good convergence properties in our simulations. Second, the close link between the AMSER and LMS algorithms suggests that they may share similar convergence properties.

VI. Conclusions

The AMSER algorithm is an adaptive equalization technique that approximates the minimum-SER equalizer for PAM and QAM, with complexity comparable to that of the LMS algorithm. The minimum-SER equalizer can be far superior to the MMSE equalizer when the number of equalizer coefficients is small. Future work should devise methods for speeding up convergence.

Appendix
Derivation of (16)

Let $q_k = I_k \text{sign}(-e_k)$, where $e_k = y_k/f_d - x_{k-d}$ is the error at the output of the normalized equalizer, and $I_k \in \{0,1\}$ is the error indicator function of (15). Thus, $q_k \in \{0, \pm 1\}$. Let $\mathbf{x}$ be uniformly distributed over the set Ω of all L^{M+N} vectors of PAM symbols, and let $\tilde{\mathbf{x}}$ be uniformly distributed over the subset for which the dth location is unity, $x_{k-d} = 1$. Let Ω_{right}, Ω_{left}, and Ω_{inner} denote the subsets of Ω for which $x_{k-d} = (L-1)$, $x_{k-d} = -(L-1)$, and $x_{k-d} \neq \pm(L-1)$, respectively. Then

$$\begin{aligned} E[q_k\mathbf{x}] &= \frac{1}{L}E[q_k\mathbf{x} \mid \mathbf{x} \in \Omega_{\text{right}}] \\ &\quad + \frac{1}{L}E[q_k\mathbf{x} \mid \mathbf{x} \in \Omega_{\text{left}}] \\ &\quad + \frac{L-2}{L}E[q_k\mathbf{x} \mid \mathbf{x} \in \Omega_{\text{inner}}] \qquad (32) \\ &= \frac{1}{L}E\Big[E\big[q_k \mid \mathbf{x}, \mathbf{x} \in \Omega_{\text{right}}\big]\mathbf{x} \mid \mathbf{x} \in \Omega_{\text{right}}\Big] \\ &\quad + \frac{1}{L}E\Big[E\big[q_k \mid \mathbf{x}, \mathbf{x} \in \Omega_{\text{left}}\big]\mathbf{x} \mid \mathbf{x} \in \Omega_{\text{left}}\Big] \\ &\quad + \frac{L-2}{L}E\Big[E\big[q_k \mid \mathbf{x}, \mathbf{x} \in \Omega_{\text{inner}}\big] \\ &\qquad \times \mathbf{x} \mid \mathbf{x} \in \Omega_{\text{inner}}\Big]. \qquad (33) \end{aligned}$$

Since $q_k = -1$ is impossible when $\mathbf{x} \in \Omega_{\text{right}}$, and since $q_k = +1$ is impossible when $\mathbf{x} \in \Omega_{\text{left}}$, we have

$$\begin{aligned} E[q_k \mid \mathbf{x}, \mathbf{x} \in \Omega_{\text{right}}] &= \Pr[q_k = +1 \mid \mathbf{x}, \mathbf{x} \in \Omega_{\text{right}}] \\ &= Q\left(\frac{f_d + \sum_{i \neq d} f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right) \qquad (34) \\ E[q_k \mid \mathbf{x}, \mathbf{x} \in \Omega_{\text{left}}] &= -\Pr[q_k = -1 \mid \mathbf{x}, \mathbf{x} \in \Omega_{\text{left}}] \\ &= -Q\left(\frac{f_d - \sum_{i \neq d} f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right) \qquad (35) \\ E[q_k \mid \mathbf{x}, \mathbf{x} \in \Omega_{\text{inner}}] &= \Pr[q_k = +1 \mid \mathbf{x}, \mathbf{x} \in \Omega_{\text{inner}}] \\ &\quad - \Pr[q_k = -1 \mid \mathbf{x}, \mathbf{x} \in \Omega_{\text{inner}}] \\ &= Q\left(\frac{f_d + \sum_{i \neq d} f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right) \\ &\quad - Q\left(\frac{f_d - \sum_{i \neq d} f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right). \qquad (36) \end{aligned}$$

Substituting (34)–(36) into (33) yields

$$\begin{aligned} E[q_k\mathbf{x}] &= \frac{1}{L}E\left[Q\left(\frac{f_d + \sum_{i \neq d} f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right)\mathbf{x} \mid \mathbf{x} \in \Omega_{\text{right}}\right] \\ &\quad - \frac{1}{L}E\left[-Q\left(\frac{f_d - \sum_{i \neq d} f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right)\mathbf{x} \mid \mathbf{x} \in \Omega_{\text{left}}\right] \\ &\quad + \frac{L-2}{L}E\Bigg[\Bigg\{Q\left(\frac{f_d + \sum_{i \neq d} f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right) \\ &\qquad -Q\left(\frac{f_d - \sum_{i \neq d} f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right)\Bigg\} \\ &\qquad \times \mathbf{x} \mid \mathbf{x} \in \Omega_{\text{inner}}\Bigg]. \qquad (37) \end{aligned}$$

However, the symmetry of the PAM alphabet implies that

$$p(-\mathbf{x} \mid \mathbf{x} \in \Omega_{\text{left}}) = p(\mathbf{x} \mid \mathbf{x} \in \Omega_{\text{right}})$$

and it also implies that

$$p(-\mathbf{x} \mid \mathbf{x} \in \Omega_{\text{inner}}) = p(\mathbf{x} \mid \mathbf{x} \in \Omega_{\text{inner}})$$

so that (37) reduces to

$$\begin{aligned} E[q_k\mathbf{x}] &= \frac{2}{L}E\left[Q\left(\frac{f_d + \sum_{i \neq d} f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right)\mathbf{x} \mid \mathbf{x} \in \Omega_{\text{right}}\right] \\ &\quad + 2\frac{L-2}{L}E\left[Q\left(\frac{f_d + \sum_{i \neq d} f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right)\mathbf{x} \mid \mathbf{x} \in \Omega_{\text{inner}}\right]. \qquad (38) \end{aligned}$$

We consider separately the dth component of $E[q_k\mathbf{x}]$ from the others. First, the dth component of (38) is

$$E[q_k\mathbf{x}]_d = \frac{2}{L}E\left[Q\left(\frac{f_d+\sum_{i\neq d}f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right)(L-1)\right] + 2\frac{L-2}{L}E\left[Q\left(\frac{f_d+\sum_{i\neq d}f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right) \times x_{k-d} \mid \mathbf{x}\in\Omega_{\text{inner}}\right] \quad (39)$$

$$= \frac{2L-2}{L}E\left[Q\left(\frac{f_d+\sum_{i\neq d}f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right)\right] \quad (40)$$

where the fact that x_{k-d} is independent of

$$Q\left(\frac{f_d+\sum_{i\neq d}f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right)$$

implies that the second expectation of (39) is zero (since $E[x_{k-d} \mid \mathbf{x}\in\Omega_{\text{inner}}]=0$). Observe that it is no longer necessary to condition the first expectation of (39) on the event $\mathbf{x}\in\Omega_{\text{right}}$ since the argument of the expectation is independent of x_{k-d}. Consider next the nth component of $E[q_k\mathbf{x}]$ with $n\neq d$; from (38), we see that it does not contain a contribution from x_{k-d} so that the conditioning in (38) on $\mathbf{x}\in\Omega_{\text{right}}$ or $\mathbf{x}\in\Omega_{\text{inner}}$ has no impact and that both expectations can equivalently be computed with respect to all $\mathbf{x}\in\Omega$. Hence, from (38)

$$E[q_k\mathbf{x}]_{n\neq d} = \frac{2}{L}E\left[Q\left(\frac{f_d+\sum_{i\neq d}f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right)x_{k-n}\right] + 2\frac{L-2}{L}E\left[Q\left(\frac{f_d+\sum_{i\neq d}f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right)x_{k-n}\right] \quad (41)$$

$$= \frac{2L-2}{L}E\left[Q\left(\frac{f_d+\sum_{i\neq d}f_i x_{k-i}}{\|\mathbf{c}\|\sigma}\right)x_{k-n}\right]. \quad (42)$$

Combining (40) and (42) yields

$$E[q_k\mathbf{x}] = \frac{2L-2}{L}E\left[Q\left(\frac{\mathbf{c}^T\mathbf{H}\tilde{\mathbf{x}}}{\|\mathbf{c}\|\sigma}\right)\hat{\mathbf{x}}\right]. \quad (43)$$

Left-multiplying both sides of (43) by the deterministic matrix $\mathbf{H}$ yields (16), which is the desired result. Q.E.D.

REFERENCES

[1] M. R. Aaron and D. W. Tufts, "Intersymbol interference and error probability," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 26–34, Jan. 1966.

[2] E. Shamash and K. Yao, "On the structure and performance of a linear decision feedback equalizer based on the minimum error probability criterion," in *Proc. IEEE Int. Conf. Commun.*, 1974, pp. 25F1–25F5.

[3] P. Galko and S. Pasupathy, "Optimization of linear receivers for data communications signals," *IEEE Trans. Inform. Theory*, vol. 34, pp. 79–92, Jan. 1988.

[4] S. Chen, E. Chng, B. Mulgrew, and G. Gibson, "Minimum-BER linear-combiner DFE," in *Proc. IEEE Int. Conf. Commun.*, 1996, pp. 1173–1177.

[5] C.-C. Yeh and J. R. Barry, "Approximate minimum bit-error rate equalization for binary signaling," in *IEEE Int. Conf. Commun.*, vol. 2, Montreal, QC, Canada, June 8–12, 1997, pp. 1095–1099.

[6] J. Tellado and J. M. Cioffi, "Quasiminimum-BER linear combiner equalizers," in *Proc. IEEE Int. Conf. Commun.*, vol. 1, Atlanta, GA, June 7–11, 1998, pp. 1–5.

[7] I. N. Psaromiligkos, S. N. Batalama, and D. A. Pados, "On adaptive minimum probability of error linear-filter receivers for DS-CDMA channels," *IEEE Trans. Commun.*, vol. 47, pp. 1092–1102, July 1999.

[8] C.-C. Yeh and J. R. Barry, "Adaptive minimum bit-error rate equalization for binary signaling," *IEEE Trans. Commun.*, vol. 48, pp. 1226–1236, July 2000.

[9] ——, "Approximate minimum bit-error rate equalization for pulse-amplitude and quadrature amplitude modulation," in *Proc. IEEE Int. Conf. Commun.*, vol. 1, Atlanta, GA, June 7–11, 1998, pp. 16–20.

[10] S. Chen, B. Mulgrew, and L. Hanzo, "Stochastic least-symbol-error-rate adaptive equalization for pulse-amplitude modulation," in *Proc. ICASSP*, Orlando, FL, May 2002, pp. 2629–2632.

[11] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney Jr, "MMSE decision-feedback equalizers and coding, Part I: Equalization results," *IEEE Trans. Commun.*, vol. 43, pp. 2582–94, Oct. 1995.

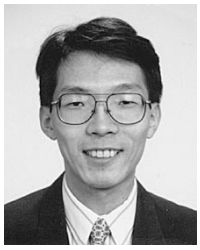
[12] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communication*, Third ed. Boston, MA: Kluwer, 2003.

[13] C.-C. Yeh, "Adaptive minimum-error-probability equalization and multiuser detection," Ph.D. dissertation, Georgia Inst. Technol., Atlanta, GA, 1998.

[14] N. A. M. Verhoeckx, H. C. Van Den Elzen, F. A. M. Snijders, and P. J. V. Gerwen, "Digital echo cancellation for baseband data transmission," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 768–781, Dec. 1979.

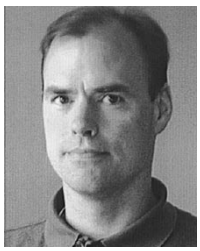
[15] S. Chen, A. Samingan, B. Mulgrew, and L. Hanzo, "Adaptive minimum-ber linear multiuser detection for DS-CDMA signals in multipath channels," *IEEE Trans. Signal Processing*, vol. 49, pp. 1240–1247, June 2001.

[16] C.-C. Yeh, R. R. Lopes, and J. R. Barry, "Approximate minimum bit-error rate multiuser detection," in *Proc. IEEE Global Commun. Conf.*, Sydney, Australia, Nov. 1998, pp. 3590–3595.



Chen-Chu Yeh (M'99) received the B.S. degree in electrical engineering from Cornell University, Ithaca, NY, in 1992 and the M.S. and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1994 and 1998, respectively.

From 1998 to 2000, he was a Systems Analyst with the Communications and Signal Processing Department, Harris Corporation, Palm Bay, FL, working on various aspects of space and airborne communications systems. Since 2000, he has been with Intersil Corporation, Palm Bay, FL, where he is a Staff Systems Engineer working on baseband processor design for wireless LAN chipsets.



John R. Barry (M'93) received the B.S. degree in electrical engineering from the State University of New York at Buffalo in 1986 and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Berkeley in 1987 and 1992, respectively.

Since 1992, he has been with the Georgia Institute of Technology, Atlanta, where he is an Associate Professor with the School of Electrical and Computer Engineering. His research interests include wireless communications, equalization, and multiuser communications. He is the author of *Wireless Infrared Communications* (Boston, MA: Kluwer, 1994) and a coauthor of *Digital Communications* (Boston, MA: Kluwer, 2004, Third ed.).